Convexities and optimal transport problems on the Wiener space

Vincent Nolot

UNIVERSITE DE BOURGOGNE

UFR Sciences et Techniques

Institut de Mathématiques de Bourgogne

THESE

pour obtenir le grade de

Docteur de l'Université de Bourgogne

Discipline : MATHEMATIQUES

par

Vincent Nolot

Convexités et problèmes de transport optimal sur l'espace de Wiener.

Soutenue publiquement le **27 Juin 2013** devant le Jury composé de

Bernard BONNARD	Université de Bourgogne	(examinateur)
Guillaume CARLIER	Université Paris Dauphine	(examinateur)
Luigi DE PASCALE	Université de Pise	(rapporteur)
Shizan FANG	Université de Bourgogne	(directeur de thèse)
Ivan GENTIL	Université de Lyon	(examinateur)
Nicolas PRIVAULT	Université de Singapour	(rapporteur)

Résumé en Français

L'objet de cette thèse est d'étudier la théorie du transport optimal sur un espace de Wiener abstrait. Les résultats qui se trouvent dans quatre principales parties, portent

- Sur la convexité de l'entropie relative. On prolongera des résultats connus en dimension finie, sur l'espace de Wiener muni d'une norme uniforme, à savoir que l'entropie relative est (au moins faiblement) $1-$convexe le long des géodésiques induites par un transport optimal sur l'espace de Wiener.

- Sur les mesures à densité logarithmiquement concaves. Le premier des résultats importants consiste à montrer qu'une inégalité de type Harnack est vraie pour le semi-groupe induit par une telle mesure sur l'espace de Wiener. Le second des résultats obtenus nous fournit une inégalité en dimension finie (mais indépendante de la dimension), contrôlant la différence de deux applications de transport optimal.

- Sur le problème de Monge. On s'intéressera au problème de Monge sur l'espace de Wiener, muni de plusieurs normes : des normes à valeurs finies, ou encore la pseudo-norme de Cameron-Martin.

- Sur l'équation de Monge-Ampère. Grâce aux inégalités obtenues précédemment, nous serons en mesure de construire des solutions fortes de l'équation de Monge-Ampère (induite par le coût quadratique) sur l'espace de Wiener, sous de faibles hypothèses sur les densités des mesures considérées.

Mots clés : transport optimal, problème de Monge, convexité, espace de Wiener, équation de Monge-Ampère, dimension infinie, mesure logarithmiquement concave.

Abstract in english

The aim of this PhD is to study the optimal transportation theory in some abstract Wiener space. You can find the results in four main parts and they are about

- The convexity of the relative entropy. We will extend the well known results in finite dimension to the Wiener space, endowed with the uniform norm. To be precise the relative entropy is (at least weakly) geodesically $1-$convex in the sense of the optimal transportation in the Wiener space.

- The measures with logarithmic concave density. The first important result consists in showing that the Harnack inequality holds for the semi-group induced by such a measure in the Wiener space. The second one provides us a finite dimensional and dimension-free inequality which gives estimate on the difference between two optimal maps.

- The Monge Problem. We will be interested in the Monge Problem on the Wiener endowed with different norms: either some finite valued norms or the pseudo-norm of Cameron-Martin.

- The Monge-Ampère equation. Thanks to the inequalities obtained above, we will be able to build strong solutions of the Monge-Ampère (those which are induced by the quadratic cost) equation on the Wiener space, provided the considered measures satisfy weak conditions.

Key words: optimal transport, Monge problem, convexity, Wiener space, Monge-Ampère equation, infinite dimension, logarithmic concave measure.

Remerciements

Mes remerciements pour l'accomplissement de ce travail s'adressent principalement à Shizan Fang, qui m'a supervisé, conseillé, orienté pendant ces trois années. Tout cela a toujours été accompagné d'enthousiasme et d'encouragements, en particulier dans les moments difficiles. Je lui adresse toute ma reconnaissance. Ce travail n'aurait jamais vu le jour sans le soutien de Patrick Gabriel, qui a co-encadré mon mémoire de recherche en master. Patrick fait partie des personnes qui m'ont scientifiquement et humainement apporté le plus, au sein du laboratoire. Je le remercie d'avoir partagé sa grande ouverture d'esprit sur les mathématiques, l'enseignement et bien au-delà.

J'ai le plaisir de remercier Nicolas Privault et Luigi De Pascale qui m'ont fait l'honneur de rapporter ma thèse, et tout autant les autres membres de mon jury, Bernard Bonnard, Guillaume Carlier et Ivan Gentil. Leur expertise dans des domaines variés est largement reconnue. Je tiens également à remercier Robert McCann qui m'accueille à l'Université de Toronto, en ce moment même où j'écris ces lignes.

Parce que faire une thèse, c'est aussi parfois rencontrer au-delà des mathématiciens, des personnalités intéressantes, ouvertes, qui n'hésitent pas à aider les jeunes chercheurs, et sans qui la motivation redescendrait trop vite; je tiens à remercier Nicolas Juillet, pour m'avoir accueilli à Strasbourg avec beaucoup de sympathie dès le début de ma thèse, ainsi que pour tous les autres bons moments que l'on a vécu aux conférences où l'on se retrouvait. Thierry Champion qui m'a grandement encouragé dans mes travaux durant un colloque à Orsay, puis dans nos rencontres Dijonaises. Pierre-André Zitt dont l'humour n'est plus à démontrer, qui était présent pour les deux premières années de ma thèse, a toujours été curieux et à l'écoute. Merci à Bernard Bonnard pour les relations d'amitié que l'on a liées tout au long de ces trois années. Je voudrais saluer mon demi-frère de thèse, Camille Tardif qui est une personne aux grandes qualités humaines, et je ne regrette que le fait qu'il aie passé plus de temps à Strasbourg plutôt qu'à Dijon. Merci finalement aux membres de mon équipe, l'équipe SPAN, pour les initiatives PodEx et tout le reste.

Les conditions de travail que le staff de l'IMB a mises à disposition étaient particulièrement adéquates. Un grand merci aux agents d'entretien, notamment Aziz pour son sourire quotidien. Un grand merci aux secrétaires pour leur dévouement, et plus spécifiquement à Caroline, qui s'est occupée avec attention de toutes mes missions, et avec qui j'ai toujours eu beaucoup de plaisir à échanger des histoires plus ou moins amusantes. A elles s'ajoutent notre bibliothécaire Pierre et notre informaticien Francis, qui sont au coeur du bon fonctionnement du laboratoire.

Trois années de vie commune avec les différents doctorants et post-doctorants du laboratoire, avec qui on pouvait partager nos sentiments sur le travail de recherche. Ces impressions que l'on découvre au cours d'une thèse et que les doctorants sont certainement les mieux à même de considérer. Merci à vous pour l'environnement agréable que vous avez créé, et j'espère que notre association tant aimée continuera son ascension. J'ai une pensée particulière à tous mes co-bureaux, et je ne citerai qu'eux (pour ne pas en oublier dautres) : Gautier, Gabriel, Pauline, Eglantine, Martin, Yi Shi et ce bon vieil Alvaro. Autant de personnes qui ont contribué à ce que le bureau 213 devienne l'un des plus emblématiques du laboratoire.

On ne devient pas docteur du jour au lendemain, mais après une succession d'évènements, une longue poursuite des études qui demandent de la persévérance, et c'est pourquoi je n'oublie pas mes amis qui m'ont permis de m'évader du monde des mathématiques et en particulier au cours de ces trois dernières années. Une pensée particulière à Gaëtan avec qui j'ai fait toute ma scolarité à l'Université de Bourgogne. Merci pour l'estime que tu as eue pour moi, cela m'a sans aucun doute encouragé dans mon parcours. A tous les autres, des pays de Langres, dijonais ou d'ailleurs pour les soirées et vacances emplies de joie et de bonne humeur. Au même niveau je remercie chaque membre du club Langres Natation 52, avec qui j'ai noué des liens très forts. Partenaires d'entraînements, de stages, de compétitions, merci ! Sous la tutelle de Rémy, quel bonheur de se retrouver dans l'eau avec vous pour souffrir physiquement, décompresser et se vider la tête. Je ne remercierai jamais assez mon ami Jean Cote, qui m'a enlevé ce fardeau de responsabilités au club, afin d'accomplir au mieux mon travail de recherche et d'enseignement. Merci Jean pour tout ce que tu m'as appris sur tant de domaines différents, en si peu de temps, et j'espère que cela n'est pas fini.

Je remercie ma famille, et notamment mes parents qui m'ont toujours poussé et m'ont à chaque fois donné les moyens de réussir mes études. Egalement mon frère qui me motivait davantage, en disant que les maths auront toujours un train de retard . Grâce à eux j'ai pu développer un esprit critique et acquérir de la rigueur.

Enfin, je voudrais remercier Alice, que j'ai rencontrée pendant ma thèse. Ses réflexions et nos discussions ont toujours été fructueuses, et je lui dois beaucoup en termes de motivation. Elle a contribué à m'ouvrir l'esprit et m'a soutenu considérablement pour la fin de ma thèse. Merci mon Alice.

Contents

Chapter 1

Introduction

Des problèmes mathématiques, laissés parfois à l'abandon pendant plusieurs siècles, peuvent refaire surface, être redécouverts et réinvestis pour prendre une envergure très importante. C'est le cas du problème économique posé par l'ingénieur-mathématicien français Monge en 1781 dans une note à l'Académie des Sciences. Gaspard Monge, né d'ailleurs non loin d'ici (Beaune), s'est demandé s'il existait un moyen de transporter un déblais vers un remblais, de façon la plus économique possible. La plus économique possible signifie que l'on connaît parfaitement le coût de transport occasionné pour déplacer une partie du déblais vers une autre du remblais. Cela revient mathématiquement à se donner une fonction (appelée fonction de coût), qui est donc au préalable de l'étude connue, et la question est de savoir s'il existe des applications mesurables (moyen de transport) envoyant une mesure (le déblais) vers une autre (le remblais). Monge a formulé ce problème à priori très concret, en des termes mathématiques rigoureux (voir ses notes à l'Académie des Sciences [52]).

Le problème qui paraît pourtant simple, s'avère particulièrement compliqué, et Monge lui-même n'a pu le résoudre à son époque. Il a fallu attendre les années 2000 (plus de deux siècles plus tard !) pour que le problème de Monge, de la manière dont son auteur l'a posé, fut résolu. Oui, il existe un moyen d'effectuer le transport (une application de transport) afin que le coût global soit le moins cher possible. La solution est apportée indépendamment par de grands mathématiciens, à savoir Ambrosio dans [3], ou Trüdinger et Wang dans [57]. Un petit bémol pourtant pour les ingénieurs, les mathématiques nous assurent l'existence d'une solution, mais ne nous donnent pas le moyen de faire en pratique ! Sauf cas bien précis, lorsque le coût de transport a une forme particulière (vaut 0 ou 1), rien ne nous permet de dire quelle quantité doit être envoyée à tel ou tel autre endroit. La curiosité mathématique a conduit à un engouement extrêmement rapide, étoffant ainsi la théorie, connue aujourd'hui sous le nom de *théorie du transport optimal.*

Au départ, il paraît naturel (et c'est comme cela que Monge l'a introduit) de dire que le prix que l'on paye pour déplacer une quantité d'un endroit à un autre, dépend de la distance entre le point de départ et celui d'arrivée. Ainsi modéliser le coût de transport entre deux points par la distance entre ces points semble raisonnable. Si ρ_0 est une mesure représentant la quantité à transporter, ρ_1 une mesure représentant le lieu d'arrivée de la quantité, et T une application (un moyen de faire) qui transporte ρ_0 sur ρ_1 alors le coût total de déplacement de ρ_0 vers ρ_1 est donné par la quantité

$$\int_{\mathbb{R}^2} |x - T(x)| d\rho_0(x).$$

Puisque notre soucis est de trouver un moyen (une application) qui minimise ce coût de transport global, le problème de Monge à résoudre s'écrit mathématiquement

$$\inf_{T_\# \rho_0 = \rho_1} \int_{\mathbb{R}^2} |x - T(x)| d\rho_0(x),$$

où la contrainte $T_\# \rho_0 = \rho_1$ correspond à envoyer la mesure ρ_0 sur la mesure ρ_1 par le biais de l'application T. Cette contrainte n'est pas agréable du tout, puisqu'elle est hautement non linéaire et non convexe, ce qui rend le problème absolument délicat à résoudre.

Les derniers auteurs cités se sont appuyés sur des travaux très conséquents réalisés à partir du milieu du 20e siècle, comme ceux de Kantorovich. Ce mathématicien et économiste russe relaxa le problème de Monge en un problème d'optimisation convexe, cela lui a valu l'obtention du Prix Nobel d'Economie. Le premier mathématicien qui proposa une preuve de l'existence de l'application optimale T fut Sudakov, mais sa preuve n'est pas correcte car elle repose sur un fait de désintégration qui ne fournit pas toujours les informations suffisantes. Ou encore le mathématicien français Brenier qui fut le premier à caractériser les applications de transport optimal dans le cadre du coût euclidien au carré.

Les mathématiciens aimant généraliser les résultats, à des ensembles de plus en plus abstraits, le problème de Monge actuel prend la forme

$$\inf_{T_\# \rho_0 = \rho_1} \int_X d(x, T(x)) d\rho_0(x),$$

où les contraintes sont les mêmes, et (X, d) est un espace (suffisamment gentil tout de même) Polonais, ou encore de longueur (voir Gigli [42]). Très vite, on trouve dans la littérature des problèmes similaires, où d'autres coûts de transports sont considérés. La raison première est que le problème de Monge faisant intervenir la distance est difficile à résoudre, de part le caractère trop peu régulier du coût : en effet la fonction distance, même si elle provient d'une norme, n'est pas strictement convexe en tant que fonction, et ne vérifie pas la condition (Twist) introduite dans le Chapitre 3. C'est ainsi qu'un des premiers travaux fournissant une application de transport *optimal* (c'est-à-dire solution du Problème) est celui de Brenier [14], où le coût considéré est la distance au carré. Le fait de regarder la distance à la puissance p où $p > 1$ simplifie grandement la résolution du problème, puisque la fonction de coût gagne suffisamment en régularité.

Revenons sur le fait que le contrainte $T_{\#}\rho_0 = \rho_1$ ne soit pas agréable. Elle correspond à imposer que l'application T envoie notre première mesure ρ_0 sur la deuxième ρ_1. Justifications à part, si nos mesures sont absolument continues (par rapport à Lebesgue par exemple) de densités respectives f_0 et f_1, la condition peut se traduire par le fait que l'application T doit résoudre une équation aux dérivées partielles bien connue, celle de Monge-Ampère :

$$f_1(T)|det(\nabla T)| = f_0.$$

Lorsqu'un problème d'optimisation est délicat à résoudre de part ses contraintes difficilement manipulables, une manière de procéder est de relaxer le problème. Il se trouve que Kantorovich a proposé un problème, qui au lieu de transporter une mesure vers une autre par une application, couple ces deux mesures ensemble. Le fait de coupler correspond mathématiquement à trouver une mesure sur l'espace produit et dont les marginales sont précisément ρ_0 et ρ_1. Il porte dorénavant le nom de Problème de Monge-Kantorovich et s'énonce ainsi

$$\min_{\Pi \in C(\rho_0,\rho_1)} \int_{X\times X} c(x,y)d\Pi(x,y),$$

avec $C(\rho_0, \rho_1)$ l'ensemble des couplages entre ρ_0 et ρ_1, et c la fonction de coût. Cette fois la contrainte est convexe, et la fonctionnelle qui à un couplage associe le coût de transport total étant linéaire, ce problème est particulièrement facile à résoudre : une solution (un *couplage optimal*) existe toujours dès lors que l'on suppose un minimum de régularité sur la fonction de coût, par exemple c étant semi-continue inférieurement. D'un point de vue pratique, la différence entre le Problème de Monge et celui de Monge-Kantorovich s'explique comme suit : le premier problème consiste à transporter chaque quantité telle quelle, tandis que le second autorise à séparer la masse du départ et envoyer les différentes parties vers différents endroits.

De ces deux problèmes (Monge et Monge-Kantorovich) nait la théorie du transport optimal. L'ampleur de la théorie est telle, qu'elle fournit d'inombrables et inattendues applications : en géométrie, en probabilité, en théorie des jeux... Dans cette thèse on s'intéresse à la théorie du transport optimal en dimension infinie. En effet malgré un gros engouement en dimension finie, on trouve peu de résultats sur les espaces de dimension infinie. On s'intéressera notamment aux espaces de Wiener abstraits, et souvent à l'espace classique de Wiener. Un espace de Wiener est le cadre naturel de généralisation des espaces de dimension finie. Il consiste en la donnée d'un espace de Hilbert H, qui s'injecte dans un espace Polonais (X, d), muni d'une Gaussienne μ portée par X, appelée mesure de Wiener et généralisant les mesures Gaussiennes sur $\mathbb{R}^n$. D'un point de vue probabiliste, la mesure de Wiener est la loi du mouvement Brownien. Rappelons qu'il n'existe pas de mesure de Lebesgue en dimension infinie, et qu'une mesure gaussienne est certainement son meilleur substitut. Les difficultés rencontrées dans ces espaces proviennent de plusieurs faits :

- l'aspect local est ardu, les compacts sont d'intérieur vide, et un outil très important en dimension finie n'est en général plus valable pour la mesure de Wiener : le théorème de différentiation de Lebesgue.

- la différentiabilité des fonctionnelles a lieu seulement dans les directions de H, à cause du fait que les mesures translatées $\mu(. + h)$ sont équivalentes à μ si et seulement si h est un élément de H. Tout cela repose sur le fameux calcul de Malliavin.

L'objectif premier de cette thèse était de résoudre le Problème de Monge sur l'espace classique de Wiener muni de la norme uniforme. En effet les seuls résultats connus jusqu'alors sur l'espace de Wiener concernent la pseudo-norme de Cameron-Martin. On pourra citer les travaux de Feyel et Üstünel ([36],[37]), de Kolesnikov ([45], [46]) ou encore de Cavalletti ([19]). Cette question naturelle est cependant particulièrement délicate et l'objectif en soi n'a pas été atteint. Nous exposons dans ce travail des résultats qui constituent certainement des avancées allant dans ce sens. Principalement nous établirons des propriétés de convexité pour l'entropie relative sur l'espace de Wiener, traiterons le problème de Monge pour un coût provenant d'une norme suffisamment agréable $\|.\|_{k,\gamma}$, et améliorerons les résultats connus sur les équations de Monge-Ampère.

Détaillons un peu plus précisément le contenu de cette thèse. Elle se décompose en plus de l'introduction en six chapitres, dont les deux et trois sont consacrés à l'introduction des outils qui nous serons nécessaires pour mener à bien notre étude. Le premier consiste à donner le cadre de notre travail, à savoir l'espace de Wiener, en rappelant les outils essentiels, le calcul de Malliavin, les opérateurs d'Ornstein-Uhlenbeck. On insistera sur l'espace de Wiener classique, c'est-à-dire l'espace des fonctions continues sur $[0, 1]$ s'annulant en 0. Etant donné qu'il s'agit d'espaces de dimension infinie, on rappelle comment on peut les approximer par des espaces de dimension finie. On finira la partie en introduisant les fonctionnels $H-$convexes, qui admettent d'agréables propriétés. Dans le deuxième chapitre des rappels, on donnera tous les éléments de la théorie du transport optimal utilisés dans la thèse. Les problèmes de Monge-Kantorovich et de Monge sont introduits sous une forme suffisamment générale et le chapitre s'achève en un bref historique des traités sur le problème de Monge. Le fait d'introduire le problème de Monge-Kantorovich avant celui de Monge est contestable, puisque cela ne respecte pas l'ordre chronologique. Cependant pour des raisons de formalisme et de compréhension, je trouve plus simple et naturel de voir directement le problème de Monge comme un cas particulier du précédent.

Voici de quoi traitent les autres chapitres, ainsi que les principales contributions de cette thèse :

- Le Chapitre 4 concerne l'étude d'une fonctionnelle particulièrement importante sur l'espace des mesures de probabilité, à savoir l'entropie relative Ent_γ par rapport à une mesure de référence γ. On se concentrera sur ses propriétés de convexité. La distance de Wasserstein est un bon outil pour mesurer l'écart entre deux probabilités, et nous fournit un cadre métrique sur l'espace des mesures de probabilité. A partir de cela, les notions de géodésiques et de convexité le long des géodésiques prennent du sens dans ce même espace. Depuis Sturm et von Renesse dans [60], dans les variétés Riemanniennes, on sait que la convexité de Ent_γ le long des géodésiques est équivalente à une borne inférieure de la courbure de Ricci. Cette caractérisation est essentielle puisqu'elle permet de définir une notion de courbure sur les espaces métriques bien plus généraux que les variétés Riemanniennes. On obtient dans ce Chapitre des propriétés sans faire appel à des théories sophistiquées telles que la stabilité par les convergens au sens de Gromov-Hausdorff mesuré (utilisée par Lott et Villani) ou au sens de Sturm. On traitera d'abord de la dimension finie, avec toujours dans l'optique de passer en dimension infinie. Sur l'espace de Wiener, on obtient le $1-$convexité de l'entropie relative par rapport à la mesure de Wiener μ, lorsque la norme considérée est la norme uniforme. Autrement dit (Théorème 4.3.5), pour tout $t \in [0,1]$

  $$Ent_\mu(\rho_t) \leq (1-t)Ent_\mu(\rho_0) + tEnt_\mu(\rho_1) - \frac{t(1-t)}{2} W_{2,\infty}^2(\rho_0,\rho_1). \quad (1.0.1)$$

  Ce même résultat a été démontré par Fang, Shao et Sturm dans [32] lorsque la norme considérée est la pseudo-norme de Cameron-Martin. Pour des raisons techniques qui nous seront utiles dans le Chapitre 6, on modifie légèrement la distance de Wasserstein, en une quantité $\mathcal{W}_\varepsilon$ qui est le résultat d'un problème de minimisation (proche de celui de Monge-Kantorovich). Avec ce $\mathcal{W}_\varepsilon$ qui n'est plus une distance, on arrive à avoir des estimées du style (1.0.1) sur un espace de Hilbert de dimension infinie, où W_2 est remplacée par $\mathcal{W}_\varepsilon$, et la géodésique ρ_t n'est plus une géodésique mais un chemin reliant ρ_0 à ρ_1 (Proposition 4.3.3).

- Le Chapitre 5 aborde un certain nombre d'inégalités. La première partie contient simplement des rappels sur l'inégalité de Talagrand. Cette inégalité contrôle la distance entre deux mesures de probabilité au sens de Wasserstein, par l'entropie relative. La suite concerne l'établissement d'une inégalité de Harnack. Celle-ci donne une approximation du semi-groupe de la chaleur (Ornstein-Uhlenbeck) (voir l'introduction de Kassmann [44]). Sur l'espace de Wiener cette inégalité a été démontrée par Shao dans [54]. Le processus standart d'Ornstein-Uhlenbeck sur l'espace de Wiener admet pour mesure invariante la mesure de Wiener. Dans cette partie nous nous intéressons à ajouter une densité à la mesure de Wiener et à considérer le processus de Ornstein-Uhlenbeck associé. Lorsque la densité n'est pas lisse, mais au moins $H-$log concave, on montre que l'inégalité de Harnack est encore vérifiée. C'est l'objet du Corollaire 5.2.3, où pour tout $\alpha > 1$, $t \geq 0$ et $f \in Cylin(X)$,

$$|\hat{P}_t f(w)|^\alpha \leq \hat{P}_t |f|^\alpha(w') \exp\left\{\frac{\alpha d_H(w,w')^2}{2(\alpha-1)(e^{2t}-1)}\right\}, \quad \forall w, w' \in X.$$

Corollaire parce qu'il découle directement de l'estimée gradient que vérifie le semi-groupe de la chaleur associé, elle-même fortement liée à la minoration de la "courbure du Ricci" de l'espace. La courbure de Ricci n'étant correctement définie que dans les variétés Riemanniennes, on lui donne néanmoins un sens dans l'espace de Wiener, grâce au Chapitre 4. Dans la dernière partie du Chapitre, on étudie la différence entre deux applications de transport optimal sur $\mathbb{R}^n$. Le coût de transport est dans cette partie toujours la norme Euclidienne au carré. Pour obtenir des estimées on part des équations de Monge-Ampère et si les densités par rapport à la mesure Gaussienne standart sont e^{-V} et e^{-W} sous les hypothèses (5.3.32), on obtient à travers le Théorème 5.3.9 :

$$\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla W|^2 e^{-W} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||_{HS}^2 e^{-W} d\gamma$$
$$\geq 2\text{Ent}_\gamma(e^{-V}) - 2\text{Ent}_\gamma(e^{-W}) + \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2 \varphi||_{HS}^2 e^{-V} d\gamma.$$

On a donc une liaison entre la norme de Hilbert-Schmidt de la Hessienne de φ, les entropies relatives des densités, leurs informations de Fisher, ainsi que la norme de Hilbert-Schmidt de la Hessienne du terme W de la mesure cible. La grande force de cette inégalité est qu'elle ne dépend pas de la dimension. Une conséquence forte de cela sera l'obtention de solution forte de l'équation de Monge-Ampère dans le Chapitre 7.

- Le Chapitre 6 est dévoué au problème de Monge en dimension infinie. Il est découpé en deux grandes parties, la première étant consacrée aux espaces de Hilbert et la seconde aux espaces de Wiener. Tout d'abord on adapte la méthode de Champion et De Pascale, avec laquelle ils prouvent l'existence dans [21] d'une application de transport optimal pour le problème de Monge sur $\mathbb{R}^n$ pour n'importe quelle norme. Cette méthode repose fondamentalement sur le théorème de différentiation de Lebesgue, qui n'est pas toujours valable en dimension infinie (voir [53]). Toutefois Tiser donne des conditions dans [56] sur les mesures Gaussiennes sur un Hilbert, pour lesquelles ce fameux théorème est vrai. Nous nous placerons dans ce cadre, et sous les hypothèses que les deux mesures ρ_0 et ρ_1 ont leur entropie relative finie, on montrera (Théorème 6.1.2), en passant par des estimées indépendantes de la dimension, que le problème

$$\inf_{T_\# \rho_0 = \rho_1} \int_H |x - T(x)| d\rho_0(x) \tag{1.0.2}$$

a au moins une solution. Une autre méthode de Champion et De Pascale [22], permet d'obtenir des applications de transport sous des hypothèses plus faibles que celles habituellement requises, à savoir la condition (NonSmooth Twist). On se proposera d'adapter cette méthode pour les espaces de Hilbert de dimension infinie. En particulier en supposant seulement que ρ_0 ne charge pas les ensembles de codimension 1, on peut montrer que (1.0.2) admet une solution lorsque le coût est donné par $|x-y|+\varepsilon\left(1+|x-y|^2\right)^{1/2}$ ($\varepsilon > 0$). Avec ces résultats et des hypothèses convenables, on arrive à avoir une stabilité (convergence en probabilité) des applications de transports.

Concernant l'espace de Wiener, on démontre d'une manière semblable à celle de Feyel et Üstünel dans [36] l'existence et l'unicité de l'application de transport dans le cas quadratique de la pseudo-norme d_H, et sous des hypothèses plus faibles. En effet dans [36], la méthode directe est donnée lorsque la première mesure est la mesure de Wiener (sans densité). L'objet du Théorème 6.2.1 est de traité d'une manière similaire le cas où l'on ajoute une densité dont l'information de Fisher est finie. Enfin sur l'espace de Wiener classique, on traite le problème de Monge lorsque le coût est issu d'une norme de type Sobolev, $\|.\|_{k,\gamma}$ pouvant être considérée comme une moyennisation des coefficients de Hölder. Si on ajoute une puissance $p > 1$ à la norme, on prouve l'existence et l'unicité (Théorème 6.3.1) de l'application de transport directement sur l'espace de Wiener, sans passer par des approximations en dimension finie. Lorsque $p = 1$ (Théorème 6.3.4), le cas est plus délicat et il s'agit d'utiliser une méthode établie par Cavalletti. Ce dernier dans [19] prouve l'existence d'une application de transport sur l'espace de Wiener pour la pseudo-norme de Cameron-Martin. Il s'agit ici de supposer que les deux mesures ρ_0 et ρ_1 sont absolument continues par rapport à la mesure de

Wiener. De plus la stratégie repose sur une désintégration et un théorème de sélection.

- Le Chapitre 7 traite des solutions fortes de l'équation de Monge-Ampère. Les résultats obtenus utilisent de façon abondante les inégalités du Chapitre 5. Lorsque le coût est la norme euclidienne au carré, on connaît grâce à Brenier la forme de l'application de transport T lorsqu'elle existe. En effet celle-ci s'écrit comme le gradient d'une fonction convexe ϕ (unique à l'ajout d'une constante près) transportant ρ_0 sur ρ_1, ou encore étant solution de l'équation de Monge-Ampère
$$f_1(\nabla\phi)det(\nabla^2\Phi) = f_0. \tag{1.0.3}$$
Et réciproquement si Φ est une fonction convexe solution de (1.0.3), alors $\nabla\Phi$ transporte ρ_0 sur ρ_1 et en plus c'est l'unique application optimale de transport pour le coût euclidien quadratique. Cette caractérisation nous permet ainsi de tirer des informations (de la régularité principalement) sur l'application optimal de transport en étudiant l'équation de Monge-Ampère (1.0.3).

  Dans ce chapitre, on traite dans un premier temps le cas de la dimension finie. On considère deux mesures de probabilité ρ_0 et ρ_1 sur $\mathbb{R}^n$ à densité dans des espaces de Sobolev convenables. Dans le but de passer en dimension infinie, le déterminant intervenant dans (1.0.3) peut être remplacé par le déterminant de Fredholm-Carleman det_2. De plus les densités respectives e^{-V} et e^{-W} sont regardées par rapport à la mesure Gaussienne standart. Le Théorème 7.1.2 sous de faibles hypothèses sur V et W (voir (7.1.1)), nous dit que l'application de transport optimal $\nabla\Phi$ est solution de l'équation de Monge-Ampère suivante
$$e^{-V} = e^{-W(\nabla\Phi)}e^{\mathcal{L}\varphi - \frac{1}{2}|\nabla\varphi|^2}\det{}_2(\mathrm{Id} + \nabla^2\varphi), \tag{1.0.4}$$
où $\nabla\Phi = Id + \nabla\varphi$. Dans un deuxième temps, on cherche à gagner le même genre de résultat sur l'espace de Wiener. Sous des contraintes similaires sur les densités, cette fois-ci par rapport à la mesure de Wiener, on obtient une solution forte de l'équation (1.0.4). Cependant, selon comment l'approximation par la dimension finie est faite, il n'est pas immédiat de voir si cette fameuse solution est l'application de transport optimale ou non.

Chapter 2

Wiener space

The aim of this chapter is to present the background of the abstract Wiener space and to prepare materials needed in the sequel.

2.1 Abstract Wiener space

It is well-known (see e.g. [12]) that on any infinite dimensional Hilbert space H, it does not exist any Gaussian measure whose Fourier transform is given by

$$x \longmapsto \exp\left(-\frac{1}{2}|x|_H^2\right).$$

The concept of the abstract Wiener space has been introduced by Gross in [43] in order to find suitable extension of H on which such Gaussian measure exists.

By an *abstract Wiener space*, we mean the triplet (X, H, μ), where X is a separable Banach space endowed with the norm $||\cdot||$, H is a separable Hilbert space endowed with the inner product $\langle\ ,\ \rangle_H$ such that H is densely embedded in X, and μ is a Borel probability measure on X such that

$$\int_X e^{i(h,x)} d\mu(x) = \exp\big(-\frac{1}{2}|j^*(h)|_H^2\big),\ \ h \in X^\star \tag{2.1.1}$$

where $X^\star$ is the dual space of X, $(h, x) := h(x)$ and $j : H \to X$ is the embedding map, so that the dual map $j^* : X^\star \to H^\star$ defined by $\langle j^*(\ell), h\rangle_H = \ell(j(h))$ is densely defined and continuous. In what follows, we will identify H with $H^\star$, H with $j(H)$ and $X^\star$ with $j^*(X^\star)$. With these identifications, we have

$$X^\star \subset H^\star = H \subset X$$

and

$$\ell(h) = \langle \ell, h\rangle_H, \quad \ell \in X^\star, h \in H. \tag{2.1.2}$$

A basic property of the Wiener space (X, H, μ) is the following quasi-invariance of μ under action of H, due to Cameron-Martin:

$$\int_X F(x+h)\, d\mu(x) = \int_X F(x)\, K_h(x)\, d\mu(x), \quad h \in H \tag{2.1.3}$$

where K_h has the expression

$$K_h(x) = \exp\big(\langle h, x\rangle - \frac{1}{2}|h|_H^2\big), \tag{2.1.4}$$

where $\langle h, x\rangle$ is a Gaussian random variable under μ, of variance $|h|_H^2$. When $h \in X^\star$, then $\langle h, x\rangle = (h, x)$ is reduced to the duality between $X^\star$ with X. Due to (2.1.3), H is called *Cameron-Martin subspace* of X, μ is called the *Wiener measure*.

Let us summarize the features of Wiener spaces:

- H is dense in X with respect to $\|.\|$.
- $\mu(H) = 0$.
- μ is a centered and non-degenerated Gaussian measure on X.
- There is a constant $a > 0$ such that

$$\|x\| \leq a|x|_H, \quad \forall x \in X.$$

2.1.1 Projections onto finite dimensional spaces

A subset C of X is called *cylindrical set* of X if it has the form

$$C = \{x \in X,\ (l_1(x), \dots, l_N(x)) \in B\},$$

where $l_i \in X^\star$, and B is a Borelian subset of $\mathbb{R}^N$. It is known that the σ-field generated by cylindrical subsets of X is the Borel σ-field $\mathcal{B}(X)$ of X.

Let $(e_j)_{j\geq 1}$ be an orthonormal basis of H whose each e_j belongs to $X^\star$. We denote by V_n the subspace of H generated by $\{e_1, \dots, e_n\}$. Let $\pi_n : H \longrightarrow V_n$ be the orthogonal projection from H onto V_n. According to (2.1.2), π_n can be extended to the whole space X, writting

$$\begin{aligned} \pi_n : X &\longrightarrow V_n \\ x &\longmapsto \sum_{j=1}^n (e_j, x) e_j. \end{aligned}$$

For each $n \in \mathbb{N}$, we have the decomposition $x = \pi_n(x) + (x - \pi_n(x))$. Denote $Y_n = Ker(\pi_n)$. Then we can write $X = V_n \oplus Y_n$. With the induced norm, Y_n is a Banach space. Let $\gamma_n := (\pi_n)_{\#}\mu$, then by (2.1.1),

$$\int_{V_n} e^{i\langle z,x\rangle_H}\, d\gamma_n(x) = e^{-\frac{1}{2}|z|_H^2}, \quad z \in V_n.$$

In other words, γ_n is the standard Gaussian measure on V_n. Denote by $\pi_n^\perp(x) = x - \pi_n(x) : X \to Y_n$. Let $\mu_n = (\pi_n^\perp)_{\#}\mu$. Then again by (2.1.1)

$$\int_{Y_n} e^{i\langle \ell,y\rangle}\, d\mu_n(y) = e^{-\frac{1}{2}|\ell|_H^2}, \quad \ell \in V_n^\perp.$$

The triplet $(Y_n, V_n^\perp, \mu_n)$ is an abstract Wiener space. We have the following factorization of the Wiener measure:

$$\mu = \gamma_n \otimes \mu_n. \tag{2.1.5}$$

2.1.2 Sobolev spaces

Let us introduce some notations in Malliavin calculus (see [48], [29]). A function $f : X \to \mathbb{R}$ is said to be *cylindrical* if it admits the expression

$$f(x) = \hat{f}(e_1(x), \ldots, e_N(x)), \quad \hat{f} \in C_b^\infty(\mathbb{R}^N), N \geq 1 \tag{2.1.6}$$

where $\{e_1, \ldots, e_N\}$ are elements in the dual space $X^\star$ of X. We denote by Cylin(X) the space of cylindrical functions on X. For $f \in$ Cylin(X) given in (2.1.6), the gradient $\nabla f(x) \in H$ is defined by

$$\nabla f(x) = \sum_{j=1}^N \partial_j \hat{f}(e_1(x), \ldots, e_N(x))\, e_j, \tag{2.1.7}$$

where ∂_j is ith-partial derivative. Then $\nabla f : X \to H$. Let K be a separable Hilbert space; a map $F : X \to K$ is cylindrical if F admits the expression

$$F = \sum_{i=1}^m f_i\, k_i, \quad f_i \in \text{Cylin}(X), k_i \in K. \tag{2.1.8}$$

We denote by Cylin(X, K) the space of K-valued cylindrical functions. For $F \in$ Cylin(X, K), define $\nabla F = \sum_{i=1}^m \nabla f_i \otimes k_i$ which is a $H \otimes K$-valued function. For $h \in H$, we denote

$$\langle \nabla F, h\rangle = \sum_{i=1}^m \langle \nabla f_i, h\rangle_H\, k_i \in K.$$

In such a way, for any $f \in \mathrm{Cylin}(X)$ and any integer $k \geq 1$, we can define, by induction,

$$\nabla^k f : X \to \otimes^k H.$$

Let $p \geq 1$; set

$$||f||_{D_k^p} = \sum_{j=0}^{k} \left(\int_X ||\nabla^j f(x)||^p_{\otimes^j H} \, d\mu(x) \right)^{1/p}, \tag{2.1.9}$$

here we used the usual convention $\otimes^0 H = \mathbb{R}, \nabla^0 f = f$.

Definition 2.1.1. *The* Sobolev space $\mathbb{D}_k^p(X)$ *is the completion of* $\mathrm{Cylin}(X)$ *under the norm defined in* (2.1.9)*. In the same way, we define the K-valued* Sobolev space $\mathbb{D}_k^p(X; K)$.

2.1.3 Ornstein-Uhlenbeck semi-group

The Ornstein-Uhlenbeck semi-group is a powerful tool in Malliavin Calculus.

Definition 2.1.2. *For $f \in C_b(X)$, we define the Ornstein-Uhlenbeck semi-group $(P_t)_{t\geq 0}$ by*

$$(P_t f)(x) := \int_X f(e^{-t}x + \sqrt{1 - e^{-2t}}y) d\mu(y).$$

This representation of P_t is called the Mehler formula. By Mehler formula, it is easy to see that

$$P_t 1 = 1, \; P_{t+s} f = P_t P_s f, \quad \forall t, s \geq 0,$$

and

$$\int_X P_t f \, g d\mu = \int_X P_t g \, f d\mu.$$

A fundamental property is that P_t regularizes integrable functions, in the sense that

Proposition 2.1.3. *For $p > 1$:*

$$f \in L^p(X, \mu) \Rightarrow P_t f \in \mathbb{D}_k^p(X), \quad \forall k \geq 1.$$

In addition for all $f \in Cylin(X)$, the following limit

$$\lim_{t \to 0} \frac{P_t f - f}{t}$$

exists in L^p and we denote its limit by $-\mathcal{L}f$. The famous Meyer formula says that

$$||f||_{\mathbb{D}_{2k}^p} \sim ||(I + \mathcal{L})^k f||_{L^p}.$$

Definition 2.1.4. *The generator $\mathcal{L}$ of P_t is called* Ornstein-Uhlenbeck operator *on the Wiener space X.*

The *divergence* δ on the Wiener space is the dual operator of the gradient, that is for all $f \in \mathbb{D}_1^2(X)$ and $v \in Dom(\delta)$:

$$\int_X f\, \delta(v) d\mu = \int_X (\nabla f, v) d\mu.$$

It is known that

$$||\delta(v)||_{L^p} \leq c_p\, ||v||_{\mathbb{D}_1^p(X,H)}.$$

We collect a few properties in

Proposition 2.1.5. *We have*

$$\mathcal{L} = \delta \circ \nabla,$$
$$\nabla \mathcal{L} f = \mathcal{L} \nabla f + \nabla f.$$

The second formula is a special form of the Weitzenböck formula.

We consider the following Dirichlet form on $\mathbb{D}_1^2(X)$,

$$\mathcal{E}_\mu(f,f) := \int_X |\nabla f|_H^2 d\mu;$$

and thanks to the property of the divergence δ, we see that $\mathcal{E}_\mu$ is associated to the operator $\mathcal{L}$:

$$\mathcal{E}_\mu(f,f) = \int_X (\nabla f, \nabla f)_H d\mu = \int_X f \delta\left(\nabla f\right) d\mu = (\mathcal{L} f, f)_\mu.$$

Let ρ be a probability measure X, absolutely continuous w.r.t. μ, with density, say $e^{-\psi}$. We consider the corresponding Dirichlet form:

$$\mathcal{E}_\rho(f,f) = \int_X (\nabla f, \nabla f)_H\, e^{-\psi} d\mu.$$

Then we have

$$\begin{aligned}\mathcal{E}_\rho(f,f) &= \int_X (\nabla f, e^{-\psi} \nabla f)_H d\mu = \int_X f \delta\left(e^{-\psi} \nabla f\right) d\mu \\ &= \int_X f \delta\left(e^{-\psi} \nabla f\right) e^{\psi} d\rho =: (L f, f)_\rho.\end{aligned}$$

Hence the generator L of $\mathcal{E}_\rho$ admits the expression

$$L(f) = \delta(e^{-\psi}\nabla f)e^{\psi} = \mathcal{L}f + (\nabla\psi, \nabla f).$$

Now we can consider $\hat{P}_t := e^{-tL}$ the semigroup associated to the infinitesimal generator L. We call $\hat{P}_t$ a *modified Ornstein-Uhlenbeck semigroup.* It turns out that $\hat{P}_t$ has ρ as invariant measure; but instead of P_t, we have no explicit formula for $\hat{P}_t$.

For more properties on the Ornstein-Uhlenbeck semi-group, we mention [29] or [12].

2.2 Classical Wiener space

Let $X = \mathcal{C}([0,1], \mathbb{R})$ be the space of continuous functions defined on $[0,1]$. Endow X with the uniform norm $\|x\|_\infty := \sup_{t\in[0,1]} |x(t)|$. Then $(X, \|.\|_\infty)$ is a separable Banach space. We denote by

$$H := \left\{ h \in X |\ h(t) = \int_0^t \dot{h}(s)ds,\ \dot{h} \in L^2([0,1]) \right\}.$$

The space H is called Cameron-Martin space, endowed with the Hilbert norm

$$|h|_H := \|\dot{h}\|_{L^2}.$$

The Wiener measure μ on X is induced by the standard Brownian motion on $\mathbb{R}$. More precisely, for any $N \geq 1$ and $0 < t_1 < \ldots < t_N \leq 1$, the measure $\mu(C)$ of the cylindrical subset C in the form

$$C = \{x \in X;\ (x(t_1), \ldots, x(t_N)) \in B\}, \quad B \in \mathcal{B}(\mathbb{R}^N),$$

is given by

$$\mu(C) = \int_B p_{t_1}(x_1)p_{t_2-t_1}(x_2 - x_1)\cdots p_{t_N - t_{N-1}}(x_N - x_{N-1})\, dx_1 \cdots dx_N,$$

where $p_t(x)$ is the Gaussian kernel: $p_t(x) = \dfrac{e^{-x^2/2t}}{\sqrt{2\pi t}}$.

The triplet (X, H, μ) is called the *classical Wiener space.* Notice that the dual space $X^\star$ of X consists of signed Borel measures on $[0,1]$. To each $\rho \in X^\star$, we associate

$$h_\rho(t) = -\int_0^t (t-s)d\rho(s) + t\rho([0,1]).$$

Then we have

$$\langle h_\rho, h\rangle_H = \int_0^1 h(s)d\rho(s), \quad h \in H,$$

which illustrates the relation (2.1.2).

We now introduce the family of *Haar functions.* For any $n \in \mathbb{N}^\star$, k odd such that $k < 2n$, we define

$$h_{k,n}(t) := \begin{cases} \sqrt{2}^{n-1} & \text{if } t \in [(k-1)2^{-n}, k2^{-n}) \\ -\sqrt{2}^{n-1} & \text{if } t \in [k2^{-n}, (k+1)2^{-n}) \\ 0 & \text{otherwise} \end{cases}$$

Consider $H_0(t) := t$,

$$H_{k,n}(t) := \int_0^t h_{k,n}(s)ds.$$

It is known that the family

$$\{H_0, H_{k,n};\ n \geq 1,\ k \text{ odd} < 2^n\ \},$$

constitutes a complete orthonormal system of H, called the *Haar basis* of H. Let

$$V_n = \operatorname{span}\big\{H_0, H_{k,m}; k \text{ odd} < 2^m, m \leq n\big\}. \tag{2.2.1}$$

Let $\pi_n : H \to V_n$ be the orthogonal projection and π_n its extension on X. Then for $x \in X$, $\pi_n(x)$ is linear on each intervall $[\ell 2^{-n}, (\ell+1)2^{-n}]$. More precisely,

$$\pi_n(x)(t) = x(\ell 2^{-n}) + 2^n(t - \ell 2^{-n})\big(x((\ell+1)2^{-n}) - x(\ell 2^{-n})\big), \quad \text{for } t \in [\ell 2^{-n}, (\ell+1)2^{-n}].$$

The subspace V_n is of dimension 2^n and

$$||\pi_n(x)||_\infty = \max\{|x(\ell 2^{-n})|;\ \ell = 1, \ldots, 2^n\}.$$

On the space X, we can consider a few of norms, for example, the L^p-norm

$$\|x\|_p := \left(\int_0^1 |x(t)|^p dt\right)^{1/p}.$$

It is obvious that

$$\|x\|_p \leq \|x\|_\infty \leq |x|_H.$$

We will also deal with another norm, introduced by Airault and Malliavin in [2]:

$$\|x\|_{k,\gamma} := \left(\int_0^1 \int_0^1 \frac{(x(t)-x(s))^{2k}}{|t-s|^{1+2k\gamma}} dtds\right)^{1/2k},$$

where $0 < \gamma < 1/2$, and k is an integer such that $2 < 1 + 2k\gamma < k$. In fact this is a pseudo-norm over W. For this reason, we consider $\hat{X} := \{x \in X;\ \|x\|_{k,\gamma} < \infty\}$. Because μ is the law of the Brownian motion, and the Brownian motion has paths which are $\alpha-$Hölder continuous (for $\alpha < 1/2$); it turns out that $\mu(\hat{X}) = 1$. Moreover $(\hat{X}, \|.\|_{k,\gamma})$ is a separable Banach space and H is still dense in $(\hat{X}, \|.\|_{k,\gamma})$.

Let $x \in H$, then $x(t) - x(s) = \int_s^t \dot{x}(u)\, du$. It follows that

$$(x(t) - x(s))^{2k} \leq |t-s|^k |x|_H^{2k},$$

so that

$$\|x\|_{k,\gamma}^{2k} \leq C_{k,\gamma}^{2k} |x|_H^{2k},$$

where $C_{k,\gamma} := \left(\int_0^1 \int_0^1 |t-s|^{k-1-2k\gamma} dtds\right)^{1/2k}$. Therefore we obtain, combining with the previous relation:

$$\|x\|_p \leq \|x\|_\infty \leq \|x\|_{k,\gamma} \leq C_{k,\gamma}|x|_H \quad \text{for all } x \in X. \tag{2.2.2}$$

The following result will be useful in Chapter 6.

Proposition 2.2.1. *Let $\tilde{F}(x) = \|x\|_{k,\gamma}$. Then we have the following properties:*

1. *$\tilde{F}$ admits a gradient $\nabla\tilde{F}(x)$ belonging to $\hat{X}^\star$ for all $x \in \hat{X}\backslash\{0\}$, where $\hat{X}^\star$ is the dual of $\hat{X}$. Moreover $\tilde{F}^p$ is everywhere differentiable for all $p > 1$.*

2. *$\tilde{F}$ is a norm on $\hat{X}$ such that its unit ball is strictly convex.*

The first part of the proof is inspired from [29].

Proof. 1. First we show the property for $F := \tilde{F}^{2k}$. Take $h \in \hat{X}$, we can write for $x \in \hat{X}$ and $\varepsilon > 0$:

$$F(x + \varepsilon h) = \int_0^1 \int_0^1 \frac{((x(t) - x(s)) + \varepsilon(h(t) - h(s)))^{2k}}{|t-s|^{1+2k\gamma}} dtds.$$

Taking the derivative at $\varepsilon = 0$, we have

$$D_h F(x) = 2k \int_0^1 \int_0^1 \frac{(x(t) - x(s))^{2k-1}(h(t) - h(s))}{|t-s|^{1+2k\gamma}}\, dtds.$$

Therefore

$$\begin{aligned} |D_h F(x)| &\leq 2k \int_0^1 \int_0^1 \frac{|x(t) - x(s)|^{2k-1}}{|t-s|^{1+2k\gamma}} |h(t) - h(s)| dtds \\ &\leq 2k \int_{[0,1]^2} \frac{|x(t) - x(s)|^{2k-1}}{|t-s|^{(1+2k\gamma)(2k-1)/(2k)}} \frac{|h(t) - h(s)|}{|t-s|^{(1+2k\gamma)/(2k)}} dtds. \end{aligned}$$

Using Hölder's inequality, we get

$$\begin{aligned}|D_hF(x)| &\leq 2k\left(\int_{[0,1]^2}\frac{|x(t)-x(s)|^{2k}}{|t-s|^{1+2k\gamma}}dtds\right)^{(2k-1)/(2k)}\left(\int_{[0,1]^2}\frac{|h(t)-h(s)|^{2k}}{|t-s|^{1+2k\gamma}}dtds\right)^{1/(2k)}\\ &= 2k\|x\|_{k,\gamma}^{2k-1}.\|h\|_{k,\gamma}.\end{aligned}$$

Hence $h\longmapsto D_hF(x)$ is a bounded operator on $\hat{X}$ for all $x\in\hat{X}$. It leads to the existence of a gradient $\nabla F(x)$ which belongs to the dual space $\hat{X}^\star\subset H^\star=H$ (by (2.2.2)). Since $\tilde{F}=F^{1/(2k)}$, its gradient satisfies $\nabla\tilde{F}(x)=F^{1/(2k)-1}(x)\nabla F(x)$ for $x\neq 0$.
$\tilde{F}$ is differentiable out of $\{0\}$, but for any $p>1$, $\tilde{F}^p$ is differentiable at 0, hence everywhere over $(\hat{X},\|.\|_{k,\gamma})$.
2. The proof for the item 2 is the same as the proof for Minkowski's inequality. Indeed for $x_1,\ x_2\in X$ and $\eta\in(0,1)$, we have:

$$\begin{aligned}&\|(1-\eta)x_1+\eta x_2\|_{k,\gamma}^{2k} = \int_{[0,1]^2}\frac{|(1-\eta)(x_1(t)-x_1(s))+\eta(x_2(t)-x_2(s))|^{2k}}{|t-s|^{1+2k\gamma}}dtds\\ &= \int_{[0,1]^2}|(1-\eta)(x_1(t)-x_1(s))+\eta(x_2(t)-x_2(s))|\\ &\times\frac{|(1-\eta)(x_1(t)-x_1(s))+\eta(x_2(t)-x_2(s))|^{2k-1}}{|t-s|^{1+2k\gamma}}dtds\\ &\leq \int_{[0,1]^2}\frac{(1-\eta)|x_1(t)-x_1(s)|}{|t-s|^{(1+2k\gamma)/(2k)}}\frac{|(1-\eta)(x_1(t)-x_1(s))+\eta(x_2(t)-x_2(s))|^{2k-1}}{|t-s|^{(1+2k\gamma-\frac{1}{2k}-\gamma)}}dtds\\ &+\int_{[0,1]^2}\frac{\eta|x_2(t)-x_2(s)|}{|t-s|^{(1+2k\gamma)/(2k)}}\frac{|(1-\eta)(x_1(t)-x_1(s))+\eta(x_2(t)-x_2(s))|^{2k-1}}{|t-s|^{(1+2k\gamma-\frac{1}{2k}-\gamma)}}dtds\\ &\leq\left((1-\eta)\|x_1\|_{k,\gamma}+\eta\|x_2\|_{k,\gamma}\right)\left(\|(1-\eta)x_1+\eta x_2\|_{k,\gamma}^{2k}\right)^{1-1/2k}.\end{aligned}$$

The two inequalities above come from the triangle inequality and Hölder's inequality. They are equality if and only if x_1 and x_2 are almost everywhere colinear. This leads to the strict convexity of our norm. □

At the end of this section, we show the limit behavior of the sequence $(\|.\|_{k,\gamma})_k$ for $0<\gamma<1/2$. For this, we introduce

$$\|x\|_{\infty,\gamma}:=\sup_{t,s\in[0,1]}\frac{|x(t)-x(s)|}{|t-s|^\gamma}.$$

That is a stronger norm than the uniform one $\|.\|_\infty$.

Lemma 2.2.2. *Let $K\subset\hat{X}$ be a compact subset of X. Then for any $0<\gamma<1/2$,*

$$\lim_{k\to\infty} \sup_{x\in K} |\|x\|_{k,\gamma} - \|x\|_{\infty,\gamma}| = 0.$$

Proof.
First we have:

$$\|x\|_{k,\gamma} = \left(\int_0^1 \int_0^1 \frac{|x(t)-x(s)|^{2k}}{|t-s|^{1+2k\gamma}} dtds\right)^{1/(2k)} \leq \sup_{t,s\in[0,1]} \frac{|x(t)-x(s)|}{|t-s|^{\frac{1}{2k}+\gamma}}.$$

Taking the limit when k goes to infinity we get:

$$\limsup_k \|x\|_{k,\gamma} \leq \sup_{t,s\in[0,1]} \frac{|x(t)-x(s)|}{|t-s|^{\gamma}} = \|x\|_{\infty,\gamma}. \tag{2.2.3}$$

Up to consider $\frac{x}{\|x\|_{\infty,\gamma}}$ we can assume $\|x\|_{\infty,\gamma} = 1$. So for $\varepsilon \in (0,1)$,

$$\begin{aligned}\|x\|_{k,\gamma}^{2k} &\geq \int_{\left\{\frac{|x(t)-x(s)|}{|t-s|^{\gamma}} > 1-\varepsilon\right\}} \frac{|x(t)-x(s)|^{2k}}{|t-s|^{1+2k\gamma}} dtds \\ &\geq (1-\varepsilon)^{2k} \int_{\left\{\frac{|x(t)-x(s)|}{|t-s|^{\gamma}} > 1-\varepsilon\right\}} \frac{1}{|t-s|} dtds.\end{aligned}$$

Because $1/|t-s| \geq 1$ for all $t,s \in [0,1]$ and because $\|x\|_{\infty,\gamma} = 1$, the set $\left\{\frac{|x(t)-x(s)|}{|t-s|^{\gamma}} > 1-\varepsilon\right\}$ has non zero Lebesgue measure. Thus

$$\|x\|_{k,\gamma} \geq (1-\varepsilon)\mathcal{L}\left(\left\{\frac{|x(t)-x(s)|}{|t-s|^{\gamma}} > 1-\varepsilon\right\}\right)^{1/(2k)},$$

where the last term tends to $(1-\varepsilon)$ when k goes to infinity. Finally because it is true for all $\varepsilon \in (0,1)$:

$$\liminf_k \|x\|_{k,\gamma} \geq 1. \tag{2.2.4}$$

Combining (2.2.3) and (2.2.4) we get the result. The uniform convergence over any compact subsets of X can be seen easily.

Note that level sets $\{x \in X;\ ||x||_{k,\gamma} \leq R\}$ are compact in X. □

2.3 $H-$convex functions on Wiener spaces

Convex functions play an important role in the theory of optimal transportation. H- convex functions on the Wiener space have been introduced by Feyel and

Üstünel. In this subsection, we will collect some results in [35] for later use. But first of all, we consider a regular case.

Let $W \in \mathbb{D}_2^2(X)$ such that e^{-W} is bounded and $\int_X e^{-W} d\mu = 1$. It is well-known that the following condition

$$\langle \nabla^2 W, h \otimes h \rangle_{H \otimes H} \geq -c\, |h|_H^2, \text{ for some } c \in [0, 1[, \tag{2.3.1}$$

implies (see [24, 35]) the *logarithmic Sobolev inequality*

$$(1-c) \int_X \frac{|f|}{||f||_{L^2(e^{-W}\mu)}}\, e^{-W} d\mu \leq \int_X |\nabla f|^2\, e^{-W} d\mu, \quad f \in \text{Cylin}(X). \tag{2.3.2}$$

It is also known (see for example [61]) that (2.3.2) is stronger than the *Poincaré inequality*

$$(1-c) \int_X (f - \mathbb{E}_W(f))^2\, e^{-W} d\mu \leq \int_X |\nabla f|^2\, e^{-W} d\mu, \tag{2.3.3}$$

where $\mathbb{E}_W$ denotes the integral with respect to the measure $e^{-W}\mu$.

In order to generalize the above inequalities to a larger class of measures, Feyel and Üstünel introduced in [35] the notion of $H-$convex functions on Wiener space.

A measurable functional $F : X \longrightarrow \mathbb{R}$ is said to be $H-$*convex* if for all $h, k \in H$, and $\alpha \in [0, 1]$,

$$F(x + \alpha h + (1-\alpha)k) \leq \alpha F(x+h) + (1-\alpha) F(x+k),$$

almost surely. For $a \in \mathbb{R}$, F is said to be $a-$*convex* if the map

$$h \to \frac{a}{2}|h|_H^2 + F(x+h)$$

is a convex map from H to $L^0(X, \mu)$ the space of measurable functions on X, that is,

$$F(x + \alpha h + (1-\alpha)k) \leq \alpha F(x+h) + (1-\alpha)F(x+k) + \alpha(1-\alpha)\, \frac{a}{2}|h-k|_H^2.$$

Let P_t be the Ornstein-Uhlenbeck semigroup. If F satisfies the above inequality, then

$$\begin{aligned} F\big(e^{-t}(x + \alpha h + (1-\alpha)k) + \sqrt{1-e^{-2t}}y\big) &\leq \alpha F(e^{-t}(x+h) + \sqrt{1-e^{-2t}}y) \\ &+ (1-\alpha)F(e^{-t}(x+k) + \sqrt{1-e^{-2t}}y) + \alpha(1-\alpha)\frac{ae^{-2t}}{2}|h-k|_H^2. \end{aligned}$$

Integrating with respect to y, we see that $P_t F$ is a $e^{-2t}a-$convex function. A characterization of a- convex functions is the following

Proposition 2.3.1. *Let $F \in L^p(\mu)$ for some $p > 1$. Then F is $a-$convex if and only if*

$$\int_X F\left(\nabla^2\varphi(x), h \otimes h\right)_{H\otimes H} d\mu(x) \geq -a|h|_H^2,$$

for any $h \in H$ and nonnegative $\varphi \in \mathbb{D}_2^\infty(X)$.

In parallel, a functional $G : X \longrightarrow \mathbb{R}$ is said to be *a-log concave* if there is a $a-$convex function F such that

$$G = e^{-F}.$$

Feyel and Üstünel gave nice properties concerning such functionals. The following result is taken from Proposition 5.1 in [35].

Proposition 2.3.2. *If $G : X \longrightarrow \mathbb{R}$ is a-log concave function, then*

- $\mathbb{E}^{V_n}(G)$ *is again a-log concave for any $n \geq 1$,*
- $P_t G$ *is again a-log concave for any $t \geq 0$.*

where $\mathbb{E}^{V_n}(G)$ denotes the conditional expectation with respect to the sub σ-field of X generated by $\pi_n = X \to V_n$, and P_t is the Ornstein-Uhlenbeck semi-group.

The following result was also proved in [35].

Proposition 2.3.3. *Let W be a $H-$convex function such that $\int_X e^{-W} d\mu = 1$. Then*

$$\int_X f^2 \left(\log f^2 - \log \|f\|^2_{L^2(e^{-W}\mu)}\right) e^{-W} d\mu \leq 2 \int_X |\nabla f|^2 e^{-W} d\mu.$$

Chapter 3

Basic tools of optimal transportation

There are a lot of monographs on the theory of optimal transportation. We refer to [5] and [58] for a broad treatement. Here we only gather some materials for later use.

3.1 Some general facts about measure theory

Let (X, d) be a *Polish space*, that is a separable complete metric space. We denote by $\mathcal{P}(X)$ the set of Borel probability measures on X. A basic fact on a Polish space is that any $\mu \in \mathcal{P}(X)$ is *tight*, that is, for any $\varepsilon > 0$, there is a compact subset K of X such that $\mu(K^c) < \varepsilon$.

Definition 3.1.1. *We say that a family Λ of probability measures on X is* tight *if for any $\varepsilon > 0$ there is a compact subset $K_\varepsilon \subset X$ such that*

$$\mu(X \backslash K_\varepsilon) \leq \varepsilon, \qquad \forall \mu \in \Lambda.$$

Prokhorov's theorem. A family $\Lambda \subset \mathcal{P}(X)$ is relatively compact for the weak topology if and only if it is tight.

Definition 3.1.2. *Let $\mu \in \mathcal{P}(X)$; we say that μ is concentrated on a Borel subset A of X if $\mu(A) = 1$.*
The support *$Supp(\mu)$ of the measure μ is the smallest closed set of X on which μ is concentrated; in other words, $X \backslash Supp(\mu)$ is $\mu-$negligible.*

An abstract Wiener space (X, H, μ) is a typical infinite dimensional example of Polish spaces. We have $Supp(\mu) = X$.

3.2 Monge-Kantorovich Problem

Let (X, d) and $(Y, \tilde{d})$ be two Polish spaces endowed with their Borel $\sigma-$algebra. Given two Borel probability measures ρ_0, ρ_1 on X and Y respectively, we say that a probability measure Π on the product space $X \times Y$ is a *coupling* of ρ_0 and ρ_1, if $(P_1)_\#\Pi = \rho_0, (P_2)_\#\Pi = \rho_1$ where $P_1 : X \times Y \to X$ is the first projection, while P_2 is the second projection. We denote by $C(\rho_0, \rho_1)$ the collection of couplings of ρ_0 and ρ_1.

Let

$$c : X \times Y \longrightarrow [0, \infty]$$

be a measurable function, which will be called *cost function.* The *Monge-Kantorovich Problem* consists of minimizing the total cost of *transportation* between ρ_0 and ρ_1 in the following sense:

$$\inf_{\Pi \in C(\rho_0,\rho_1)} \int_{X\times Y} c(x, y)d\Pi(x, y) := W_c(\rho_0, \rho_1), \tag{MKP}$$

Here are a few obvious remarks:

- $C(\rho_0, \rho_1)$ is never empty, since $\rho_0 \otimes \rho_1 \in C(\rho_0, \rho_1)$.
- $C(\rho_0, \rho_1)$ is convex.
- $C(\rho_0, \rho_1)$ is tight.
- If c is lower semi-continuous then the functional

  $$F(\Pi) = \int_{X\times Y} c(x, y)d\Pi(x, y)$$

  is also lower semi-continuous with respect to the weak topology on $C(\rho_0, \rho_1)$. By Prokhorov's theorem, F attains its minimum over $C(\rho_0, \rho_1)$.

The last point in the previous remark says that the infimum in **(MKP)** can be replaced by the minimum provided the cost function is lower semi-continuous.

3.2.1 Characterization of optimal couplings

In what follows, we always assume that the cost function is lower semi-continuous.

Definition 3.2.1. *A coupling $\Pi_0 \in C(\rho_0, \rho_1)$ is said to be optimal, relative to the cost c, if it realizes the minimum in **(MKP)**:*

$$\int_{X\times Y} c(x, y)d\Pi_0(x, y) = \min_{\Pi\in C(\rho_0,\rho_1)} \int_{X\times Y} c(x, y)d\Pi(x, y).$$

We denote by $C_0(\rho_0, \rho_1)$ the (non empty) set of *optimal couplings* between ρ_0 and ρ_1. Again it is easy to see that $C_0(\rho_0, \rho_1)$ is a *convex* subset of $C(\rho_0, \rho_1)$.

The following notion of cyclical monotonicity plays an important role in the characterization of the optimality of couplings.

Definition 3.2.2. *A subset $\Gamma \subset X \times Y$ is said to be $c-$*cyclically monotone *if for any finite number of couples of points $(x_1, y_1), \ldots, (x_N, y_N) \in \Gamma$, it holds that*

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1}),$$

with the convention $y_{N+1} = y_1$.
*We say that a coupling $\Pi \in C(\rho_0, \rho_1)$ is $c-$*cyclically monotone *if its support $Supp(\Pi)$ is $c-$cyclically monotone.*

Here is the useful characterization to be optimal for a coupling.

Proposition 3.2.3. *Let $c : X \times Y \longrightarrow [0, \infty]$ be a cost function.*

- *If c is lower semi-continuous, then any optimal coupling is $c-$cyclically monotone.*
- *If moreover c is real-valued and continuous, then a coupling $\Pi \in C(\rho_0, \rho_1)$ is optimal if and only if it is c-cyclically monotone.*

Proof. We refer to [58] Theorem 5.10. □

Now we only consider the case $(X, d) = (Y, \tilde{d})$ and we assume that $x \to d(x, x_0)$ is in $L^1(\rho_0) \cap L^1(\rho_1)$.

Another important tool in optimal transportation is the Kantorovich duality formula. First, we introduce the notion of $c-$convex function. Let $\varphi : X \longrightarrow \mathbb{R}$ be a measurable function. We say that φ is $c-$*convex* if

$$\varphi(x) = \sup_{y \in X} \left(\varphi^c(y) - c(x, y)\right) \qquad \forall x \in X,$$

where φ^c, called $c-$*transform* of φ, is defined by:

$$\varphi^c(y) = \inf_{x \in X} \left(\varphi(x) + c(x, y)\right) \qquad \forall y \in X.$$

Proposition 3.2.4. *Let $c : X \times X \longrightarrow [0, \infty)$ be a cost function such that $W_c(\rho_0, \rho_1) < +\infty$. Assume that $c(x, y) \leq \alpha(x) + \beta(y)$ with $\alpha \in L^1(\rho_0)$ and $\beta \in L^1(\rho_1)$, then we have the equivalence between the two points:*

- Π *is optimal in* ***(MKP)*** *(for c)*
- *there exist a* $c-$*convex* $\varphi \in L^1(\rho_0)$ *and a Borel subset* $\Gamma \subset X \times X$ *such that* $\Pi(\Gamma) = 1$ *and*

$$\begin{cases} \varphi^c(y) - \varphi(x) = c(x,y), & \forall (x,y) \in \Gamma \\ \varphi^c(y) - \varphi(x) \leq c(x,y), & \forall (x,y) \in X \times X. \end{cases}$$

Proof. We refer to [58] Theorem 5.10. □

The original Monge problem concerns the cost induced by a distance $c(x,y) = d(x,y)$. In this case we have a better proposition than above:

Proposition 3.2.5. *Let* $c : X \times X \longrightarrow [0, \infty)$ *a cost function induced by the distance on* X *i.e.* $c(x,y) = d(x,y)$. *Let* ρ_0, ρ_1 *be two probability measures on* X *such that* $x \to d(x, x_0)$ *is integrable with respect to* ρ_0 *and to* ρ_1. *If* Π *is optimal for the Monge-Kantorovich problem between* ρ_0 *and* ρ_1 *with respect to the cost* c, *then we can find a* $1-$*Lipschitz map* $u : X \longrightarrow \mathbb{R}$ *such that:*

$$\begin{cases} u(x) - u(y) = c(x,y), & \forall (x,y) \in Supp(\Pi) \\ u(x) - u(y) \leq c(x,y), & otherwise. \end{cases} \tag{3.2.1}$$

In particular, under conditions in Proposition 3.2.5, the *Kantorovich-Rubinstein formula*:

$$\min_{\Pi \in C(\rho_0,\rho_1)} \int_{X \times X} d(x,y) d\Pi(x,y) = \max_{u \in Lip(X)} \left\{ \int_X u d\rho_0 - \int_X u d\rho_1 \right\}$$

holds.

3.2.2 Stability

Lemma 3.2.6. *Let* $(\mu_k)_k$ *be a sequence of probability measures on* X, *which converges weakly to a measure* μ. *Then for any* $x \in Supp(\mu)$, *there exists a sequence of points* x_k *such that* $x_k \in Supp(\mu_k)$ *and* $\lim_{k \to +\infty}(x_k) = x$.

Proof. Let $x \in Supp(\mu) \subset X$. Thus for any $p \in \mathbb{N}^\star$, we have $\mu(B(x, 1/p)) > 0$. By weak convergence and the fact that $B(x, 1/p)$ is open, we have:

$$\liminf_{k \longrightarrow +\infty} \mu_k(B(x, 1/p)) \geq \mu(B(x, 1/p)) > 0.$$

This inequality allows us to define an increasing sequence $(j_p)_p$ such that: $j_0 := 0$ and for $p > 0$

$$j_p := \min\{q \in \mathbb{N},\ q > j_{p-1},\ \forall n \geq q\ : Supp(\mu_n) \cap B(x, 1/p) \neq \varnothing\}.$$

For all $q \geq 1$, there exists $p \in \mathbb{N}$ such that $j_p \leq q < j_{p+1}$, so that we can pick up a point $x_q \in Supp(\mu_q) \cap B(x, 1/p)$. The sequence $(x_q)_q$ converges to x. □

The following proposition claims in particular that for a convergent sequence of cost functions, any sequence of corresponding optimal couplings converges as well, to a coupling optimal for the limit cost function.

Proposition 3.2.7. *Let $c_k, c : X \times X \longrightarrow [0, \infty)$ be continuous costs such that $(c_k)_k$ converges uniformly on compact subsets to c. If $\Pi_k \in C_0(\mu_k, \nu_k)$ (such as the total cost w.r.t. c_k is finite) whith $(\mu_k)_k, (\nu_k)_k \subset \mathcal{P}(X)$ which converge weakly respectively to μ and ν ; then up to a subsequence, $(\Pi_k)_k$ converges weakly to some coupling $\Pi \in C(\mu, \nu)$. In addition if*

$$\int c d\Pi < \infty$$

then Π is optimal.

Proof. Since $(\mu_k)_k$ and $(\nu_k)_k$ are convergent sequences, they are tight sets. It turns out that $(\Pi_k)_k$ is tight; therefore up to a subsequence, Π_k converges weakly to some $\Pi \in C(\mu, \nu)$.
By Proposition 3.2.3, it is sufficient to prove that $Supp(\Pi)$ is $c-$cyclically monotone. Let $N \in \mathbb{N}^\star$ and $(x_1, y_1), \ldots, (x_N, y_N) \in Supp(\Pi)$. Since $(\Pi_k)_k$ converges weakly to Π, we can apply Lemma 3.2.6: for all $i = 1, \ldots N$, there exists $(x_i^k, y_i^k) \in Supp(\Pi_k)$ such that $\lim_{k \to +\infty}(x_i^k, y_i^k) = (x_i, y_i)$. Thus $(x_1^k, y_1^k), \ldots, (x_N^k, y_N^k) \in Supp(\Pi_k)$ which is c_k-cyclically monotone, because Π_k is optimal for the cost c_k. Then the inequality

$$\sum_{i=1}^{N} c_k(x_i^k, y_i^k) \leq \sum_{i=1}^{N} c_k(x_i^k, y_{i+1}^k) \tag{3.2.2}$$

holds, with $y_{N+1} := y_1$. And it is elementary to check that the sets

$$\cup_{k \geq 1} \{(x_1^k, y_1^k), \ldots, (x_N^k, y_N^k)\} \bigcup \{(x_1, y_1), \ldots, (x_N, y_N)\},$$
$$\cup_{k \geq 1} \{(x_1^k, y_2^k), \ldots, (x_N^k, y_1^k)\} \bigcup \{(x_1, y_2), \ldots, (x_N, y_1)\},$$

are compact of $\mathbb{R}^n \times \mathbb{R}^n$. But since $(c_k)_k$ converges uniformly on compact subsets of $X \times X$ to c, we get from (3.2.2), taking the limit with $k \to +\infty$:

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1}).$$

That is exactly the definition of $c-$cyclically monotone for $Supp(\Pi)$. The result follows from Proposition 3.2.3 . □

3.3 Wasserstein distances

Let X be a Polish space and

$$d : X \times X \longrightarrow [0, \infty],$$

be a distance or a pseudo-distance on X. For example, on the Wiener space (X, H, μ), the d_H distance defined by

$$d_H(x, y) = \begin{cases} |x - y|_H & \text{if} \quad x - y \in H; \\ +\infty & \text{otherwise.} \end{cases}$$

is a pseudo-distance, which is lower semi-continuous.

We will introduce the Wasserstein distance on $\mathcal{P}(X)$. Let ρ_0 and $\rho_1 \in \mathcal{P}(X)$ be two probability measures.

Definition 3.3.1. *We define the L^p- Wasserstein distance between ρ_0 and ρ_1 as:*

$$W_{p,d}(\rho_0, \rho_1) := \left(\inf_{\Pi \in C(\rho_0, \rho_1)} \int_{X \times X} d(x, y)^p d\Pi(x, y) \right)^{1/p}.$$

Note that $W_{p,d}$ could take the value infinity.

- Notice that if d is a true distance, and $\Pi \in C(\rho_0, \rho_1)$, we have:

  $$\int_{X \times X} d(x, y)^p d\Pi(x, y) \leq 2^{p-1} \int_X d(x, x_0)^p d\rho_0(x) + \int_X d(x_0, y)^p d\rho_1(y).$$

  It follows that $W_{p,d}$ is finite provided ρ_0 and ρ_1 have finite moment of order p. We denote by

  $$\mathcal{P}_p(X) := \{\rho \in \mathcal{P}(X),\ m_p(\rho) < \infty\},$$

  where $m_p(\rho) := \int_X d(x, x_0)^p d\rho(x)$ for some fixed $x_0 \in X$.

- For d_H on the Wiener space, the notion of moment is not suitable since $d_H(x, x_0) = +\infty$ for μ-almost everywhere. However, in this case, the Talagrand inequality

  $$W_{2,d_H}(\mu, \rho)^2 \leq 2\mathrm{Ent}_\mu(\rho),$$

  holds where $\mathrm{Ent}_\mu(\rho) = \int_X f \log f \, d\mu$ if $\rho = f\mu$, otherwise to be $+\infty$. So $W_{2,d_H}(\rho_0, \rho_1)$ is finite if ρ_0 and ρ_1 have finite entropy. We denote

  $$D(Ent_m) = \{\rho \in \mathcal{P}(X);\ Ent_m(\rho) < +\infty\}.$$

In what follows, we will use the notation $\mathcal{P}(X)[p]$ for $\mathcal{P}_p(X)$ if m admits the moment of order p. In the case where the moment of order 2 of m is infinite, but the Talagrand inequality holds for m, de denote $\mathcal{P}(X)[2] = D(Ent_m)$.

The following proposition justify the term of *distance* for W_p.

Proposition 3.3.2. *$W_{p,d}$ is a distance over $\mathcal{P}(X)[p]$.*

Here are some Wasserstein distances that we will deal with:

Space (X, d)	Wasserstein distance	$\mathcal{P}(X)[p]$
$(\mathbb{R}^n, \|.\|_q)$	$W_{p,q}$	$\mathcal{P}_p(\mathbb{R}^n)$
(X, H, d_H)	W_2	$D(Ent_\mu)$
$(X, H, \|.\|_\infty)$	$W_{p,\infty}$, $1 \leq p \leq 2$	$D(Ent_\mu)$
$(X, H, \|.\|_{k,\gamma})$	$W_{p,(k,\gamma)}$, $1 \leq p \leq 2$	$\mathcal{P}_p(X)$

3.4 The Monge Problem

3.4.1 Optimal transportation theory

Let X be a Polish space endowed with the Borel $\sigma-$algebra, and ρ_0, ρ_1 be two Borel probability measures on X.

The *Monge Problem* with respect to the cost c consists of finding a measurable map $T : X \to X$, which minimizes the quantity

$$\int_X c(x, T(x)) d\rho_0(x), \qquad \text{(MP)}$$

where the constraint is taken such that $T_\# \rho_0 = \rho_1$, that is, $\rho_0(T^{-1}(A)) = \rho_1(A)$ for all Borel subsets A of X. We say that T *pushes* ρ_0 *forward to* ρ_1. Originally Monge himself stated in 1781 the problem for the Euclidian norm in $\mathbb{R}^3$.
This constraint is **fully non linear**. Indeed on the Eulidean space $\mathbb{R}^n$, when both measures ρ_0 and ρ_1 are absolutely continuous with respect to the Lebesgue measure m, solving $T_\# \rho_0 = \rho_1$ is equivalent (at least formally) to solve the partial derivative equation

$$f_0 = f_1(T)\, |det(\nabla T)|.$$

In Chapter 7, we will study the above *Monge-Ampère equation*.

So the Monge Problem is difficult to solve. The Monge-Kantorovich Problem **(MKP)** gives a relaxed version of it. In fact, if a Borel map T solves the Monge problem, then the coupling between ρ_0 and ρ_1 defined by $(id \times T)_\# \rho_0$ is a solution to the Monge-Kantorovich problem. From the Monge-Kantorovich problem to the Monge problem, we have to prove that the optimal coupling is indeed supported by the graph of a measurable map T which pushes ρ_0 forward to ρ_1.

Definition 3.4.1. *A measurable map $T : X \longrightarrow X$ minimizing the quantity in **(MP)** will be called an* optimal transport map*.*

It makes sense to search a *Monge solution* whenever (MKP) (or the Wasserstein distance $W_c(\rho_0, \rho_1)$) is finite.

In what follows, we will give a brief review of results concerning the Monge problem. Perhaps the most famous one has been obtained by Brenier in [14], where he solved the Monge Problem when the cost is induced by the square of the Euclidian norm in $\mathbb{R}^n$. Besides he proved that the *optimal transport map* is given by the gradient of convex functions and gave a link with Monge-Ampère equations. We omit the second indice in the Wasserstein distance when it is induced by the Euclidian norm. Here is his result.

Theorem. (Brenier) *Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^n)$ having moment of order 2. Assume that ρ_0 is absolutely continuous with respect to the Lebesgue measure of $\mathbb{R}^n$. Then there is a convex function $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that $T := \nabla\Phi$ is an optimal transport map from ρ_0 to ρ_1. In addition $(I \times T)_{\#}\rho_0$ is the unique optimal plan in (MKP) and T is the unique optimal transport map .*

Later R. McCann [51] solved Monge problem on compact Riemmanian manifolds when the cost is given by the square of the Riemmanian distance, and the first measure is absolutely continuous with respect to the volume measure. The optimal transport map T again admits an explicit expression using the geodesic exponential map

$$T(x) = \exp_x(\nabla\varphi(x)).$$

In case of compact Lie groups, an alternative proof of R. McCann's result has been given by Fang and Shao [31].

The assumption on the absolute continuity of the first measure ρ_0 is weakened, first by McCann in [49] where he proved that it is enough that ρ_0 does not charge any subset of Hausdorff dimension less than $n-1$. Recently Gigli [41] gave a sharp condition on the first measure.

A straighforward generalization of the square of Euclidean norm is a cost $c : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$, which is a differentiable function satisfying the twist condition:

(Twist) $\qquad \forall x \in \mathbb{R}^n, \quad y \longmapsto \nabla_x c(x, y)$ is injective.

A more precise statement is (see Villani's book [58]):

Theorem 3.4.2. *Let $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^n)$ such that $\rho_0 << \mathcal{L}$ and*

$$W_{2,c}(\rho_0, \rho_1) < \infty.$$

*If the cost function c satisfies the above twist condition **(Twist)** and that* $\nabla_x c(x, y)$ *is bounded locally in* x*, uniformly in* $y \in \mathbb{R}^n$*. Then there is a locally Lipschitz function* $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$*, such that* $T(x) := (\nabla_x c(x, .))^{-1}(-\nabla\phi(x))$ *is the unique (up to a* ρ_0-*negligible set) optimal map from* ρ_0 *to* ρ_1*. In addition* $(I \times T)_{\#}\rho_0$ *is the unique optimal plan in (MKP) .*

Remark 3.4.3. *A typical example of above twist costs is*

$$c(x, y) = |x - y|^p, \qquad \forall p > 1.$$

The regularity of optimal transport maps is of great interest. We finish the section talking about *approximate differentiability*. This notion plays a great role to get properties concerning optimal maps. Recall that in $\mathbb{R}^n$, we call *density* of a measurable subset $\Omega \subset \mathbb{R}^n$ at a point $x \in \Omega$, the quantity

$$\lim_{r \to 0} \frac{\mathcal{L}(B(x, r) \cap \Omega)}{\mathcal{L}(B(x, r))},$$

which equals 1 $\mathcal{L}$-almost surely (thanks to the Lebesgue differentiation theorem).

Proposition 3.4.4. *Let* $\rho_0,\ \rho_1 \in \mathcal{P}(\mathbb{R}^n)$ *be two probability measures, absolutely continuous w.r.t. the Lebesgue measure* $\mathcal{L}$*. Assume that the cost* c *is given by* $c(x, y) = h(x - y)$ *where the function* $h : \mathbb{R}^n \to [0, +\infty[$ *is strictly convex with superlinear growth and satisfies*

- $h \in C^1(\mathbb{R}^n) \cap C^2(\mathbb{R}^n \backslash \{0\})$
- $\nabla^2 h$ *is positive definite in* $\mathbb{R}^n \backslash \{0\}$.

Then the optimal map T *between* ρ_0 *and* ρ_1 *is approximately differentiable at* ρ_0*-almost everywhere point* x*. In other words, there exists a differentiable function* $\tilde{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *such that for* ρ_0-*a.e.* $x \in \mathbb{R}^n$*, the set* $\{T = \tilde{T}\}$ *has density* 1 *at* x*, that is,*

$$\lim_{r \to 0} \frac{\mathcal{L}(B(x, r) \cap \{T = \tilde{T}\})}{\mathcal{L}(B(x, r))} = 1.$$

In addition $\nabla\tilde{T}$ *is diagonalizable with nonnegative eigenvalues.*

Proof.
See Theorem 6.2.7. in [6]. □

The approximatively differentiable functions also enjoy the formula of change of variable. More precisely

Proposition 3.4.5. *Let $\rho \in \mathcal{P}(\mathbb{R}^n)$ be absolutely continuous w.r.t. to $\mathcal{L}$ with density f. For $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ approximately differentiable on Ω, such that $\tilde{T}_{|\Omega}$ is injective and $\mathcal{L}(\{f > 0\}\backslash\Omega) = 0$, we have:*

$$T_{\#}\rho << \mathcal{L} \Leftrightarrow det(\tilde{\nabla}T) > 0 \quad \mathcal{L} - \text{a.s.}$$

In this case the density can be written as

$$T_{\#}\rho = \frac{f}{|det(\tilde{\nabla}T)|} \circ \tilde{T}^{-1}_{|T(\Omega)}\mathcal{L}. \tag{3.4.1}$$

Proof. See for instance Lemma 5.5.3 in [6]. □

3.4.2 Historical background

The Monge Problem **(MP)** has been introduced by Monge in 1781 ([52]). The relaxed Monge-Kantorovich Problem **(MKP)** has been introduced by Kantorovich in 1948. From these two problems the theory of optimal transportation has been largely invested.

Below I put a (non exhaustive) list of contributions in solving Monge problems during the last decades, in order to illustrate the art of the stage. We will denote by $|.|$ for the Euclidian norm (or Hilbert norm), $\|.\|$ for some general norm on $\mathbb{R}^n$, $\mathcal{L}$ for the Lebesgue measure (respectively for the volume measure) on $\mathbb{R}^n$ (respectively on a Riemannian manifold M). Sometimes the cost c is not necessarly induced by a distance. Let $\rho_0,\ \rho_1 \in \mathcal{P}(X)$. When we write ρ_0 *compact*, it means that the measure ρ_0 is concentrated on a compact subset of X.

Space	Cost	Main assumptions	Year	Author(s)	Paper
$\mathbb{R}^n$	$\|.\|^2$	$\rho_0 << \mathcal{L}$	1991	Brenier	[14]
$\mathbb{R}^n$	c	c strict. conv. + $\rho_0 << \mathcal{L}$	1996	Gangbo, McCann	[39]
$\mathbb{R}^n$	$\|.\|$	$\rho_0, \rho_1 << \mathcal{L}$ Lipschitz densities	1999	Evans, Gangbo	[28]
$\mathbb{R}^n$	$\|.\|$	$\rho_0, \rho_1 << \mathcal{L}$	2001	Trudinger, Wang	[57]
(M, d)	d^2	M compact, smooth + $\rho_0 << \mathcal{L}$	2001	McCann	[51]
$\mathbb{R}^n$	$\\|.\\|$	$\\|.\\|$ unif. conv. + $\rho_0, \rho_1 << \mathcal{L}$ compact	2002	Caffarelli, Feldman, McCann	[16]
M	d	$\rho_0 << \mathcal{L}$ compact	2002	Feldman, McCann	[34]
$\mathbb{R}^n$	$\|.\|$	$\rho_0 << \mathcal{L}$	2003	Ambrosio	[4]
$\mathbb{R}^n$	$\\|.\\|$	$\\|.\\|$ unif. conv. + $\rho_0 << \mathcal{L}$	2003	Ambrosio, Pratelli	[8]
(X, H)	d_H^2	$\rho_0 << \mathcal{L}$	2004	Feyel, Ustünel	[36]
$\mathbb{R}^n$	$\\|.\\|$	$\\|.\\|$ crystalline + $\rho_0 << \mathcal{L}$	2004	Ambrosio, Kirchheim, Pratelli	[7]
(H, γ)	$\|.\|^p$	$\rho_0 << \gamma$	2005	Ambrosio, Gigli, Savare	[6]
(M, d)	c	M compact + c TL + $\rho_0 << \mathcal{L}$	2007	Bernard, Buffoni	[9]
(M, d)	d	$\rho_0 << \mathcal{L}$	2007	Figalli	[38]
(M, d)	c	c TL + $\rho_0 << \mathcal{L}$	2010	Fathi, Figalli	[33]
$\mathbb{R}^n$	$\\|.\\|$	$\\|.\\|$ strict. conv. + $\rho_0 << \mathcal{L}$	2010	Champion, De Pascale	[20]
$\mathbb{R}^n$	$\\|.\\|$	$\rho_0 << \mathcal{L}$	2011	Champion, De Pascale	[21]
$\mathbb{R}^n$	$\\|.\\|$	$\rho_0 << \mathcal{L}$	2011	Caravenna	[17]
(X, H)	d_H	$\rho_0, \rho_1 << \mathcal{L}$	2012	Cavalletti	[19]
(X, d)	d^2	X CD(K,N) NB space + $\rho_0 << \mathcal{L}$	2012	Gigli	[42]

CD(K,N) means that X satisfies the *curvature-dimension condition.*
NB space means *non branching* space.
TL means cost induced by a *Tonelli Lagrangian* on the manifold.

Chapter 4

Convexity of relative entropy on infinite dimensional space

It has been proved by Sturm and von Renesse in [60] that on a Riemannian manifold, the Ricci curvature has a lower bound $K \in \mathbb{R}$ if and only if the relative entropy Ent_m relative to the Riemannian volume is $K-$convex along geodesics (see definition below). This is a starting point that Sturm, Lott and Villani studied the geometry for a measured metric space (X, d, m): the space (X, d, m) has a Ricci lower bound K if and only if the entropy Ent_m relative to m is K convex along geodesics. Shortly earlier, Otto arrived at describing solutions to heat equations, to porous medium equations or to a large class of non linear partial equations as gradient flows with respect to convex functionals on the space of probability measures. A general study on gradient flows over a metric space, especially on a Wasserstein space of probability measures has been done in [6], but the norm considered in the latter situation is strictly convex, satisfying conditions in Proposition 3.4.4.

The main objectif of this part is to prove that the classical Wiener space (X, H, μ) endowed with the uniform norm, seen as a measure metric space has 1 as the Ricci lower bound. The following result will be concerned with two norms: $|\cdot|_H, ||\cdot||_\infty$ introduced in Chapter 1.

Theorem 4.0.6. *Let ρ_0 and ρ_1 be two probability measures on X of finite entropy with respect to μ. Then there exists some constant speed geodesic ρ_t induced by an optimal coupling between ρ_0 and ρ_1 such that:*

$$Ent_\mu(\rho_t) \leq (1-t)Ent_\mu(\rho_0) + tEnt_\mu(\rho_1) - \frac{Kt(1-t)}{2} W_p^2(\rho_0, \rho_1) \qquad \forall t \in [0,1],$$

for $1 \leq p \leq 2$, where

- *$K = 1$, for $|.|_H$ and $p = 1$,*

- *$K = 1$, for $\|.\|_\infty$.*

Note that the notion of $K-$convexity of relative entropy introduced in [47] by Lott and Villani is stronger: they required that the above inequality holds for *all* constant speed geodesics. In many situations, there is unicity of geodesics between two given measures. However for the case of branching spaces (see [10]), the optimal coupling is not unique. Following [10], $\mathcal{P}(X)[p]$ is said to be a *non-branching space*, if any geodesic $\gamma : [0,1] \longrightarrow \mathcal{P}(X)[p]$ is uniquely determined by its restriction on a smaller interval. For example, Banach space with a strictly convex norm is non-branching, while Banach space with a non strictly convex norm is branching.

Instead of using powerfull tools like *Gromov-Hausdorff* convergence or $\mathbb{D}-$convergence introduced by Sturm in [55], we will use finite dimensional approximations as Fang, Shao and Sturm in [32], who have treated the case of the Cameron-Martin norm.

In the current language, we say that $(X, \|.\|_\infty)$ is a $CD(1,\infty)$ space. As consequences over space $(X, \|.\|_\infty)$, we can get Brunn-Minkowski, Bishop-Gromov or Log-Sobolev inequalities (see [5]).

The organization of this chapter is as follows. We start with some definitions and properties of the relative entropy with respect to a reference measure on a Polish space. In the second section we prove some results on finite dimensional spaces, with the standard Gaussian measure as the reference measure. We also get inequalities for some slightly modified *Wasserstein distance* : They are not true distance, but this kind of inequalities will be used to prove Theorem 6.1.6. At last we deal with the main purpose of this chapter, that is to get $K-$convexity of the relative entropy on infinite dimensional spaces.

4.1 Relative entropy

4.1.1 Definition and properties

Let (X, d, m) be a measured metric space, that is, (X, d) is a Polish space and m is a probability measure on X. The *relative entropy w.r.t. m* is the functional $Ent_m : \mathcal{P}(X) \longrightarrow [0,\infty]$ defined as

$$Ent_m(\rho) := \begin{cases} \int f \log(f) dm & \text{if } \rho \text{ admits the density } f \text{ w.r.t } m, \\ +\infty & \text{otherwise} \end{cases} \tag{4.1.1}$$

Denote by $D(Ent_m)$ the *domain* in $\mathcal{P}(X)$ on which the relative entropy Ent_μ is well-defined. That is: $\rho \in D(Ent_m)$ if and only if $\mathrm{Ent}_m(\rho) < +\infty$. In particular any probability measure belonging to $D(Ent_m)$ is absolutely continuous w.r.t. m.

A basic result concerning $\rho \to \mathrm{Ent}_m(\rho)$ is

Proposition 4.1.1. *With respect to the weak topology,*

1. $\rho \to Ent_m(\rho)$ *is lower semicontinuous.*
2. *The subset* $\{\rho \in \mathcal{P}(X),\ Ent_m(\rho) \leq R\}$ *is compact in* $\mathcal{P}(X)$.

Proof. The item 1 is well-known (see for instance Lemma 9.4.3) in [6], while the item 2 is a direct consequence of Vallé-Poussin lemma, which says that any uniformly integrable family is a sequentially relatively compact subset with respect to the weak topology of $L^1(X, m)$. □

4.1.2 Convexity along geodesics

Here and thereafter (X, d) will stand for either a Polish space or a Wiener space (X, H, d_H). Let $p \geq 1$; consider the Wasserstein distance W_p, that is,

$$W_p(\rho_0, \rho_1) = \left(\inf_{\Pi \in C(\rho_0, \rho_1)} \int_{X \times X} d(x, y)^p d\Pi(x, y) \right)^{1/p}.$$

Thanks to the Proposition 3.3.2, $(\mathcal{P}(X)[p], W_p)$ is a complete metric space. Therefore we can introduce a notion of *geodesics* over this space. A curve $t \in [0, 1] \longmapsto \rho_t \in \mathcal{P}(X)[p]$ is said to be a *constant speed geodesic*, provided

$$W_p(\rho_t, \rho_s) = (t - s) W_p(\rho_0, \rho_1), \quad \forall 0 \leq s \leq t \leq 1.$$

One can obtain a constant speed geodesic by picking an optimal coupling Π (for the cost d^p) between ρ_0 and ρ_1 and letting

$$\rho_t := ((1 - t)P_1 + tP_2)_\# \Pi, \quad \forall t \in [0, 1], \tag{4.1.2}$$

where $P_1 : X \times X \to X$ is the first projection, while P_2 is the second projection. The curve $t \to \rho_t$ obtained in (4.1.2) is a constant speed geodesic, that we will call the *McCann's interpolation* between ρ_0 and ρ_1. We refer to [58] for a general theory about dynamical optimal couplings which provides constant speed geodesics in $(\mathcal{P}(X)[p], W_p)$. However for our purpose we will focus on geodesics defined in (4.1.2).

Definition 4.1.2. *Let $\rho_0,\ \rho_1 \in \mathcal{P}(X)[p]$; We say that the relative entropy with respect to a reference measure m, is $K-$*geodesically convex *in $(\mathcal{P}(X)[p], W_p)$ if there exists a constant speed geodesic ρ_t between ρ_0 and ρ_1 such that:*

$$Ent_m(\rho_t) \leq (1-t)Ent_m(\rho_0) + tEnt_m(\rho_1) - \frac{Kt(1-t)}{2} W_p^2(\rho_0, \rho_1), \qquad \forall t \in [0,1].$$

We say that relative entropy is strongly $K-$geodesically convex *in $(\mathcal{P}(X)[p], W_p)$ if the latter inequality holds* <u>*for all*</u> *constant speed geodesics ρ_t between ρ_0 and ρ_1.*

Throughout this chapter, we denote by $T_t := (1-t)P_1 + tP_2$ for $t \in [0,1]$. Moreover the interpolation between two probability measures ρ_0 and ρ_1, will always be the following

$$\rho_t := (T_t)_\# \Pi = ((1-t)P_1 + tP_2)_\# \Pi,$$

for any optimal coupling $\Pi \in C_0(\rho_0, \rho_1)$, in the sense that Π minimizes **(MKP)**

$$\inf_{\Pi \in C(\rho_0,\rho_1)} \int_{X\times X} c(x,y) d\Pi(x,y). \tag{MKP}$$

4.2 The case of finite dimension

This section is devoted to establish some convexity results in finite dimensional spaces, say $\mathbb{R}^n$. These results depend on

- the reference measure m, because of the definition of the relative entropy,
- the metric considered on $\mathbb{R}^n$, because of the definition of the Wasserstein distance.

We will use m to denote for either the Lebesgue measure $\mathcal{L}$ or the standard Gaussian measure γ_n. Metrics considered are always norms in $\mathbb{R}^n$. For the purpose in Chapter 6 (see Theorem 6.1.6), we have to consider a cost function, which is not induced by a distance. In this situation, instead of considering constant speed geodesics which are not defined, we will consider the McCann's interpolation defined in (4.1.2).

In order to extend results in infinite dimensional spaces, we will take Gaussian measures as reference measures. Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$. We consider two probability measures ρ_0 and ρ_1 on $\mathbb{R}^n$ belonging to $D(Ent_{\gamma_n})$.

The following Proposition states that the relative entropy with respect to the Lebesgue measure on $\mathbb{R}^n$ is *geodesically convex* in $(\mathcal{P}_p(\mathbb{R}^n), W_p)$ whatever $p > 1$. It will play a fundamental role in getting other results of *convexity* of the relative entropy, when the reference measure is absolutely continuous with respect to the Lebesgue measure.

Proposition 4.2.1. *Let $||\cdot||$ be a strictly convex norm, C^2 on $\mathbb{R}^n\backslash\{0\}$. Then for any optimal coupling Π between ρ_0, ρ_1 for $c := ||\cdot||^p$, the McCann's interpolation $\rho_t := (T_t)_\#\Pi$ satisfies*

$$Ent_{\mathcal{L}}(\rho_t) \leq (1-t)Ent_{\mathcal{L}}(\rho_0) + tEnt_{\mathcal{L}}(\rho_1), \qquad \forall t \in [0,1]. \tag{4.2.1}$$

Proof. For the sake of self-contained, we will give a sketch of proof, which is taken from [6], page 213. By assumptions on c, the Theorem 3.4.2 provides us an optimal transport map T which pushes ρ_0 forward to ρ_1. Moreover it is well known that $T_t := (1-t)Id + tT$ is an optimal transport map which pushes ρ_0 forward to $\rho_t := (T_t)_\#\rho_0$.
By Proposition 3.4.4, T is approximately differentiable ρ_0-a.s. and its approximate differential $\tilde{\nabla}T$ is diagonalizable with nonnegative eigenvalues. Besides $det(\tilde{\nabla}T(x)) > 0$ ρ_0-a.s. in $x \in \mathbb{R}^n$. Therefore $\tilde{\nabla}T_t$ is diagonalizable too, with positive eigenvalues and denote by f_t the density of ρ_t (for $t \in [0,1]$). It follows by (3.4.1),

$$Ent_{\mathcal{L}}(\rho_t) = \int_{\mathbb{R}^n} f_t \log f_t d\mathcal{L} = \int_{\mathbb{R}^n} f_0(x) \log \frac{f_0(x)}{det(\tilde{\nabla}T_t(x))} dx.$$

Since the map $t \in [0,1] \longmapsto det((1-t)Id + t\tilde{\nabla}T)^{1/n}$ is concave, $t \longmapsto f_0(x) \log \frac{f_0(x)}{t^n}$ is convex and non increasing, we get

$$f_0(x) \log \frac{f_0(x)}{det(\tilde{\nabla}T_t(x))} \leq (1-t)f_0(x)\log f_0(x) + tf_0(x) \log \frac{f_0(x)}{det(\tilde{\nabla}T(x))}.$$

Integrating w.r.t. $\mathcal{L}$ gives the result. □

Let $\|.\|$ be a norm, C^2-differentiable on $\mathbb{R}^n\backslash\{0\}$ satisfying

$$\|x\| \leq \frac{1}{\sqrt{K}}|x|. \tag{4.2.2}$$

Recall that

$$W_{p,||\cdot||}(\rho_0,\rho_1) = \inf_{\Pi \in C(\rho_0,\rho_1)} \Big(\int_{\mathbb{R}^n\times\mathbb{R}^n} ||x-y||^p \, d\Pi(x,y)\Big)^{1/p}.$$

Proposition 4.2.2. *Let $1 < p \leq 2$; then for any optimal coupling Π between ρ_0, ρ_1 for $||\cdot||^p$, the McCann's interpolation $\rho_t := (T_t)_\#\Pi$ satisfies*

$$\mathrm{Ent}_{\gamma_n}(\rho_t) \leq (1-t)\mathrm{Ent}_{\gamma_n}(\rho_0) + t\mathrm{Ent}_{\gamma_n}(\rho_1) - \frac{K(1-t)}{2} W^2_{p,\|\cdot\|}(\rho_0,\rho_1). \tag{4.2.3}$$

For $p = 1$, there is an optimal coupling Π between ρ_0, ρ_1 for $||\cdot||$ such that the above inequality holds.
In particular if $\rho_0, \rho_1 \in D(Ent_{\gamma_n})$ then also $\rho_t \in D(Ent_{\gamma_n})$ for any $t \in (0,1)$.

Proof. We have:

$$Ent_{\gamma_n}(\rho_i) = Ent_{\mathcal{L}}(\rho_i) + \mathcal{V}(\rho_i) + \frac{n}{2}\log(2\pi),$$

where $\mathcal{V}(\rho_i) := \frac{1}{2}\int |x|^2 d\rho_i(x)$. By 1−convexity of the Euclidian norm, it is easy to see that

$$\mathcal{V}(\rho_t) \leq (1-t)\mathcal{V}(\rho_0) + t\mathcal{V}(\rho_1) - \frac{t(1-t)}{2}\int |x-y|^2 d\Pi(x,y).$$

Now by the Hölder inequality (because $2/p \geq 1$) and (4.2.2):

$$\mathcal{V}(\rho_t) \leq (1-t)\mathcal{V}(\rho_0) + t\mathcal{V}(\rho_1) - \frac{Kt(1-t)}{2} W^2_{p,\|.\|}(\rho_0,\rho_1). \tag{4.2.4}$$

For $p > 1$, the cost $\|.\|^p$ is strictly convex and we can apply Proposition 4.2.1 and take the sum with (4.2.4).

The case $p = 1$ is a little more tricky. Let $p \downarrow 1$; then $||x||^p$ converges to $||x||$ uniformly on any compact subsets of $\mathbb{R}^n$. We consider a sequence of optimal couplings $\Pi^p \in C(\rho_0,\rho_1)$ for $||\cdot||^p$. The interpolation $\rho^p_t := (T_t)_\#\Pi^p$ satisfies (4.2.3). Up to a subsequence, Π^p converges to $\Pi \in C(\rho_0,\rho_1)$ which is optimal for $||\cdot||$. Also ρ^p_t converges weakly to $\rho_t = (T_t)_\#\Pi$. Now by lower semi continuity of the relative entropy, the result

$$Ent_{\gamma_n}(\rho_t) \leq (1-t)Ent_{\gamma_n}(\rho_0) + tEnt_{\gamma_n}(\rho_1) - \frac{K(1-t)}{2} W^2_{1,\|.\|}(\rho_0,\rho_1).$$

□

In terms of Definition 4.1.2, the relative entropy w.r.t. the Gaussian measure γ_n on $(\mathbb{R}^n, \|.\|)$ is strongly K-geodesically convex in $(\mathcal{P}_p(\mathbb{R}^n), W_p)$ for any $1 < p \leq 2$ and it is convex for $p = 1$.

Note that for any $q \geq 2$, the norm $|x|_q = (\sum_{i=1}^n |x_i|^q)^{1/q} \leq |x|$; so the constant K in (4.2.2) for the norm $|\cdot|_q$ is equal to 1. On the classical Wiener space, $||x||_{k,\gamma} \leq C_{k,\gamma}|x|_H$; so their restriction on any finite dimensional subspace V_n satisfy the relation (4.2.2) with $K = 1/\sqrt{C_{k,\gamma}}$.

In what follows, we will extend the previous result to the uniform norm $|x|_\infty = \sup_{i=1,\dots,n} |x_i|$. Note that $|x-y|^p_\infty$ $(1 \leq p \leq 2)$ is neither strictly convex nor differentiable on $\mathbb{R}^n\backslash\{0\}$.

When one changes the cost function, the Wasserstein distance changes accordingly, as well as the constant speed geodesics.

Fix two probability measures ρ_0 and ρ_1 on $\mathbb{R}^n$ with finite second moments. For the sake of simplicity, we denote by $W_{p,q}$ the $p-$Wasserstein distance induced by the $q-$norm $|.|_q$. By hypothesis on ρ_0 and ρ_1, it is obvious that $W_{p,q}(\rho_0, \rho_1) < \infty$ for all $q \geq 2$ and all $1 \leq p \leq 2$.

Fix $1 \leq p \leq 2$. We know that for $q \geq 2$, there exists a unique coupling $\Pi_0^{(q)}$ between ρ_0 and ρ_1 optimal for the cost function $c_q(x,y)^p := |x-y|_q^p$. Let us first get a look on the behavior of the sequence $(\Pi_0^{(q)})_q$. We know that, when $q \to +\infty$, $|x|_q \to |x|_\infty$ uniformly on any compact subsets of $\mathbb{R}^n$. On the other hand, up to a subsequence, $(\Pi_0^{(q)})_q$ converges weakly to a probability measure which will be an optimal coupling for the cost $|\cdot|_\infty^p$. This fact, combined with the property of lower semicontinuity of the relative entropy, and the nonincreasing of the following sequence

$$q \in \mathbb{N} \longmapsto W_{p,q}^2(\rho_0, \rho_1),$$

will yield 1−convexity of relative entropy along geodesics with respect to $|\cdot|_\infty^p$.

Because of non strict convexity of $|.|_\infty$, $(\mathbb{R}^n, |.|_\infty)$ is a branching space: there exists many constant speed geodesics between two probability measures.

Proposition 4.2.3. *Let $1 \leq p \leq 2$; then there is an optimal coupling $\Pi \in C_o(\rho_0, \rho_1)$ with respect to the cost $c^p(x,y) := |x-y|_\infty^p$, such that for any $t \in (0,1)$:*

$$Ent_{\gamma_n}(\rho_t) \leq (1-t)Ent_{\gamma_n}(\rho_0) + tEnt_{\gamma_n}(\rho_1) - \frac{t(1-t)}{2} W_{p,\infty}^2(\rho_0, \rho_1), \tag{4.2.5}$$

where $\rho_t = ((1-t)P_1 + tP_2)_\# \Pi$. In particular if $\rho_0, \rho_1 \in D(Ent_{\gamma_n})$ then also $\rho_t \in D(Ent_{\gamma_n})$ for any $t \in (0,1)$.

Proof. To prove the weak convergence of $(\Pi_0^{(q)})_q$, we remark that the sequence is tight. By Prokohov's Theorem, there exists a subsequence $(\Pi_0^{(q_k)})_{q_k}$ that we will denote by $(\Pi_0^{(q)})_q$ again, converging weakly to a measure Π^∞. It is easy to check that Π^∞ is a coupling of ρ_0 and ρ_1. For the optimality of Π^∞, we apply the Proposition 3.2.7, taking $\mu_k = \rho_0$ and $\nu_k = \rho_1$.
For $q \in [2, +\infty)$ we consider associated constant speed geodesics

$$\rho_t^{(q)} := (T_t)_\# \Pi_0^q.$$

Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a bounded continuous function. We have

$$\int_{\mathbb{R}^n} \psi(x)\, d\rho_t^{(q)}(dx) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \psi(tx + (1-t)y)\, d\Pi_0^q(x,y),$$

which converges to $\int_{\mathbb{R}^n\times\mathbb{R}^n} \psi(tx+(1-t)y)\, d\Pi_0^\infty(x,y)$. Hence the sequence $(\rho_t^{(q)})_q$ converges weakly to ρ_t^∞ for all $t\in[0,1]$. Applying Proposition 4.2.2 with $|.|_q$ norms, we get:

$$Ent_{\gamma_n}(\rho_t^{(q)}) \leq (1-t)Ent_{\gamma_n}(\rho_0) + tEnt_{\gamma_n}(\rho_1) - \frac{t(1-t)}{2}W_{p,q}^2(\rho_0,\rho_1), \qquad (4.2.6)$$

for all $q\geq 2$. Note that

$$W_{p,q}(\rho_0,\rho_1)\geq W_{p,\infty}(\rho_0,\rho_1).$$

Since the relative entropy is lower semi-continuous, it holds

$$\liminf_q Ent_{\gamma_n}(\rho_t^{(q)}) \geq Ent_{\gamma_n}(\rho_t^\infty).$$

Finally, combining this two arguments, taking the liminf in the inequality (4.2.6) with respect to q, we get the result:

$$Ent_{\gamma_n}(\rho_t^\infty) \leq (1-t)Ent_{\gamma_n}(\rho_0) + tEnt_{\gamma_n}(\rho_1) - \frac{t(1-t)}{2}W_{p,\infty}^2(\rho_0,\rho_1).$$

□

For a C^2 differentiable norm $\|.\|$ on $\mathbb{R}^n\backslash\{0\}$, we introduce the quantity:

$$\mathcal{W}_{\varepsilon,\|.\|}(\rho_0,\rho_1) := \inf_{\Pi\in C(\rho_0,\rho_1)} \int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\| + \varepsilon\alpha(x-y)d\Pi(x,y),$$

where

$$\alpha(x-y) := \left(1+||x-y||^2\right)^{1/2}.$$

Note that α is a strictly convex and differentiable function on $\mathbb{R}^n$. Under the condition (4.2.2), we have the relation:

$$c_{\varepsilon,\|.\|}(x-y) := \|x-y\| + \varepsilon\alpha(x-y) \leq \varepsilon + \frac{1+\varepsilon}{\sqrt{K}}|x-y|, \qquad (4.2.7)$$

where $|\cdot|$ denotes the Euclidean norm of $\mathbb{R}^n$.
It is obvious that

$$\mathcal{W}_{\varepsilon,\|.\|}(\rho_0,\rho_1) \geq W_{1,\|.\|}(\rho_0,\rho_1).$$

So for $\rho_0\neq\rho_1$, there is a small $\varepsilon>0$ such that

$$\mathcal{W}_{\varepsilon,\|.\|}(\rho_0,\rho_1) - \varepsilon \geq W_{1,\|.\|}(\rho_0,\rho_1) - \varepsilon > 0.$$

Proposition 4.2.4. *There is an optimal coupling* Π *with respect to the cost* $c_{\varepsilon,\|\cdot\|}$*, such that for any* $t \in (0,1)$*,*

$$Ent_{\gamma_n}(\rho_t) \le (1-t)Ent_{\gamma_n}(\rho_0)+tEnt_{\gamma_n}(\rho_1)-\frac{t(1-t)}{2}\frac{K}{(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,\|\cdot\|}(\rho_0,\rho_1)-\varepsilon\right)^2. \tag{4.2.8}$$

In particular if $\rho_0, \rho_1 \in D(Ent_{\gamma_n})$*, then also* $\rho_t \in D(Ent_{\gamma_n})$ *for any* $t \in (0,1)$*.*

Proof. Let $p \downarrow 1$, and $\Pi^{(p)}$ be an optimal coupling with respect to $||\cdot||^p + \varepsilon\alpha$. As $p \to 1$, $||x||^p + \varepsilon\alpha(x)$ converges uniformly to $c_{\varepsilon,||.||}(x)$ over any compact subsets of $\mathbb{R}^n$. So up to a subsequence, $\Pi^{(p)}$ converges weakly to an optimal coupling Π with respect to $c_{\varepsilon,||.||}(x)$, also $\rho_t^{(p)}$ converges weakly to $\rho_t = ((1-t)P_1 + tP_2)_{\#}\Pi$. We can assume that $\rho_0, \rho_1 \in D(Ent_{\gamma_n})$; otherwise the inequality is obvious. Since ρ_0 and ρ_1 are two probability measures absolutely continuous with respect to γ_n, they are also absolutely continuous with respect to the Lebesgue measure $\mathcal{L}$. Moreover

$$\mathrm{Ent}_{\gamma_n}(\rho_i) = \mathrm{Ent}_{\mathcal{L}}(\rho_i) + \frac{n}{2}\log(2\pi) + \mathcal{V}(\rho_i),$$

where $\mathcal{V}(\rho) := \frac{1}{2}\int |x|^2 d\rho(x)$. By $1-$convexity of the Euclidian norm, it is easy to see that:

$$\mathcal{V}(\rho_t^{(p)}) \le (1-t)\mathcal{V}(\rho_0) + t\mathcal{V}(\rho_1) - \frac{t(1-t)}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\Pi^{(p)}(x,y).$$

For the cost $||\cdot||^p + \varepsilon\alpha$, we can apply (4.2.1), so that

$$\mathrm{Ent}_{\gamma_n}(\rho_t^{(p)}) \le (1-t)\mathrm{Ent}_{\gamma_n}(\rho_0) + t\mathrm{Ent}_{\gamma_n}(\rho_1) - \frac{t(1-t)}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\Pi^{(p)}(x,y).$$

Letting $p \to 1$ yields

$$\mathrm{Ent}_{\gamma_n}(\rho_t) \le (1-t)\mathrm{Ent}_{\gamma_n}(\rho_0) + t\mathrm{Ent}_{\gamma_n}(\rho_1) - \frac{t(1-t)}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2 d\Pi(x,y).$$

The result (4.2.8) follows, by Cauchy-Schwarz's inequality and remarking that

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y| d\Pi(x,y) \ge \frac{\sqrt{K}}{1+\varepsilon}(\mathcal{W}_{\varepsilon,\|\cdot\|}(\rho_0,\rho_1)-\varepsilon).$$

□

4.3 On infinite dimensional spaces

Let (X, H, μ) be an abstract Wiener space. Let V_n be a subspace of H introduced as in section 2.1.1; we have finite dimensional approximations $\pi_n : X \longrightarrow V_n$ and the decomposition $X = V_n \oplus V_n^\perp$, with $\mu = \gamma_n \otimes \nu$, where ν is the Wiener measure on $(V_n^\perp, V_n^\perp \cap H, \nu)$. Let c be a cost function induced by a power of pseudo-norm on X. Let $\rho_0,\ \rho_1 \in \mathcal{P}(X)$ such that

$$\mathcal{W}(\rho_0, \rho_1) := \inf_{\Pi \in C(\rho_0,\rho_1)} \int_{X\times X} c(x-y)d\Pi(x,y) > 0.$$

We denote by $\rho_i^n := (\pi_n)_\# \rho_i$ for $i = 0, 1$. We assume that

$$c(\pi_n(x), \pi_n(y)) \le c(x,y). \tag{4.3.1}$$

Proposition 4.3.1. *Let c_n be the restriction of c on $V_n \times V_n$; then*

$$\lim_{n\to\infty} \mathcal{W}_{c_n}(\rho_0^n, \rho_1^n) = \mathcal{W}_c(\rho_0, \rho_1).$$

Proof. Take an optimal coupling $\Pi \in C(\rho_0, \rho_1)$ for c. Then for $n \in \mathbb{N}$, $\Pi_n := (\pi_n \times \pi_n)_\# \Pi \in C(\rho_0^n, \rho_1^n)$ and thanks to (4.3.1),

$$\begin{aligned}\int_{V_n\times V_n} c_n(x,y)d\Pi_n &= \int_{X\times X} c(\pi_n(x), \pi_n(y))_{V_n} d\Pi \\ &\le \int_{X\times X} c(x,y)d\Pi = \mathcal{W}_c(\rho_0,\rho_1).\end{aligned}$$

Taking the sup on $n \in \mathbb{N}$, we get

$$\sup_n \mathcal{W}_{c_n}(\rho_0^n, \rho_1^n) \le \mathcal{W}_c(\rho_0, \rho_1). \tag{4.3.2}$$

On the other hand, for $n \in \mathbb{N}$, take $\Pi_n \in C(\rho_0^n, \rho_1^n)$ optimal for c_n and we define $\hat{\Pi}_n$ in such a way: for any bounded continuous function $\psi : X \times X \longrightarrow \mathbb{R}$,

$$\int_{X\times X} \psi(x,y)d\hat{\Pi}_n = \int_{V_n^\perp} \left(\int_{V_n\times V_n} \psi(x_n+\xi, y_n+\xi)d\Pi_n(x_n,y_n) \right) d\nu(\xi). \tag{4.3.3}$$

Then $\hat{\Pi}_n \in C(\rho_0^n \circ \pi_n, \rho_1^n \circ \pi_n)$. Since the sequence $(\rho_0^n \circ \pi_n)_n$ converges to ρ_0 and $(\rho_1^n \circ \pi_n)_n$ converges to ρ_1 in $L^1(X)$, there exists a subsequence of $(\hat{\Pi}_n)_n$ which converges weakly to $\hat{\Pi} \in C(\rho_0, \rho_1)$. We have

$$\int_{X\times X} c(x,y)\,d\hat{\Pi}_n(x,y) = \int_{V_n}\Big[\int_{V_n\times V_n} c(x_n+\xi, y_n+\xi) d\Pi_n(x_n,y_n)\Big] d\nu(\xi) = W_{c_n}(\rho_0^n,\rho_1^n).$$

Therefore

$$\liminf_{n\to\infty} \mathcal{W}_{c_n}(\rho_0^n,\rho_1^n) \geq \int_{X\times X} c(x,y)\,d\Pi(x,y) \geq \mathcal{W}_c(\rho_0,\rho_1).$$

Combining with (4.3.2), the result follows. □

Remark 4.3.2. Let $\tilde{\rho}_0^n = \rho_0^n \circ \pi_n, \tilde{\rho}_1^n = \rho_1^n \circ \pi_n$. The above computation shows that

i) $W_{c_n}(\rho_0^n,\rho_1^n) = W_c(\tilde{\rho}_0^n,\tilde{\rho}_1^n)$,

ii) If Π_n is an optimal coupling in $C(\rho_0^n,\rho_1^n)$, then $\hat{\Pi}_n$ defined in (4.3.3) is an optimal coupling in $C(\tilde{\rho}_0^n\tilde{\rho}_1^n)$.

4.3.1 On a Hilbert space

Let X be a separable Hilbert space with inner product $\langle\ ,\ \rangle_X$. A Borel probability measure γ on X is said to be (centered) Gaussian measure if

$$\int_X e^{i\langle x,y\rangle_X}\,d\gamma(y) = e^{-\frac{1}{2}\langle Bx,x\rangle_X},$$

where B is a positive symmetric trace operator. Let $\{e_n;\ n\geq 0\}$ be an orthonormal basis of X, of eigenvectors of B such that

$$Be_n = c_n\,e_n, \quad c_n > 0.$$

Then we have

$$\int_H e^{i\xi\langle e_n,y\rangle_X}\,d\gamma(y) = e^{-(c_n\xi^2)/2},$$

which means that the projection $x \to \langle x, e_n\rangle_X$ pushes γ forward to a Gaussian measure on $\mathbb{R}$, of variance c_n. Let c denote the sequence $(c_n)_{n\geq 0}$. Then

$$\sum_{n\geq 0} c_n < +\infty.$$

Consider the application $\Phi : X \to \mathbb{R}^{\mathbb{N}}$ defined by

$$x \to (\langle e_n, x\rangle_H/\sqrt{c_n})_{n\geq 0}.$$

Let
$$l^2(c) := \{x \in \mathbb{R}^{\mathbb{N}},\ \sum_{n\geq 0} c_n x_n^2 < \infty\}.$$
Then Φ sends X onto $l^2(c)$ and $\mu = \Phi_{\#}\gamma$ is the countable product of standard Gaussian measures on $\mathbb{R}$. It is known that the measure μ is quasi-invariant under translation of elements in
$$l^2 = \{x \in \mathbb{R}^{\mathbb{N}},\ \sum_{n\geq 0} x_n^2 < \infty\}.$$
More precisely, for $h \in l^2$ and $\tau_h(x) = x + h$, then $d(\tau_h)_{\#}\mu = \rho_h\, d\mu$, with
$$\rho_h(x) = e^{-\frac{1}{2}|h|_{l^2}^2 - \langle h,x\rangle},$$
where $\langle h, x\rangle = \sum_{n\geq 0} h_h x_n$. Note that
$$l^2 \subset l^2(c), \quad |x|_{l^2(c)} \leq \max\{c_n\} \times |x|_{l^2}.$$
In other words, $(l^2(c), l^2, \mu)$ is an abstract Wiener space. For the simplicity, we will suppose that
$$\max\{c_n; n \geq 0\} \leq 1;$$
so the constant K in (4.2.2) is equal to 1.
Let $V_n = (x_0, x_1, \ldots, x_n, 0, \cdots)$ and $\pi_n : X \to V_n$ be the canonical projection. Then we have
$$|x|_{V_n}^2 := |\pi_n(x)|_{l^2(c)}^2 = \sum_{k=0}^{n} c_k x_k^2 \leq |x|_{l^2(c)}^2. \tag{4.3.4}$$

In what follows, we will set $X = l^2(c)$, $H = l^2$ and $||\cdot||$ the Hilbertian norm of X. Let $\rho_0,\ \rho_1 \in \mathcal{P}(X)$ such that $W_{1,||.||}(\rho_0, \rho_1) > 0$. In the sequel, $\varepsilon > 0$ is taken small enough so that $W_{1,||.||}(\rho_0, \rho_1) - \varepsilon > 0$. By Proposition 4.3.1, for n big enough $W_{1,||.||_n}(\rho_0, \rho_1) - \varepsilon$ is still positive, where $||.||_n$ denotes the restriction of $||.||$ on V_n.

In Chapter 6, we will consider the following variational problem:
$$\min_{\Pi\in C(\rho_0,\rho_1)} \Big[\int_{X\times X} ||x-y|| d\Pi(x,y) + \varepsilon \int_{X\times X} \alpha(x-y) d\Pi(x,y)\Big], \tag{P_ε}$$
where α is defined by
$$\alpha(x-y) := \left(1 + ||x-y||^2\right)^{1/2}.$$
Thanks to (4.3.4), it holds
$$||\pi_n(x)|| + \varepsilon\alpha(\pi_n(x)) \leq ||x|| + \varepsilon\alpha(x). \tag{4.3.5}$$

The following result extends the Proposition 4.2.4 to the infinite dimensional Hilbert space.

Proposition 4.3.3. *There is a solution Π_ε to (P_ε), such that, If $\rho_t := ((1-t)P_1 + tP_2)_\#\Pi_\varepsilon$ then for any $t \in (0,1)$, $\rho_t \in D(Ent_\mu)$ and:*

$$Ent_\mu(\rho_t) \le (1-t)Ent_\mu(\rho_0) + tEnt_\mu(\rho_1) - \frac{t(1-t)}{2(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,\|.\|}(\rho_0,\rho_1) - \varepsilon\right)^2. \quad (4.3.6)$$

Proof. For any $n \ge 1$, we consider $\rho_i^n = (\pi_n)_\#\rho_i$ as above. By Proposition 4.2.4, there is an optimal coupling $\Pi_n \in C(\rho_0^n, \rho_1^n)$ such that

$$Ent_{\gamma_n}(\rho_t^n) \le (1-t)Ent_{\gamma_n}(\rho_0^n) + tEnt_{\gamma_n}(\rho_1^n) - \frac{t(1-t)}{2(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,\|.\|_n}(\rho_0^n,\rho_1^n) - \varepsilon\right)^2,$$

where $\rho_t^n := ((1-t)P_1 + tP_2)_\#\Pi_n$ for $t \in (0,1)$. Let $\hat{\Pi}_n$ be defined in (4.3.3), and $\hat{\rho}_t^n = ((1-t)P_1 + tP_2)_\#\hat{\Pi}_n$. Then for any bounded continuous function $\psi : X \to \mathbb{R}$,

$$\begin{aligned}
&\int_{X\times X} \psi((1-t)x + ty)\, d\hat{\Pi}_n(x,y)\\
&= \int_{V_n^\perp}\Big[\int_{V_n\times V_n} \psi((1-t)(x_n+\xi) + t(y_n+\xi))d\Pi_n(x_n,y_n)\Big]\, d\nu(\xi)\\
&= \int_{V_n^\perp}\Big[\int_{V_n} \psi(x+\xi)d\rho_t^n(x)\Big]\, d\nu(\xi) = \int_X \psi(x)\, f_t^n \circ \pi_n(x) d\mu(x)
\end{aligned}$$

where f_t^n denotes the density of ρ_t^n with respect to γ_n. It follows that $\hat{\rho}_t^n$ has $f_t^n \circ \pi_n$ as density with respect to μ. Therefore

$$\mathrm{Ent}_\mu(\hat{\rho}_t^n) = \mathrm{Ent}_{\gamma_n}(\rho_t^n), \qquad \forall t \in [0,1],$$

and combining with Remark 4.3.2, we have for all $t \in [0,1]$:

$$\mathrm{Ent}_\mu(\hat{\rho}_t^n) \le (1-t)\mathrm{Ent}_\mu(\tilde{\rho}_0^n) + t\mathrm{Ent}_\mu(\tilde{\rho}_1^n) - \frac{t(1-t)}{2(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,\|.\|}(\tilde{\rho}_0^n,\tilde{\rho}_1^n) - \varepsilon\right)^2.$$

Now $d\tilde{\rho}_i^n = \mathbb{E}^{V_n}(\rho_i)\, d\mu$ for $i = 0,1$; then by Jensen inequality, $\mathrm{Ent}_\mu(\tilde{\rho}_i^n) \le \mathrm{Ent}_\mu(\rho_i)$. Since $(\hat{\Pi}_n)_n$ converges weakly to Π, so that $(\rho_t^n)_n$ converges weakly to ρ_t. Letting $n \to +\infty$ in above inequality yields

$$\mathrm{Ent}_\mu(\rho_t) \le (1-t)\mathrm{Ent}_\mu(\rho_0) + t\mathrm{Ent}_\mu(\rho_1) - \frac{t(1-t)}{2(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,\|.\|}(\rho_0,\rho_1) - \varepsilon\right)^2.$$

Since the cost function c is continuous on $X \times X$, the coupling $\Pi \in C(\rho_0,\rho_1)$ is optimal with respect to c. □

In the next Corollary, we deal with the true *Wasserstein distance* $W_{1,\|.\|}$ on $\mathcal{P}(X)$. In this case for any optimal coupling $\Pi \in C(\rho_0,\rho_1)$, the McCann's interpolation ρ_t is a constant speed geodesic, namely

$$W_{1,\|.\|}(\rho_t,\rho_s) = |t-s| W_{1,\|.\|}(\rho_0,\rho_1), \qquad \forall t \in [0,1].$$

Corollary 4.3.4. *There is an optimal coupling* $\Pi \in C(\rho_0, \rho_1)$ *such that for any* $t \in (0,1)$, $\rho_t \in D(Ent_\mu)$ *and:*

$$Ent_\mu(\rho_t) \leq (1-t)Ent_\mu(\rho_0) + tEnt_\mu(\rho_1) - \frac{t(1-t)}{2} W^2_{1,||\cdot||}(\rho_0, \rho_1). \tag{4.3.7}$$

In the literature, this proposition can be reformulated as: *the relative entropy is geodesically* $1-$*convex in* $(\mathcal{P}(X), W_{1,|.|})$.

Proof. Using Proposition 4.2.2 and the same proof as above yields the result. □

4.3.2 On a Wiener space

In this section we will deal with the classical Wiener space (X, H, μ) with its Wiener measure μ. Note that X endowed with the uniform norm $||.||_\infty$, together with the Wiener measure μ is the simplest example of infinite dimensional measured metric space. When the cost is arised from the square of the Cameron-Martin norm, the 1- convexity of entropy with respect to μ has been given in [32].
Now let V_n be the subspace introduced in (2.2.1), constitued of continuous functions which are linear on each intervall $[l2^{-n}, (l+1)2^{-n}]$ for $l = 0, \ldots, 2^n - 1$. Let $\pi_n : X \to V_n$ be the projection and note that, in this case,

$$\|\pi_n(x)\|_\infty \leq \|x\|_\infty,$$

so that the Proposition 4.3.1 holds.

Theorem 4.3.5. *Let* ρ_0 *and* ρ_1 *be two probability measures in* $\mathcal{P}(X)$*. For* $p \in [1,2]$*, there exists an optimal coupling* Π *(with respect to* $\|.\|^p_\infty$*), for which the McCann interpolation* $\rho_t := (T_t)_\#\Pi$ *satisfies, for any* $t \in [0,1]$, $\rho_t \in D(Ent_\mu)$ *and:*

$$\mathrm{Ent}_\mu(\rho_t) \leq (1-t)\mathrm{Ent}_\mu(\rho_0) + t\mathrm{Ent}_\mu(\rho_1) - \frac{t(1-t)}{2} W^2_{p,\infty}(\rho_0, \rho_1). \tag{4.3.8}$$

In the literature, this proposition can be reformulated as: *the relative entropy is geodesically* $1-$*convex in* $(\mathcal{P}(X), W_{p,\infty})$.

Proof. As above, let $\rho_i^n = (\pi_n)_\#\mu$ for $i = 0, 1$. On the subsapce V_n, we first consider the norm

$$||x||_q^q = \int_0^1 |x(t)|^q \, dt,$$

which converges uniformy to $||x||_\infty$ on any compact subsets of V_n, as $q \to +\infty$. Proceeding as in the proof of the Proposition 4.2.3, we get an optimal coupling

$\Pi_n \in C(\rho_0^n, \rho_1^n)$ (with respect to $||.||_\infty$), for which the McCann interpolation ρ_t^n satisfies

$$\mathrm{Ent}_{\gamma_n}(\rho_t^n) \leq (1-t)\mathrm{Ent}_{\gamma_n}(\rho_0^n) + t\mathrm{Ent}_{\gamma_n}(\rho_1^n) - \frac{t(1-t)}{2} W_{p,\infty}^2(\rho_0^n, \rho_1^n).$$

Denote by $\hat{\rho}_i^n = \rho_i^n \circ \pi_n$, for $i = 0, 1$. Let $\hat{\Pi}_n \in C(\hat{\rho}_0^n, \hat{\rho}_1^n)$ be defined in (4.3.3). By Remark 4.3.2, $\hat{\Pi}_n$ is still optimal for $\|.\|_\infty^p$. We denote by $(\hat{\Pi}_{n_k})_k$ which converges weakly to some coupling $\hat{\Pi}$ between ρ_0 and ρ_1, optimal for $\|.\|_\infty^p$. We apply the Proposition 4.2.3 to obtain:

$$Ent_{\gamma_{n_k}}(\rho_t^{n_k}) \leq (1-t)Ent_{\gamma_{n_k}}(\rho_0^{n_k}) + tEnt_{\gamma_{n_k}}(\rho_1^{n_k}) - \frac{t(1-t)}{2} W_{p,\infty}^2(\rho_0^{n_k}, \rho_1^{n_k}) \quad \forall t \in [0,1]. \tag{4.3.9}$$

Now proceeding as in the proof of Proposition 4.3.3, we get the result by letting $k \to +\infty$. □

Remark 4.3.6. *For the norm $\|.\|_{k,\gamma}$ the proposition 4.3.1 does not hold anymore. Indeed it is not clear if $\|\pi_n(x)\|_{k,\gamma} \leq \|x\|_{k,\gamma}$ for any $x \in X$.*

Chapter 5

Logarithmic concave measures on the Wiener space

Let (X, H, μ) be an abstract Wiener space. A probability measure ν on X is said to be logarithmic concave, if there exists a a-convex function W on X such that

$$d\nu = e^{-W}\, d\mu,$$

for some $a \in [0, 1)$. This class of measures plays an important role in Analysis on the Wiener space. For example, the logarithmic Sobolev inequality still holds for such a measure ν (see the chapter 1).

It is now well-known (see [47]) that the convexity of relative entropy implies Talagrand's inequality. For the sake of self-contained, we will show this implication in section 1. In section 2, we will prove that the Wang's Harnack inequality is still true for a logarithmic concave measure: from the general theory of functional inequalities, the Harnack inequality implies the logarithmic Sobolev inequality. In section 3, we will study the stability of optimal transports when the target measure is logarithmic concave.

5.1 Talagrand's inequality

Talagrand's inequality with respect to the square of Cameron-Martin norm has been discussed in PhD thesis by I. Gentil. The implication from logarithmic Sobolev inequality to Talagrand's inequality has been estalished by Otto-Villani and Bobokov, Gentil and Ledoux. In this section, we only show the implication of the inequality (4.3.8) to

$$W_{2,\infty}^2(\rho_0, \mu) \le 2\mathrm{Ent}_\mu(\rho_0).$$

If there is a probability measure ρ_0 such that

$$\mathrm{Ent}_\mu(\rho_0) < \frac{1}{2} W_{2,\infty}^2(\rho_0, \mu),$$

then in the inequality (4.3.8), taking $\rho_1 = \mu$, we get

$$\mathrm{Ent}_\mu(\rho_t) \leq (1-t)\left(Ent_\mu(\rho_0) - \frac{t}{2} W_{2,\infty}^2(\rho_0, \mu) \right).$$

For a t close enough to 1 we have

$$Ent_\mu(\rho_0) < \frac{t}{2} W_{2,\infty}^2(\rho_0, \mu).$$

Then for this t,

$$\mathrm{Ent}_\mu(\rho_t) < 0.$$

But $\mathrm{Ent}_\mu(\rho_t) \geq 0$. We get a contradiction. Therefore for any probability measure ρ_0,

$$W_{2,\infty}^2(\rho_0, \mu) \leq 2\mathrm{Ent}_\mu(\rho_0).$$

□

5.2 Harnack's inequality

Harnack's inequalities was introduced by F. Wang in order to prove the logarithmic Sobolev inequality on complete Riemannian manifolds. There are now many applications of such an inequality, we refer to the paper of Bobkov, Gentil and Ledoux [11] and the book of Wang [61]. In infinite dimensional spaces, we refer to Shao [54] and to Aida and Zhang [1].

Let $V \in \mathbb{D}_1^2(X)$ be a positive function on the Wiener space X such that $\int_X V\, d\mu = 1$. Assume that

$$\int_X \frac{|\nabla V|^2}{V}\, d\mu < +\infty. \tag{5.2.1}$$

The condition (5.2.1) says that the Ficher information of the probability measure $\nu := V\mu$ is finite. Under this condition, the quadatic form

$$\mathcal{E}_V(f, f) = \int_X |\nabla f|^2\, V d\mu, \quad f \in \mathrm{Cylin}(X),$$

is closable, where $\mathrm{Cylin}(X)$ denotes the space of cylindrical functions on X. We will denote by $\mathbb{D}_1^2(X, \nu)$, or $Dom(\mathcal{E}_V)$ the minimal extension of $(\mathcal{E}_V, \mathrm{Cylin}(X))$. Set

$V = e^{-W}$. For the sake of simplicity, we denote $\mathcal{E}_W$ instead of $\mathcal{E}_V$. Let $\mathcal{L}_W$ be the generator of $\mathcal{E}_W$, that is associated to

$$\int_X |\nabla f|^2 \, e^{-W} d\mu = \int_X \mathcal{L}_W f \, f \, e^{-W} d\mu.$$

We have

$$\mathcal{L}_W f = \mathcal{L} f + \langle \nabla W, \nabla f \rangle_H \tag{5.2.2}$$

for all $f \in Cylin(X)$, where $\mathcal{L}$ is the Ornstein-Uhlenbeck operator on X.

Assume that

$$0 < \delta_1 \leq e^{-W} \leq \delta_2 < \infty. \tag{5.2.3}$$

Under (5.2.3) we have:

$$Dom(\mathcal{E}_W) = Dom(\mathcal{E}).$$

Now let $P_t^W = e^{-t\mathcal{L}_W}$ be the semigroup associated to $\mathcal{L}_W$. Then $P_t^W : L^p(X, e^{-W}\mu) \to L^p(X, e^{-W}\mu)$ is a contraction for any $1 \leq p \leq +\infty$, i.e. $\forall f \in L^p(X, e^{-W}\mu)$,

$$\|P_t f\|_{L^p(e^{-W}\mu)} \leq \|f\|_{L^p(e^{-W}\mu)}, \qquad \forall t \geq 0. \tag{5.2.4}$$

Proposition 5.2.1. *Let $W \in \mathbb{D}_1^2(X)$ and $(W_n)_n \subset \mathbb{D}_\infty^2(X)$ a sequence of functions satisfying* (5.2.3) *, which converges to W in $\mathbb{D}_1^2$. If P_t^n denotes the semigroup associated to $\mathcal{L}^{W_n}$, then*

$$\lim_{n\to\infty} \|P_t^n f - P_t^W f\|_{L^2(\mu)} = 0, \qquad \forall f \in \text{Cylin}(X).$$

Proof. Let $f \in \text{Cylin}(X)$ and $\nu_n := e^{-W_n}\mu$. Because $\frac{d}{dt} P_t f = -\mathcal{L}(P_t f)$, we have

$$\begin{aligned}
\frac{d}{dt} \int_X |P_t^n f - P_t^W f|^2 \, d\nu_n &= -2 \int_X \left(P_t^n f - P_t^W f\right) \left(\mathcal{L}_n P_t^n f - \mathcal{L}_W P_t^W f\right) \, d\nu_n \\
&= -2 \int_X \left(P_t^n f - P_t^W f\right) \mathcal{L}_n \left(P_t^n f - P_t^W f\right) d\nu_n \\
&\quad - 2 \int_X \left(P_t^n f - P_t^W f\right) \left(\mathcal{L}_n P_t^W f - \mathcal{L}_W P_t^W f\right) d\nu_n \\
&= I_1 + I_2,
\end{aligned}$$

By definition of $\mathcal{L}_n$, the first term is negative, that is, $I_1 \leq 0$. To estimate I_2, we remark

$$\mathcal{L}_n f - \mathcal{L}_W f =< \nabla(W_n - W), \nabla f >_H .$$

Hence by (5.2.3),

$$|I_2| \leq 2\delta_2 \|P_t^n f - P_t^W f\|_\infty \|\nabla W_n - \nabla W\|_{L^2(\mu)} \|\nabla P_t^W f\|_{L^2(\mu)}.$$

Moreover using (5.2.3), (5.2.4) and (5.2.2),

$$\begin{aligned}
||\nabla P_t^W f||^2_{L^2(\mu)} &\leq \frac{1}{\delta_1}\int_X |\nabla P_t^W f|^2\, e^{-W} d\mu = -\frac{1}{\delta_1}\int_X \mathcal{L}_W(P_t^W f)\, P_t^W f\, e^{-W} d\mu \\
&= -\frac{1}{\delta_1}\int_X P_t^W(\mathcal{L}_W f)\, P_t^W f\, e^{-W} d\mu \leq \frac{1}{\delta_1}||P_t^W(\mathcal{L}_W f)||_{L^2(e^{-W}\mu)}||P_t^W f||_{L^2(e^{-W}\mu)} \\
&\leq \frac{1}{\delta_1}||\mathcal{L}_W f||_{L^2(e^{-W}\mu)} \cdot ||f||_{L^2(e^{-W}\mu)} \leq \frac{\delta_2}{\delta_1}||\mathcal{L}_W f||_{L^2(\mu)} \cdot ||f||_{L^2(\mu)} \\
&\leq \frac{\delta_2}{\delta_1}(||\mathcal{L} f||_{L^2(\mu)} + ||\nabla f||_\infty\, ||\nabla W||_{L^2(\mu)}) \cdot ||f||_{L^2(\mu)}.
\end{aligned}$$

Combining above computations, there is a constant C, dependent on

$$\delta_1, \delta_2, ||f||_\infty, ||\nabla f||_\infty, ||\mathcal{L} f||_{L^2(\mu)}, ||\nabla W||_{L^2(\mu)}$$

such that

$$\frac{d}{dt}\int_X |P_t^n f - P_t^W f|^2\, d\nu_n \leq C\, ||\nabla W_n - \nabla W||_{L^2(\mu)}.$$

It follows that for $t > 0$,

$$\int_X |P_t^n f - P_t^W f|^2\, d\nu_n \leq t\, C\, ||\nabla W_n - \nabla W||_{L^2(\mu)} \to 0 \quad \text{as } n \to +\infty.$$

Finally note that $\delta_1 \leq e^{-W_n}$,

$$\|P_t^n f - P_t^W f\|_{L^2(\mu)} \leq \frac{1}{\delta_1}\|P_t^n f - P_t^W f\|_{L^2(\nu_n)} \to 0 \quad \text{as } n \to +\infty.$$

□

Let $K \in \mathbb{R}$ be a real number and $W \in \mathbb{D}_1^2(X)$ is a $K-$convex function on X satisfying the condition (5.2.3). Using the Ornstein-Uhlenbeck semi-group, we can get a sequence of K_n-convex functions $W_n \in \mathbb{D}_\infty^2(X)$ satisfying also (5.2.3), which converges to W in $\mathbb{D}_1^2(X)$, with

$$\lim_{n\to+\infty} K_n = K.$$

Theorem 5.2.2. *Let $K \in \mathbb{R}$, and $W \in \mathbb{D}_1^2(X)$ is a $K-$ convex function on X satisfying* (5.2.3). *Then for each $t > 0$*

$$|\nabla P_t^W f| \leq e^{-(K+1)t} P_t^W |\nabla f|, \quad \forall f \in Cylin(X).$$

Proof. For a K_n-convex function $W_n \in \mathbb{D}^2_\infty(X)$, we have

$$|\nabla P_t^n f| \leq e^{-(K_n+1)t} P_t^n |\nabla f|.$$

Let $\varepsilon > 0$ small. We can assume that $K_n \geq K - \varepsilon$. Hence integrating with respect to $\nu_n = e^{-W_n}\mu$,

$$\int_X |\nabla P_t^n f|^2 d\nu_n \leq e^{-2(K-\varepsilon+1)t} \int |\nabla f|^2 d\nu_n,$$

therefore

$$\int_X |\nabla P_t^n f|^2 d\mu \leq \frac{\delta_2}{\delta_1} e^{-2(K-\varepsilon+1)t} \int |\nabla f|^2 d\mu.$$

It follows that $(P_t^n f)_n$ is bounded in $\mathbb{D}^2_1(X)$; therefore there exists a subsequence (still denoted by $P_t^n f$), which converges weakly to some element $g \in \mathbb{D}^2_1(\mu)$. By Banach-Saks theorem, up to a subsequence,

$$\left(\frac{1}{n}\sum_{k=1}^n P_t^k f\right)_n$$

converges strongly to g in $\mathbb{D}^2_1(\mu)$. By the Proposition 5.2.1, the sequence $(P_t^k f)_k$ converges to $P_t^W f$ in $L^2(\mu)$, which yields

$$g = P_t^W f.$$

But

$$\left|\nabla\Big(\frac{1}{n}\sum_{k=1}^n P_t^k f\Big)\right| \leq \frac{1}{n}\sum_{k=1}^n |\nabla P_t^k f| \leq e^{-t(K-\varepsilon+1)}\frac{1}{n}\sum_{k=1}^n P_t^k |\nabla f|.$$

Letting $n \to \infty$ yields the result:

$$|\nabla P_t^W f| \leq e^{-(K-\varepsilon+1)t} P_t^W |\nabla f|.$$

The result follows by letting $\varepsilon \to 0$. □

As a consequence of gradient estimate, we get the following Harnack's inequality.

Proposition 5.2.3. *Let $W \in \mathbb{D}^2_1(X)$ be a K-convex function W on X satisfying (5.2.3). Assume that $\int_X e^{-W}\mu = 1$. Then for any $\alpha > 2$, any $t \geq 0$ and $f \in Cylin(X)$,*

$$|P_t^W f(w)|^\alpha \leq P_t^W |f|^\alpha(w') \exp\left\{\frac{\alpha(K+1)d_H(w,w')^2}{2(\alpha-1)(e^{2t}-1)}\right\}, \quad \forall w, w' \in X,$$

where

$$d_H(w,w') := \begin{cases} |w-w'|_H & \text{if} \quad w-w' \in H; \\ +\infty & \text{otherwise.}\end{cases}$$

Proof. The proof follows in the same line as in [61] or in [54]. □

Remark 5.2.4. The novelty in above proposition is we assume only that $W \in \mathbb{D}_1^2(X)$ instead of $W \in \mathbb{D}_2^2(X)$ in the literature. The technical condition $e^{-W} \geq \delta_1$ making the calculation easier, could be dropped.

5.3 Variation of optimal transport maps in Sobolev spaces

Another good behaviour of logarithmic concave measure is it insures the stability of optimal transport maps when the target measure satisfies such a property: It is the purpose of this section. The word *optimal* will always refer to optimality with respect to the cost being the *square of the Euclidian norm*, that is:

$$c(x, y) = |x - y|^2.$$

Let $e^{-V}dx$ and $e^{-W}dx$ be two probability measures on $\mathbb{R}^n$ having second moment, then there is a convex function Φ (Brenier's theorem) such that $\nabla\Phi$ is the optimal transport map which pushes $e^{-V}dx$ to $e^{-W}dx$. If moreover

1. the functions V and W are smooth, bounded from below,
2. the Hessian $\nabla^2 V$ of V is bounded from above and $\nabla W \geq K_1 \,\mathrm{Id}$ with $K_1 > 0$,

then Φ is smooth (see [15, 45]) and

$$\sup_{x\in\mathbb{R}^n} ||\nabla^2\Phi(x)||_{HS} < +\infty,$$

where

$$\|A\|_{HS} := Tr|A^{\star}A|,$$

denotes the Hilbert-Schmidt norm. The above upper bound is dimension-dependent. In a recent work [45], A.V. Kolesnikov proved the inequality

$$\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V}\,dx \geq K_1 \int_{\mathbb{R}^n} ||\nabla^2\Phi||_{HS}^2\, e^{-V} dx. \tag{5.3.1}$$

Although the constant K_1 in (5.3.1) is of dimension free, but on infinite dimensional spaces, $\nabla^2\Phi$ usually is not of Hilbert-Schmidt class. Let $\nabla\Phi(x) = x + \nabla\varphi(x)$. A dimension free inequality for $||\nabla^2\varphi||_{HS}^2$ has been established in [45] under the hypothesis

$$\nabla^2 W \leq K_2 \,\mathrm{Id}. \tag{5.3.2}$$

The main contribution of this section is to remove the condition (5.3.2).

Firstly we get a priori estimate, following the ideas in [45], mainly combining change of variables formula. It turns out that it can be extended in suitable Sobolev spaces. And this estimate leads to the main result of the section:

Theorem. *Let $e^{-V}d\gamma$ and $e^{-W}d\gamma$ be two probability measures on $\mathbb{R}^n$, where γ is the standard Gaussian measure on $\mathbb{R}^n$. Suppose that $\nabla^2 W \geq -c\,\mathrm{Id}$ with $c \in [0,1)$. Then*

$$\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla W|^2 e^{-W} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||_{HS}^2\, e^{-W} d\gamma$$
$$\geq 2\mathrm{Ent}_\gamma(e^{-V}) - 2\mathrm{Ent}_\gamma(e^{-W}) + \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2\, e^{-V} d\gamma.$$

5.3.1 A priori estimates

Consider a probability measure $d\mu = e^{-\alpha(x)}\,dx$ on the Euclidean space $(\mathbb{R}^n, |\cdot|)$, where $\alpha : \mathbb{R}^n \to \mathbb{R}$ is smooth. Let h, f be two positive functions on $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n} h\,d\mu = \int_{\mathbb{R}^n} f\,d\mu = 1$. Under some smooth conditions on h and f (see [15, 45] or p. 561 in [59]), there exists a smooth convex function $\Phi : \mathbb{R}^n \to \mathbb{R}$ such that $\nabla\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism which pushes $h\mu$ forwards to $f\mu$: $(\nabla\Phi)_\#(h\mu) = f\mu$ and

$$W_2^2(h\mu, f\mu) = \int_{\mathbb{R}^n} |x - \nabla\Phi(x)|^2\, h(x) d\mu(x), \tag{5.3.3}$$

where $W_2(h\mu, f\mu)$ denotes the $2-$Wasserstein distance for the Euclidian norm between the probability measures $h\mu$ and $f\mu$, which is defined by

$$W_2^2(h\mu, f\mu) = \inf_{\Pi \in C(h\mu, f\mu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\, d\Pi(x,y),$$

the set $C(h\mu, f\mu)$ being the totality of probability measures on the product space $\mathbb{R}^n \times \mathbb{R}^n$ such that $h\mu$ and $f\mu$ are marginals.
By formula of change of variables (proved by McCann in [50]), $\nabla\Phi$ satisfies a.e. the following equation

$$f(\nabla\Phi)e^{-\alpha(\nabla\Phi)}\det(\nabla^2\Phi) = he^{-\alpha}. \tag{5.3.4}$$

Now consider two couples of positive functions (h_1, f_1) and (h_2, f_2) satisfying same conditions as (h, f). Let Φ_1 and Φ_2 be the associated optimal maps, namely

$$(\nabla\Phi_1)_\# : h_1\mu \longrightarrow f_1\mu,$$
$$(\nabla\Phi_2)_\# : h_2\mu \longrightarrow f_2\mu.$$

Then we have

$$f_1(\nabla\Phi_1)e^{-\alpha(\nabla\Phi_1)}\mathrm{det}(\nabla^2\Phi_1) = h_1 e^{-\alpha}, \tag{5.3.5}$$

$$f_2(\nabla\Phi_2)e^{-\alpha(\nabla\Phi_2)}\mathrm{det}(\nabla^2\Phi_2) = h_2 e^{-\alpha}. \tag{5.3.6}$$

Let S_2 be the inverse map of $\nabla\Phi_2$, that is, $\nabla\Phi_2(S_2(x)) = x$ on $\mathbb{R}^n$; then we have

$$\nabla^2\Phi_2(S_2(x))\,\nabla S_2(x) = \mathrm{Id}, \text{ or } \quad \nabla S_2(x) = (\nabla^2\Phi_2)^{-1}(S_2(x)).$$

Acting on the right by S_2 the two hand sides of (5.3.5), as well as of (5.3.6), we get

$$f_1(\nabla\Phi_1(S_2))e^{-\alpha(\nabla\Phi_1(S_2))}\mathrm{det}(\nabla^2\Phi_1(S_2)) = h_1(S_2)e^{-\alpha(S_2)}, \tag{5.3.7}$$

$$f_2\, e^{-\alpha}\,\mathrm{det}(\nabla^2\Phi_2(S_2)) = h_2(S_2)e^{-\alpha(S_2)}. \tag{5.3.8}$$

It follows that

$$\frac{f_1}{f_2}\cdot\frac{f_1(\nabla\Phi_1(S_2))e^{-\alpha(\nabla\Phi_1(S_2))}}{f_1 e^{-\alpha}}\cdot\mathrm{det}\Big[(\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big](S_2) = \frac{h_1(S_2)}{h_2(S_2)}.$$

Taking the logarithm on the two sides yields

$$\begin{aligned}\log(\frac{f_1}{f_2}) + \log(f_1e^{-\alpha})(\nabla\Phi_1(S_2)) - \log(f_1e^{-\alpha})\\ + \log\mathrm{det}\Big[(\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big](S_2) = \log(\frac{h_1}{h_2})(S_2).\end{aligned} \tag{5.3.9}$$

Integrating the two sides of (5.3.9) with respect to the measure $f_2\mu$, we get

$$\begin{aligned}\int_{\mathbb{R}^n}\log(\frac{h_1}{h_2})(S_2)\,f_2d\mu - \int_{\mathbb{R}^n}\log(\frac{f_1}{f_2})\,f_2d\mu = \int_{\mathbb{R}^n}\log\mathrm{det}\Big[(\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big](S_2)\,f_2d\mu\\ + \int_{\mathbb{R}^n}\Big[\log(f_1e^{-\alpha})(\nabla\Phi_1(S_2)) - \log(f_1e^{-\alpha})\Big]\,f_2d\mu.\end{aligned} \tag{5.3.10}$$

By Taylor formula up to order 2,

$$\begin{aligned}\log(f_1e^{-\alpha})(\nabla\Phi_1(S_2)) - \log(f_1e^{-\alpha}) = \langle\nabla\log(f_1e^{-\alpha}), \nabla\Phi_1(S_2(x)) - x\rangle\\ + \int_0^1(1-t)\Big[\nabla^2\log(f_1e^{-\alpha})((1-t)x + t\nabla\Phi_1(S_2(x))\Big]\cdot(\nabla\Phi_1(S_2(x)) - x)^2\,dt.\end{aligned} \tag{5.3.11}$$

We have

$$\int_{\mathbb{R}^n} \langle \nabla \log(f_1 e^{-\alpha}), \nabla\Phi_1(S_2(x)) - x\rangle\, f_2 d\mu$$
$$= \int_{\mathbb{R}^n} \langle \nabla(f_1 e^{-\alpha}), \nabla\Phi_1(S_2(x)) - x\rangle\, \frac{f_2}{f_1}\, dx.$$

By integration by parts, this last term goes to

$$-\int_{\mathbb{R}^n} f_1 e^{-\alpha}\,\mathrm{div}\Big(\nabla\Phi_1(S_2(x)) - x\Big)\,\frac{f_2}{f_1}\,dx - \int_{\mathbb{R}^n} f_1 e^{-\alpha}\langle \nabla\Phi_1(S_2(x)) - x, \nabla(\frac{f_2}{f_1})\rangle\,dx$$
$$= -\int_{\mathbb{R}^n} \mathrm{div}\Big(\nabla\Phi_1(S_2(x)) - x\Big)\, f_2 d\mu - \int_{\mathbb{R}^n} \langle \nabla\Phi_1(S_2(x)) - x, \nabla(\log\frac{f_2}{f_1})\rangle\, f_2 d\mu.$$

Note that $\nabla\Big[(\nabla\Phi_1)(S_2)\Big] = \nabla^2\Phi_1(S_2)\,\nabla S_2 = \nabla^2\Phi_1(S_2)\cdot(\nabla^2\Phi_2)^{-1}(S_2)$, and

$$\mathrm{div}\Big(\nabla\Phi_1(S_2(x)) - x\Big) = \mathrm{Trace}\Big[\nabla^2\Phi_1(S_2)\cdot(\nabla^2\Phi_2)^{-1}(S_2) - \mathrm{Id}\Big].$$

Combining above computations yields

$$\begin{aligned}&\int_{\mathbb{R}^n} \langle \nabla \log(f_1 e^{-\alpha}), \nabla\Phi_1(S_2(x)) - x\rangle\, f_2 d\mu\\ &= -\int_{\mathbb{R}^n} \mathrm{Trace}\Big[\nabla^2\Phi_1(S_2)\cdot(\nabla^2\Phi_2)^{-1}(S_2) - \mathrm{Id}\Big]\, f_2 d\mu\\ &\quad - \int_{\mathbb{R}^n} \langle \nabla\Phi_1(S_2(x)) - x, \nabla(\log\frac{f_2}{f_1})\rangle\, f_2 d\mu.\end{aligned} \tag{5.3.12}$$

For a matrix A on $\mathbb{R}^n$, the Fredholm-Carleman determinant $\det_2(A)$ is defined by

$$\det_2(A) = e^{\mathrm{Trace}(\mathrm{Id}-A)}\det(A).$$

It is easy to check that if A is symmetric positive, then $0 \le \det_2(A) \le 1$. We have

$$\mathrm{Trace}\Big((\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big) = \mathrm{Trace}\Big((\nabla^2\Phi_2)^{-1/2}\,\nabla^2\Phi_1\,(\nabla^2\Phi_2)^{-1/2}\Big),$$

and

$$\det\Big((\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big) = \det\Big((\nabla^2\Phi_2)^{-1/2}\,\nabla^2\Phi_1\,(\nabla^2\Phi_2)^{-1/2}\Big).$$

Therefore

$$\log\det_2\Big((\nabla^2\Phi_1)(\nabla^2\Phi_2)^{-1}\Big) = \log\det_2\Big((\nabla^2\Phi_2)^{-1/2}\,\nabla^2\Phi_1\,(\nabla^2\Phi_2)^{-1/2}\Big) \le 0. \tag{5.3.13}$$

Now combining (5.3.10), (5.3.11) and (5.3.12), we get the following result.

Theorem 5.3.1. *Let $\alpha \in C^\infty(\mathbb{R}^n)$ and $d\mu = e^{-\alpha}dx$ be a probability measure on $\mathbb{R}^n$. Then*

$$\begin{aligned}
&\mathrm{Ent}_{h_1\mu}(\frac{h_2}{h_1}) - \mathrm{Ent}_{f_1\mu}(\frac{f_2}{f_1}) = \int_{\mathbb{R}^n} \langle \nabla\Phi_1 - \nabla\Phi_2, \nabla(\log\frac{f_2}{f_1})(\nabla\Phi_2)\rangle\, h_2 d\mu \\
&- \int_{\mathbb{R}^n} \log\det{}_2\Big((\nabla^2\Phi_2)^{-1/2}\,\nabla^2\Phi_1\,(\nabla^2\Phi_2)^{-1/2}\Big)\, h_2 d\mu \\
&+ \int_0^1 (1-t)dt \int_{\mathbb{R}^n} \Big[-\nabla^2\log(f_1 e^{-\alpha})((1-t)\nabla\Phi_2 + t\nabla\Phi_1)\Big]\cdot(\nabla\Phi_1 - \nabla\Phi_2)^2\, h_2 d\mu.
\end{aligned} \tag{5.3.14}$$

Corollary 5.3.2. *Suppose that*

$$\nabla^2\big(-\log(f_1 e^{-\alpha})\big) \geq c\,\mathrm{Id}, \quad c > 0. \tag{5.3.15}$$

Then

$$\begin{aligned}
\int_{\mathbb{R}^n} |\nabla\Phi_1 - \nabla\Phi_2|^2\, h_2 d\mu \leq \frac{4}{c}\Big(\mathrm{Ent}_{h_1\mu}(\frac{h_2}{h_1}) - \mathrm{Ent}_{f_1\mu}(\frac{f_2}{f_1})\Big) \\
+ \frac{4}{c^2}\int_{\mathbb{R}^n} |\nabla\log\frac{f_2}{f_1}|^2\, f_2 d\mu.
\end{aligned} \tag{5.3.16}$$

If moreover $f_1 = f_2$, then it holds more precisely

$$\frac{c}{2}\int_{\mathbb{R}^n} |\nabla\Phi_1 - \nabla\Phi_2|^2\, h_2 d\mu \leq \mathrm{Ent}_{h_1\mu}(\frac{h_2}{h_1}).$$

Proof. Note that

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^n} \langle \nabla\Phi_1 - \nabla\Phi_2, \nabla(\log\frac{f_2}{f_1})(\nabla\Phi_2)\rangle\, h_2 d\mu\Big| \\
&\leq \Big(\int_{\mathbb{R}^n} |\nabla\Phi_1 - \nabla\Phi_2|^2\, h_2 d\mu\Big)^{1/2}\Big(\int_{\mathbb{R}^n} |\nabla\log\frac{f_2}{f_1}|^2\, f_2 d\mu\Big)^{1/2} \\
&\leq \frac{c}{4}\int_{\mathbb{R}^n} |\nabla\Phi_1 - \nabla\Phi_2|^2\, h_2 d\mu + \frac{1}{c}\int_{\mathbb{R}^n} |\nabla\log\frac{f_2}{f_1}|^2\, f_2 d\mu.
\end{aligned}$$

Under condition (5.3.15), the last term in (5.3.14) is bounded from below by

$$\frac{c}{2}\int_{\mathbb{R}^n} |\nabla\Phi_1 - \nabla\Phi_2|^2\, h_2 d\mu.$$

Now according to (5.3.14), we get the result from (5.3.16). □

Here are some technical lemmas.

Lemma 5.3.3. *Let A be a symmetric positive definite matrix and B be a symmetric matrix on $\mathbb{R}^n$; then*

$$||A^{-1/2}BA^{-1/2}||_{HS} \geq \frac{||B||_{HS}}{||A||_{op}}, \tag{5.3.17}$$

where $||\cdot||_{op}$ denotes the norm of matrices.

Proof. Let $C = A^{-1/2}BA^{-1/2}$, then $C = A^{1/2}BA^{1/2}$. Let $\{e_1, \cdots, e_n\}$ be an orthonormal basis of $\mathbb{R}^n$, of eigenvalues of A: $A^{1/2}e_i = \sqrt{\lambda_i}\, e_i$. We have $Be_i = \sqrt{\lambda_i}\, A^{1/2}Ce_i$ and

$$|Be_i|^2 \leq \max(\lambda_i)\, |A^{1/2}Ce_i|^2 = \max(\lambda_i)\, \langle Ce_i, ACe_i\rangle \leq ||A||_{op}^2\, |Ce_i|^2.$$

It follows that $||B||_{HS}^2 \leq ||A||_{op}^2\, ||C||_{HS}^2$. The result (5.3.17) follows. ☐

Lemma 5.3.4. *Let A, B be symmetric matrices such that $I + A$ and $I + B$ are positive definite. Then*

$$\begin{aligned} &-\log \det{}_2\Big((I+A)(I+B)^{-1}\Big) \\ &= \int_0^1 (1-t)||(I+(1-t)B+tA)^{-1/2}(A-B)(I+(1-t)B+tA)^{-1/2}||_{HS}^2\, dt. \end{aligned} \tag{5.3.18}$$

Proof. Note first $I - (I+A)(I+B)^{-1} = (B-A)(I+B)^{-1}$ and

$$(i) \qquad \text{Trace}\Big[I - (I+A)(I+B)^{-1}\Big] = \langle B-A, (I+B)^{-1}\rangle_{HS}.$$

Let $\chi(t) = \log\det\Big(I + (1-t)B + tA\Big)$ for $t \in [0,1]$. We have

$$\chi'(t) = \text{Trace}\Big[(A-B)(I+(1-t)B+tA)^{-1}\Big] = \langle A-B, (I+(1-t)B+tA)^{-1}\rangle_{HS}.$$

Then

$$\log\det(I+A) - \log\det(I+B) = \langle A-B, \int_0^1 (I+(1-t)B+tA)^{-1}\, dt\rangle_{HS}.$$

According to above (i) and definition of $\det_2$, we get

$$\begin{aligned} -\log\det{}_2\Big((I+A)(I+B)^{-1}\Big) &= \langle A-B, \int_0^1 \Big[(I+B)^{-1} - (I+(1-t)B+tA)^{-1}\Big]\, dt\rangle_{HS} \\ &= \langle A-B, \int_0^1 \Big[\int_0^t (I+(1-s)B+sA)^{-1}\,(A-B)\,(I+(1-s)B+sA)^{-1} ds\Big] dt\rangle_{HS} \end{aligned}$$

which is equal to $\int_0^1 (1-t)\langle A-B, (I+(1-t)B+tA)^{-1}(A-B)(I+(1-t)B+tA)^{-1}\rangle_{HS}\,dt$, implying (5.3.18). $\square$

In what follows, we will consider the standard Gaussian measure γ as the reference measure on $\mathbb{R}^n$. Let e^{-V} and e^{-W} be two density functions with respect to γ, that is, $\int_{\mathbb{R}^n} e^{-V} d\gamma = \int_{\mathbb{R}^n} e^{-W} d\gamma = 1$. Let Φ be a smooth convex function such that $\nabla\Phi$ pushes $e^{-V}\gamma$ forward to $e^{-W}\gamma$, that is,

$$\int_{\mathbb{R}^n} F(\nabla\Phi)\, e^{-V} d\gamma = \int_{\mathbb{R}^n} F\, e^{-W} d\gamma.$$

Let $a \in \mathbb{R}^n$; then

$$\int_{\mathbb{R}^n} F(\nabla\Phi(x+a)) e^{-V(x+a)} e^{-\langle x,a\rangle - \frac{1}{2}|a|^2}\, d\gamma = \int_{\mathbb{R}^n} F(\nabla\Phi) e^{-V}\, d\gamma.$$

Denote by τ_a the translation by a, and $M_a(x) = e^{-\langle x,a\rangle - \frac{1}{2}|a|^2}$, then the above relations imply that

$$\nabla(\tau_a\Phi)_\# : e^{-\tau_a V} M_a \gamma \to e^{-W}\gamma.$$

Let $h_1 = e^{-\tau_a V} M_a, h_2 = e^{-V}$. Then $\mathrm{Ent}_{h_1\mu}\big(\frac{h_2}{h_1}\big) = \int_{\mathbb{R}^n} (\tau_a V - V + \langle x,a\rangle + \frac{1}{2}|a|^2) e^{-V} d\gamma$. Applying Theorem 5.3.1 , we get

$$\begin{aligned}
&\int_{\mathbb{R}^n} (\tau_a V - V + \langle x,a\rangle + \frac{1}{2}|a|^2) e^{-V} d\gamma \\
&= -\int_{\mathbb{R}^n} \log\det{}_2\Big[(\nabla^2\Phi)^{-1/2}\,\nabla^2(\tau_a\Phi)\,(\nabla^2\Phi)^{-1/2}\Big]\, e^{-V} d\gamma \\
&\quad + \int_0^1 (1-t)dt \int_{\mathbb{R}^n} \Big[(\mathrm{Id} + \nabla^2 W)(\Lambda(t,x,a))\Big]\cdot(\nabla\Phi(x) - \nabla\Phi(x+a))^2 e^{-V} d\gamma,
\end{aligned}$$

where $\Lambda(t,x,a) = (1-t)\nabla\Phi(x) + t\nabla\Phi(x+a)$. Note that as $a \to 0$, $\Lambda(t,x,a) \to \nabla\Phi(x)$.

Replacing a by $-a$, and summing respectively the two hand sides of these equalities, we get

$$\begin{aligned}
&\int_{\mathbb{R}^n} \big(V(x+a) + V(x-a) - 2V(x) + |a|^2\big)\, e^{-V} d\gamma = J(a) + J(-a) \\
&+ \int_0^1 (1-t)dt \int_{\mathbb{R}^n} \Big[(\mathrm{Id} + \nabla^2 W)(\Lambda(t,x,a))\Big]\cdot(\nabla\Phi(x) - \nabla\Phi(x+a))^2 e^{-V} d\gamma \\
&+ \int_0^1 (1-t)dt \int_{\mathbb{R}^n} \Big[(\mathrm{Id} + \nabla^2 W)(\Lambda(t,x,-a))\Big]\cdot(\nabla\Phi(x) - \nabla\Phi(x-a))^2 e^{-V} d\gamma,
\end{aligned} \tag{5.3.19}$$

where

$$J(a) = -\int_{\mathbb{R}^n} \log \det{}_2 \Big[(\nabla^2\Phi)^{-1/2}\,\nabla^2(\tau_a\Phi)\,(\nabla^2\Phi)^{-1/2}\Big]\, e^{-V} d\gamma.$$

By explicit formula given by the Lemma 5.3.3, and write $\nabla\Phi(x) = x + \nabla\varphi(x)$, we have

$$\begin{aligned}\frac{1}{\varepsilon^2} J(\varepsilon a) = \int_0^1 (1-t)dt \int_{\mathbb{R}^n} &||(I + (1-t)\nabla^2\varphi + t\nabla^2\varphi(x+\varepsilon a))^{-1/2} \\ &\varepsilon^{-1}\Big(\nabla^2\varphi(x+\varepsilon a) - \nabla^2\varphi(x)\Big)(I + (1-t)\nabla^2\varphi + t\nabla^2\varphi(x+\varepsilon a))^{-1/2}||_{HS}^2 e^{-V} d\gamma.\end{aligned}$$

So that, by Fatou lemma

$$\underline{\lim}_{\varepsilon\to 0} \frac{J(\varepsilon a)}{\varepsilon^2} \geq \frac{1}{2}\int_{\mathbb{R}^n} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma. \quad (5.3.20)$$

Now replacing a by εa and dividing by ε^2 the two hand sides of (5.3.19), letting $\varepsilon \to 0$ yields

$$\begin{aligned}\int_{\mathbb{R}^n} \Big[D_a^2 V + |a|^2\Big]\, e^{-V} d\gamma &\geq \int_{\mathbb{R}^n} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma \\ &+ \int_{\mathbb{R}^n} (\mathrm{Id} + \nabla^2 W)(\nabla\Phi)\,(D_a\nabla\Phi, D_a\nabla\Phi)\, e^{-V} d\gamma \\ &= \int_{\mathbb{R}^n} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma \\ &+ \int_{\mathbb{R}^n} |D_a\nabla\Phi|^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} (\nabla^2 W)(\nabla\Phi)(D_a\nabla\Phi, D_a\nabla\Phi)\, e^{-V} d\gamma.\end{aligned} \quad (5.3.21)$$

By integration by parts,

$$\int_{\mathbb{R}^n} D_a^2 V\, e^{-V} d\gamma = \int_{\mathbb{R}^n} (D_a V)^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} D_a V\, \langle a, x\rangle\, e^{-V} d\gamma.$$

Using (5.3.21) and $|D_a\nabla\Phi|^2 = |a|^2 + 2\langle a, D_a\nabla\varphi\rangle + |D_a\nabla\varphi|^2$, we get

$$\begin{aligned}&\int_{\mathbb{R}^n} (D_a V)^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} D_a V\, \langle a, x\rangle\, e^{-V} d\gamma \\ &\geq \int_{\mathbb{R}^n} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma \\ &+ 2\int_{\mathbb{R}^n} \langle a, D_a\nabla\varphi\rangle\, e^{-V} d\gamma + \int_{\mathbb{R}^n} |D_a\nabla\varphi|^2\, e^{-V} d\gamma + \int_{\mathbb{R}^n} \nabla^2 W(\nabla\Phi)(D_a\nabla\Phi, D_a\nabla\Phi)\, e^{-V} d\gamma.\end{aligned}$$

Summing a on an orthonormal basis $\mathcal{B}$, it follows

$$\begin{aligned}
&\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} \langle x, \nabla V\rangle\, e^{-V} d\gamma \\
\geq & \int_{\mathbb{R}^n} \sum_{a\in\mathcal{B}} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma \\
+ 2 & \int_{\mathbb{R}^n} \Delta\varphi\, e^{-V} d\gamma + \int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2 e^{-V} d\gamma + \sum_{a\in\mathcal{B}} \int_{\mathbb{R}^n} \nabla^2 W(\nabla\Phi)(D_a\nabla\Phi, D_a\nabla\Phi)\, e^{-V} d\gamma.
\end{aligned} \tag{5.3.22}$$

Let

$$N_W(\nabla^2\varphi) = \sum_{a\in\mathcal{B}} \nabla^2 W_{\nabla\Phi}(D_a\nabla\varphi, D_a\nabla\varphi). \tag{5.3.23}$$

Then

$$\begin{aligned}
&\sum_{a\in\mathcal{B}} \int_{\mathbb{R}^n} \nabla^2 W_{\nabla\Phi}(D_a\nabla\Phi, D_a\nabla\Phi)\, e^{-V} d\gamma \\
= & \int_{\mathbb{R}^n} (\Delta W)(\nabla\Phi)\, e^{-V} d\gamma + 2\int_{\mathbb{R}^n} \langle \nabla^2 W(\nabla\Phi), \nabla^2\varphi\rangle_{HS}\, e^{-V} d\gamma + \int_{\mathbb{R}^n} N_W(\nabla^2\varphi)\, e^{-V} d\gamma.
\end{aligned}$$

This equality, together with (5.3.22) yield

$$\begin{aligned}
&\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} \langle x, \nabla V\rangle\, e^{-V} d\gamma \\
\geq & \int_{\mathbb{R}^n} \sum_{a\in\mathcal{B}} ||(I+\nabla^2\varphi)^{-1/2}\, D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\, e^{-V} d\gamma \\
+ 2 & \int_{\mathbb{R}^n} \Delta\varphi\, e^{-V} d\gamma + \int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2 e^{-V} d\gamma + \int_{\mathbb{R}^n} (\Delta W)(\nabla\Phi)\, e^{-V} d\gamma \\
+ 2 & \int_{\mathbb{R}^n} \langle \nabla^2 W(\nabla\Phi), \nabla^2\varphi\rangle_{HS}\, e^{-V} d\gamma + \int_{\mathbb{R}^n} N_W(\nabla^2\varphi)\, e^{-V} d\gamma.
\end{aligned} \tag{5.3.24}$$

In order to obtain desired terms, we first use the relation

$$\int_{\mathbb{R}^n} |x + \nabla\varphi(x)|^2\, e^{-V} d\gamma = \int_{\mathbb{R}^n} |x|^2\, e^{-W} d\gamma$$

which gives that

$$2\int_{\mathbb{R}^n} \langle x, \nabla\varphi(x)\rangle\, e^{-V} d\gamma = \int_{\mathbb{R}^n} |x|^2\, e^{-W} d\gamma - \int_{\mathbb{R}^n} |x|^2\, e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla\varphi(x)|^2\, e^{-V} d\gamma.$$

Let $\mathcal{L}$ be the Ornstein-Uhlenbeck operator: $\mathcal{L}f(x) = \Delta f(x) - \langle x, \nabla f\rangle$. Remark that

$$\mathcal{L}(\frac{1}{2}|x|^2) = d - |x|^2.$$

Then $\int_{\mathbb{R}^n}|x|^2\,e^{-W}d\gamma-\int_{\mathbb{R}^n}|x|^2\,e^{-V}d\gamma=-\int_{\mathbb{R}^n}\mathcal{L}(\frac{1}{2}|x|^2)e^{-W}d\gamma+\int_{\mathbb{R}^n}\mathcal{L}(\frac{1}{2}|x|^2)e^{-V}d\gamma$, which is equal to

$$-\int_{\mathbb{R}^n}\langle x,\nabla W\rangle\,e^{-W}d\gamma+\int_{\mathbb{R}^n}\langle x,\nabla V\rangle\,e^{-V}d\gamma.$$

Therefore

$$\begin{aligned}2\int_{\mathbb{R}^n}\langle x,\nabla\varphi(x)\rangle\,e^{-V}d\gamma=&-\int_{\mathbb{R}^n}\langle x,\nabla W\rangle\,e^{-W}d\gamma\\&+\int_{\mathbb{R}^n}\langle x,\nabla V\rangle\,e^{-V}d\gamma-\int_{\mathbb{R}^n}|\nabla\varphi|^2\,e^{-V}d\gamma.\end{aligned}\tag{5.3.25}$$

On the other hand, from Monge-Ampère equation,

$$e^{-V}=e^{-W(\nabla\Phi)}e^{\mathcal{L}\varphi-\frac{1}{2}|\nabla\varphi|^2}\det{}_2(\mathrm{Id}+\nabla^2\varphi),$$

we have

$$-V=-W(\nabla\Phi)+\mathcal{L}\varphi-\frac{1}{2}|\nabla\varphi|^2+\log\det{}_2(\mathrm{Id}+\nabla^2\varphi).$$

Integrating the two hand sides with respect to $e^{-V}d\gamma$, we get

$$\begin{aligned}\int_{\mathbb{R}^n}\mathcal{L}\varphi\,e^{-V}d\gamma=&\mathrm{Ent}_\gamma(e^{-V})-\mathrm{Ent}_\gamma(e^{-W})+\frac{1}{2}\int_{\mathbb{R}^n}|\nabla\varphi|^2\,e^{-V}d\gamma\\&-\int_{\mathbb{R}^n}\log\det{}_2(\mathrm{Id}+\nabla^2\varphi)\,e^{-V}d\gamma.\end{aligned}\tag{5.3.26}$$

Combining (5.3.25) and (5.3.26), we get

$$\begin{aligned}2\int_{\mathbb{R}^n}\Delta\varphi\,e^{-V}d\gamma&=2\int_{\mathbb{R}^n}\mathcal{L}\varphi\,e^{-V}d\gamma+2\int_{\mathbb{R}^n}\langle x,\nabla\varphi\rangle\,e^{-V}d\gamma\\&=2\mathrm{Ent}_\gamma(e^{-V})-2\mathrm{Ent}_\gamma(e^{-W})-2\int_{\mathbb{R}^n}\log\det{}_2(\mathrm{Id}+\nabla^2\varphi)\,e^{-V}d\gamma\\&\quad-\int_{\mathbb{R}^n}\langle x,\nabla W\rangle\,e^{-W}d\gamma+\int_{\mathbb{R}^n}\langle x,\nabla V\rangle\,e^{-V}d\gamma.\end{aligned}$$

Replacing $\int_{\mathbb{R}^n}\Delta\varphi\,e^{-V}d\gamma$ in (5.3.24) by above expression, we obtain

$$\begin{aligned}\int_{\mathbb{R}^n}|\nabla V|^2\,e^{-V}d\gamma\geq&\,2\mathrm{Ent}_\gamma(e^{-V})-2\mathrm{Ent}_\gamma(e^{-W})-2\int_{\mathbb{R}^n}\log\det{}_2(\mathrm{Id}+\nabla^2\varphi)\,e^{-V}d\gamma\\&+\int_{\mathbb{R}^n}\sum_{a\in\mathcal{B}}||(I+\nabla^2\varphi)^{-1/2}\,D_a\nabla^2\varphi(x)\,(I+\nabla^2\varphi)^{-1/2}||_{HS}^2\,e^{-V}d\gamma+\int_{\mathbb{R}^n}||\nabla^2\varphi||_{HS}^2\,e^{-V}d\gamma\\&+\int_{\mathbb{R}^n}\mathcal{L}W\,e^{-W}d\gamma+2\int_{\mathbb{R}^n}\langle\nabla^2W(\nabla\Phi),\nabla^2\varphi\rangle_{HS}\,e^{-V}d\gamma+\int_{\mathbb{R}^n}N_W(\nabla^2\varphi)\,e^{-V}d\gamma.\end{aligned}$$

So we get

Theorem 5.3.5. *We have*

$$\int_{\mathbb{R}^n} |\nabla V|^2 \, e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla W|^2 \, e^{-W} d\gamma$$
$$\geq 2\mathrm{Ent}_\gamma(e^{-V}) - 2\mathrm{Ent}_\gamma(e^{-W}) - 2\int_{\mathbb{R}^n} \log\det{}_2(\mathrm{Id} + \nabla^2\varphi) \, e^{-V} d\gamma$$
$$+ \int_{\mathbb{R}^n} \sum_{a\in\mathcal{B}} ||(I+\nabla^2\varphi)^{-1/2} \, D_a\nabla^2\varphi(x) \, (I+\nabla^2\varphi)^{-1/2}||_{HS}^2 \, e^{-V} d\gamma + \int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2 \, e^{-V} d\gamma$$
$$+ 2\int_{\mathbb{R}^n} \langle \nabla^2 W(\nabla\Phi), \nabla^2\varphi\rangle_{HS} \, e^{-V} d\gamma + \int_{\mathbb{R}^n} N_W(\nabla^2\varphi) \, e^{-V} d\gamma.$$

Theorem 5.3.6. *Assume that $\nabla^2 W \geq -c\,\mathrm{Id}$ with $c \in [0,1[$; then*

$$\begin{aligned}&\int_{\mathbb{R}^n} |\nabla V|^2 \, e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla W|^2 \, e^{-W} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||_{HS}^2 e^{-W} d\gamma \\ &\geq 2\mathrm{Ent}_\gamma(e^{-V}) - 2\mathrm{Ent}_\gamma(e^{-W}) + \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2 \, e^{-V} d\gamma.\end{aligned} \tag{5.3.27}$$

Proof. It is sufficient to notice that

$$2\int_{\mathbb{R}^n} |\langle \nabla^2 W(\nabla\Phi), \nabla^2\varphi\rangle_{HS}| \, e^{-V} d\gamma \leq \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi||_{HS}^2 \, e^{-V} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||_{HS}^2 \, e^{-W} d\gamma.$$

The inequality (5.3.27) follows from Theorem 5.3.5. □

Theorem 5.3.7. *Let $1 \leq p < 2$. Denote by $||\cdot||_{op}$ the norm of operator, then*

$$||\nabla^3\varphi||_{L^p(e^{-V}\gamma)}^2 \leq \Big\| ||I+\nabla^2\varphi||_{op} \Big\|_{L^{\frac{2p}{2-p}}(e^{-V}\gamma)}^2 \Big(||\nabla V||_{L^2(e^{-V}\gamma)}^2 + \frac{2}{1-c}||\nabla^2 W||_{L^2(e^{-W}\gamma)}^2 \Big). \tag{5.3.28}$$

Proof. By Hölder inequality

$$\int_{\mathbb{R}^n} ||\nabla^3\varphi||_{HS}^p \, e^{-V} d\gamma \leq \Big(\int_{\mathbb{R}^n} \frac{||\nabla^3\varphi||_{HS}^2}{||I+\nabla^2\varphi||_{op}^2} \, e^{-V} d\gamma \Big)^{p/2} \Big(\int_{\mathbb{R}^n} ||I+\nabla^2\varphi||_{op}^{\frac{2p}{2-p}} \, e^{-V} d\gamma \Big)^{\frac{2-p}{2}}.$$

By (5.3.17),

$$\frac{||\nabla^3\varphi||_{HS}^2}{||I+\nabla^2\varphi||_{op}^2} \leq \sum_{a\in\mathcal{B}} ||(I+\nabla^2\varphi)^{-1/2} \, D_a\nabla^2\varphi(x) \, (I+\nabla^2\varphi)^{-1/2}||_{HS}^2.$$

Remark that $\int_{\mathbb{R}^n} |\nabla W|^2 e^{-W} d\gamma \geq 2\mathrm{Ent}_\gamma(e^{-W})$. Now by Theorem 5.3.5, we get the result. □

In what follows, we will compute the variation of optimal transport maps in Sobolev spaces. Consider

$$(\nabla\Phi_1)_\#: e^{-V_1}d\gamma \to e^{-W_1}d\gamma, \quad (\nabla\Phi_2)_\#: e^{-V_2}d\gamma \to e^{-W_2}d\gamma.$$

We will explore the term $-\log\det_2\Big[(\nabla^2\Phi_2)^{-1/2}\nabla^2\Phi_1(\nabla^2\Phi_2)^{-1/2}\Big]$ in Theorem 5.3.1. Let $\nabla\Phi_1(x) = x + \nabla\varphi_1(x)$ and $\nabla\Phi_2(x) = x + \nabla\varphi_2(x)$; then

$$\nabla^2\Phi_1 = I + \nabla^2\varphi_1, \quad \nabla^2\Phi_2 = I + \nabla^2\varphi_2.$$

Theorem 5.3.8. *Let* $1 \le p < 2$ *and*

$$M(\nabla^2\varphi_1, \nabla^2\varphi_2) = \max\Big(\Big\| ||I + \nabla^2\varphi_1||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)}, \Big\| ||I + \nabla^2\varphi_2||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)}\Big). \tag{5.3.29}$$

Assume that $\nabla^2 W_1 \ge -c\,\mathrm{Id}$ *with* $c \in [0,1[$. *Then we have*

$$\begin{aligned} ||\nabla^2\varphi_1 - \nabla^2\varphi_2||^2_{L^p(e^{-V_2}\gamma)} \le & 2M(\nabla^2\varphi_1, \nabla^2\varphi_2)\Big[2\int_{\mathbb{R}^n}(V_1 - V_2)e^{-V_2}d\gamma \\ & + \frac{2}{1-c}\int_{\mathbb{R}^n}|\nabla(W_1 - W_2)|^2 e^{-W_2}d\gamma\Big]. \end{aligned} \tag{5.3.30}$$

Proof. Applying Lemma 5.3.3 to $B = \nabla^2\varphi_1 - \nabla^2\varphi_2$ and $A = I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1$ yields

$$\begin{aligned} &||(I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1)^{-1/2}(\nabla^2\varphi_1 - \nabla^2\varphi_2)(I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1)^{-1/2}||^2_{HS} \\ &\qquad \ge \frac{||\nabla^2\varphi_1 - \nabla^2\varphi_2||^2_{HS}}{||I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1||^2_{op}}. \end{aligned}$$

As above, by Hölder inequality, we have

$$\int_{\mathbb{R}^n}\frac{||\nabla^2\varphi_1 - \nabla^2\varphi_2||^2_{HS}}{||I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1||^2_{op}}\,e^{-V_2}d\gamma \ge \frac{||\nabla^2\varphi_1 - \nabla^2\varphi_2||^2_{L^p(e^{-V_2}\gamma)}}{\Big\| ||I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)}}.$$

Now by convexity,

$$\begin{aligned} &\Big\| ||I + (1-t)\nabla^2\varphi_2 + t\nabla^2\varphi_1||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)} \\ \le (1-t)&\Big\| ||I + \nabla^2\varphi_2||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)} + t\Big\| ||I + \nabla^2\varphi_1||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)} \le M(\nabla^2\varphi_1, \nabla^2\varphi_2). \end{aligned}$$

According to Lemma 5.3.4, we have

$$\begin{aligned}
&\int_{\mathbb{R}^n} -\log\det{}_2\Big((\nabla^2\Phi_2)^{-1/2}\,\nabla^2\Phi_1\,(\nabla^2\Phi_2)^{-1/2}\Big)\,e^{-V_2}d\gamma \\
&\geq \int_0^1 (1-t)dt \int_{\mathbb{R}^n} \frac{||\nabla^2\varphi_1-\nabla^2\varphi_2||_{HS}^2}{||I+(1-t)\nabla^2\varphi_2+t\nabla^2\varphi_1||_{op}^2}\,e^{-V_2}d\gamma \\
&\geq \frac{1}{2}\frac{||\nabla^2\varphi_1-\nabla^2\varphi_2||_{L^p(e^{-V_2}\gamma)}^2}{M(\nabla^2\varphi_1,\nabla^2\varphi_2)}.
\end{aligned} \tag{5.3.31}$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^n} \langle\nabla\Phi_1-\nabla\Phi_2,\nabla(W_1-W_2)(\nabla\Phi_2)\rangle\,e^{-V_2}d\gamma\Big| \\
&\leq \Big(\int_{\mathbb{R}^n}|\nabla\Phi_1-\nabla\Phi_2|^2\,e^{-V_2}d\gamma\Big)^{1/2}\Big(\int_{\mathbb{R}^n}|\nabla(W_1-W_2)|^2\,e^{-W_2}d\gamma\Big)^{1/2} \\
&\leq \frac{1-c}{4}\int_{\mathbb{R}^n}|\nabla\Phi_1-\nabla\Phi_2|^2\,e^{-V_2}d\gamma+\frac{1}{1-c}\int_{\mathbb{R}^n}|\nabla(W_1-W_2)|^2\,e^{-W_2}d\gamma.
\end{aligned}$$

Under the hypothesis $\nabla^2 W_1 \geq -c\mathrm{Id}$ with $c<1$, the inequality (5.3.16) implies

$$\int_{\mathbb{R}^n}|\nabla\Phi_1-\nabla\Phi_2|^2\,e^{-V_2}d\gamma \leq \frac{4}{1-c}\int_{\mathbb{R}^n}(V_1-V_2)e^{-V_2}d\gamma+\frac{4}{(1-c)^2}\int_{\mathbb{R}^n}|\nabla(W_1-W_2)|^2e^{-W_2}d\gamma,$$

so that

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^n} \langle\nabla\Phi_1-\nabla\Phi_2,\nabla(W_1-W_2)(\nabla\Phi_2)\rangle\,e^{-V_2}d\gamma\Big| \\
&\leq \int_{\mathbb{R}^n}(V_1-V_2)e^{-V_2}d\gamma+\frac{2}{1-c}\int_{\mathbb{R}^n}|\nabla(W_1-W_2)|^2e^{-W_2}d\gamma.
\end{aligned}$$

Now combinig (5.3.14) and (5.3.31), we conclude (5.3.30). □

5.3.2 Extension to Sobolev spaces

In this subsection, we will assume that $V\in\mathbb{D}_1^2(\mathbb{R}^n,\gamma), W\in\mathbb{D}_2^2(\mathbb{R}^n,\gamma)$ and there exist constants $\delta_2>0$ and $c\in[0,1[$ such that

$$e^{-V}\leq\delta_2,\quad e^{-W}\leq\delta_2\quad \text{and } \nabla^2W\geq -c\,\mathrm{Id}. \tag{5.3.32}$$

It turns out that V and W are bounded from below. Consider the Ornstein-Uhlenbeck semi-group P_ε

$$P_\varepsilon f(x)=\int_{\mathbb{R}^n} f(e^{-\varepsilon}x+\sqrt{1-e^{2\varepsilon}}\,y)\,d\gamma(y).$$

If $f \in \mathbb{D}_2^2(\mathbb{R}^n, \gamma)$, then

$$\nabla P_\varepsilon f(x) = e^{-\varepsilon} \int_{\mathbb{R}^n} \nabla f(e^{-\varepsilon} x + \sqrt{1-e^{2\varepsilon}}\, y)\, d\gamma(y),$$

and

$$\nabla^2 P_\varepsilon f(x) = e^{-2\varepsilon} \int_{\mathbb{R}^n} \nabla^2 f(e^{-\varepsilon} x + \sqrt{1-e^{2\varepsilon}}\, y)\, d\gamma(y).$$

It follows that $||\nabla P_\varepsilon f||_{L^2(\gamma)} \le ||\nabla f||_{L^2(\gamma)}$ and $||\nabla^2 P_\varepsilon f||_{L^2(\gamma)} \le ||\nabla^2 f||_{L^2(\gamma)}$ and

$$\lim_{\varepsilon \to 0} ||P_\varepsilon f - f||_{\mathbb{D}_2^2(\gamma)} = 0. \tag{5.3.33}$$

Now we use P_ε to regularize V and W. Let

$$V_m = \chi_m\, P_{\frac{1}{m}} V + \log \int_{\mathbb{R}^n} e^{-\chi_m P_{\frac{1}{m}} V} d\gamma\,, \quad W_m = P_{\frac{1}{m}} W + \log \int_{\mathbb{R}^n} e^{-P_{\frac{1}{m}} W}\, d\gamma,$$

where $\chi_m \in C_c^\infty(\mathbb{R}^n)$ is a smooth function with compact support satisfying usual conditions: $0 \le \chi_m \le 1$ and

$$\chi_m(x) = 1 \text{ if } |x| \le m, \quad \chi_m(x) = 0 \text{ if } |x| \ge m+2, \quad \sup_{m \ge 1} ||\nabla \chi_m||_\infty \le 1.$$

Then the functions V_m, W_m satisfy conditions in (5.3.32) with $2\delta_2$ for n big enough, and ∇V_m converges to ∇V in $L^2(\gamma)$. In fact,

$$\nabla V_m - \nabla V = \nabla \chi_m P_{\frac{1}{m}} V + \chi_m \left(\nabla P_{\frac{1}{m}} V - \nabla V\right) + \nabla V\, (\chi_m - 1).$$

It is only to check that $\lim_{m \to +\infty} \int_{\mathbb{R}^n} |\nabla \chi_m|^2 P_{\frac{1}{m}} |V|^2\, d\gamma = 0$. But

$$(*) \qquad \int_{\mathbb{R}^n} |\nabla \chi_m|^2 P_{\frac{1}{m}} |V|^2\, d\gamma = \int_{\mathbb{R}^n} |V|^2\, P_{\frac{1}{m}} |\nabla \chi_m|^2\, d\gamma.$$

For $x \in \mathbb{R}^n$ fixed, let $r_m(x) = \dfrac{m - (1-e^{-1/m})|x|}{\sqrt{1-e^{-2/m}}}$, then

$$P_{\frac{1}{m}} |\nabla \chi_m|^2(x) \le \int_{\mathbb{R}^n} \mathbf{1}_{\{|e^{-1/m} x + \sqrt{1-e^{-2/m}} y| \ge m\}} d\gamma(y) \le \gamma(|y| \ge r_m(x)) \to 0,$$

as $m \to +\infty$. Now dominated Lebesgue convergence theorem, together with above $(*)$ yield the result.

Let $x \to x + \nabla\varphi_m(x)$ be the optimal transport map which pushes $e^{-V_m}\gamma$ forward to $e^{-W_m}\gamma$. By Theorem 5.3.6, we have

$$\begin{aligned}\int_{\mathbb{R}^n} |\nabla V_m|^2 e^{-V_m} d\gamma - \int_{\mathbb{R}^n} |\nabla W_m|^2 e^{-W_m} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W_m||_{HS}^2 e^{-W_m} d\gamma \\ \geq 2\mathrm{Ent}_\gamma(e^{-V_m}) - 2\mathrm{Ent}_\gamma(e^{-W_m}) + \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi_m||_{HS}^2 e^{-V_m} d\gamma.\end{aligned} \tag{5.3.34}$$

It follows that, according to (5.3.32),

$$(i) \qquad \sup_{m\geq 1}\int_{\mathbb{R}^n} ||\nabla^2\varphi_m||_{HS}^2 e^{-V_m} d\gamma < +\infty.$$

On the other hand,

$$\int_{\mathbb{R}^n} |\nabla\varphi_m|^2\, e^{-V_m} d\gamma = W_2^2(e^{-V_m}\gamma, e^{-W_m}\gamma).$$

Note that, by transport cost inequality for Gaussian measure: $W_2^2(e^{-V_m}\gamma, \gamma) \leq 2\mathrm{Ent}_\gamma(e^{-V_m})$, the right hand side of above equality is dominated by $4(\mathrm{Ent}_\gamma(e^{-V_m}) + \mathrm{Ent}_\gamma(e^{-W_m}))$ which is bounded with respect to n, due to (5.3.32). Therefore

$$(ii) \qquad \sup_{m\geq 1}\int_{\mathbb{R}^n} |\nabla\varphi_m|^2 e^{-V_m} d\gamma < +\infty.$$

For the moment, we suppose that

$$(H) \qquad 0 < \delta_1 \leq e^{-V}.$$

Under (H), above (i), (ii) imply that

$$\sup_{m\geq 1}\Big[\int_{\mathbb{R}^n} |\nabla\varphi_m|^2 d\gamma + \int_{\mathbb{R}^n} ||\nabla^2\varphi_m||_{HS}^2 d\gamma\Big] < +\infty.$$

Now by Poincaré inequality $\int_{\mathbb{R}^n} |\varphi_m - \mathbb{E}(\varphi_m)|^2\, d\gamma \leq \int_{\mathbb{R}^n} |\nabla\varphi_m|^2 d\gamma$ where $\mathbb{E}(\varphi_m)$ denotes the integral of φ_m with respect to γ. Up to changing φ_m by $\varphi_m - \mathbb{E}(\varphi_m)$, we get

$$\sup_{m\geq 1} ||\varphi_m||_{\mathbb{D}_2^2(\gamma)} < +\infty. \tag{5.3.35}$$

Therefore there exists $\varphi \in \mathbb{D}_2^2(\gamma)$ such that $\varphi_m \to \varphi, \nabla\varphi_m \to \nabla\varphi$ and $\nabla^2\varphi_m \to \nabla^2\varphi$ weakly in $L^2(\gamma)$. Now by Theorem 5.3.8 (for $p = 1$), there exists a constant $K > 0$ (independent of n), such that

$$||\nabla^2\varphi_m - \nabla^2\varphi_q||_{L^1(\gamma)}^2 \leq K\left(||V_m - V_q||_{L^1(\gamma)} + ||\nabla W_m - \nabla W_q||_{L^2(\gamma)}^2\right) \to 0, \tag{5.3.36}$$

as $m, q \to +\infty$. Also by (5.3.16),

$$||\nabla\varphi_m - \nabla\varphi_q||^2_{L^2(\gamma)} \leq \frac{4}{1-c}||V_m - V_q||_{L^1(\gamma)} + \frac{4}{(1-c)^2}||\nabla W_m - \nabla W_q||^2_{L^2(\gamma)} \to 0, \tag{5.3.37}$$

as $m, q \to +\infty$. It follows that $\nabla^2\varphi_m$ converges to $\nabla^2\varphi$ in $L^1(\gamma)$ and $\nabla\varphi_m$ converges to $\nabla\varphi$ in $L^2(\gamma)$, as $m \to +\infty$. Up to a subsequence, $\nabla^2\varphi_m$ converges to $\nabla^2\varphi$ and $\nabla\varphi_m$ converges to $\nabla\varphi$ almost everwhere. Therefore $x + \nabla\varphi(x)$ pushes $e^{-V}\gamma$ to $e^{-W}\gamma$ and $\mathrm{Id} + \nabla^2\varphi$ is positive.

Theorem 5.3.9. *Let $V \in \mathbb{D}^2_1(\mathbb{R}^n, \gamma)$ and $W \in \mathbb{D}^2_2(\mathbb{R}^n, \gamma)$ satisfying conditions (5.3.32) and (H), then the optimal transport map $x \to x + \nabla\varphi(x)$ which pushes $e^{-V}\gamma$ to $e^{-W}\gamma$ is such that $\varphi \in \mathbb{D}^2_2(\mathbb{R}^n, \gamma)$ and*

$$\begin{aligned}\int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma - \int_{\mathbb{R}^n} |\nabla W|^2 e^{-W} d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||^2_{HS} e^{-W} d\gamma \\ \geq 2\mathrm{Ent}_\gamma(e^{-V}) - 2\mathrm{Ent}_\gamma(e^{-W}) + \frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi||^2_{HS} e^{-V} d\gamma.\end{aligned} \tag{5.3.38}$$

Proof. Again due to (5.3.32), as $m \to +\infty$, at least for a subsequence,

$$\int_{\mathbb{R}^n} |\nabla V_m|^2 e^{-V_m} d\gamma \to \int_{\mathbb{R}^n} |\nabla V|^2 e^{-V} d\gamma, \quad \int_{\mathbb{R}^n} |\nabla W_m|^2 e^{-W_m} d\gamma \to \int_{\mathbb{R}^n} |\nabla W|^2 e^{-W} d\gamma.$$

On the other hand, for an almost everywhere convergent subsequence, by Fatou lemma,

$$\lim_{m\to+\infty} \int_{\mathbb{R}^n} ||\nabla^2\varphi_m||^2_{HS} e^{-V_m} d\gamma \geq \int_{\mathbb{R}^n} ||\nabla^2\varphi||^2_{HS} e^{-V} d\gamma.$$

At the limit, (5.3.34) leads to (5.3.38). □

In what follows, we will drop the condition (H), but assume (5.3.32). Let $m \geq 1$, consider

$$V_m = V \wedge m.$$

Then $V_m \leq V$, $|\nabla V_m| \leq |\nabla V|$ and V_m converge to V in $\mathbb{D}^2_1(\mathbb{R}^n, \gamma)$. Let $a_m = \int_{\mathbb{R}^n} e^{-V_m} d\gamma$; then $a_m \to 1$, as $m \to +\infty$. Let $x \to x + \nabla\varphi_m(x)$ be the optimal map which pushes $e^{-V_m}/a_m \, d\gamma$ forward to $e^{-W} d\gamma$. Then by (5.3.38),

$$\frac{1-c}{2}\int_{\mathbb{R}^n} ||\nabla^2\varphi_m||^2_{HS} \frac{e^{-V_m}}{a_m} d\gamma \leq \delta_2 \int_{\mathbb{R}^n} |\nabla V|^2 \, d\gamma + \frac{2}{1-c}\int_{\mathbb{R}^n} ||\nabla^2 W||^2_{HS} e^{-W} d\gamma.$$

On the other hand,

$$\int_{\mathbb{R}^n} |\nabla\varphi_m|^2 \frac{e^{-V}}{a_m} d\gamma \leq \int_{\mathbb{R}^n} |\nabla\varphi_m|^2 \frac{e^{-V_m}}{a_m} d\gamma = W_2^2(\frac{e^{-V_m}}{a_m}\gamma, e^{-W}\gamma).$$

It follows that

$$\sup_{m\geq 1}\Big[\int_{\mathbb{R}^n}|\nabla\varphi_m|^2 e^{-V}d\gamma+\int_{\mathbb{R}^n}||\nabla^2\varphi_m||_{HS}^2 e^{-V}d\gamma\Big]<+\infty. \tag{5.3.39}$$

Since the Dirichlet form $\mathcal{E}(f,f)=\int_{\mathbb{R}^n}|\nabla f|^2\, e^{-V}d\gamma$ is closed, then there exists $Y\in\mathbb{D}_1^2(\mathbb{R}^n,\mathbb{R}^n;e^{-V}\gamma)$ such that

$$\nabla\varphi_m\to Y,\quad \nabla^2\varphi_m\to\nabla Y$$

weakly in $L^2(e^{-V}\gamma)$. Then, for any $\xi\in L^\infty(\mathbb{R}^n,\mathbb{R}^n;e^{-V}\gamma)$,

$$(i)\qquad \lim_{m\to+\infty}\int_{\mathbb{R}^n}\langle\xi,\nabla\varphi_m\rangle\, e^{-V}d\gamma=\int_{\mathbb{R}^n}\langle\xi,Y\rangle\, e^{-V}d\gamma.$$

On the other hand, by stability of optimal transport plans, there exists a 1-convex function $\varphi\in L^1(e^{-V}\gamma)$ such that $x\to x+\nabla\varphi(x)$ is the unique optimal transport map which pushes $e^{-V}d\gamma$ forward to $e^{-W}d\gamma$ (see [58],p.74), such that, up to a subsequence,

$$(ii)\qquad \lim_{m\to+\infty}\int_{\mathbb{R}^n}\psi(x,x+\nabla\varphi_m(x))\,\frac{e^{-V_m}}{a_m}d\gamma=\int_{\mathbb{R}^n}\psi(x,x+\nabla\varphi(x))\, e^{-V}d\gamma,$$

for any bounded continuous function $\psi:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$. Let α_R be a cut-off function on $\mathbb{R}$: $\alpha_R\in C_b(\mathbb{R})$ such that $0\leq\alpha_R\leq 1$ and $\alpha_R=1$ over $[0,R]$ and $\alpha_R=0$ over $[2R,+\infty[$. Take ξ as a bounded continuous function $\mathbb{R}^n\to\mathbb{R}^n$ and consider

$$\psi(x,y)=\langle\xi(x),y\rangle\alpha_R(|y|).$$

By above (ii), and noting $\nabla\Phi_m(x)=x+\nabla\varphi_m(x)$ and $\nabla\Phi(x)=x+\nabla\varphi(x)$, we have

(iii)

$$\lim_{m\to+\infty}\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi_m(x)\rangle\alpha_R(|\nabla\Phi_m(x)|)\frac{e^{-V_m}}{a_m}d\gamma=\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi(x)\rangle\alpha_R(|\nabla\Phi(x)|)e^{-V}d\gamma.$$

Note that

$$\begin{aligned}&\Big|\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi_m(x)\rangle\big(1-\alpha_R(|\nabla\Phi_m(x)|)\big)\,\frac{e^{-V_m}}{a_m}d\gamma\Big|\\ &=\Big|\int_{\mathbb{R}^n}\langle\xi((\nabla\Phi_m)^{-1}(y)),y\rangle\big(1-\alpha_R(|y|)\big)\, e^{-W}d\gamma\Big|\leq\delta_2\,||\xi||_\infty\int_{\{|y|\geq R\}}|y|\, d\gamma(y),\end{aligned}$$

Combining this estimate with above (iii), we get

$$\lim_{m\to+\infty}\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi_m(x)\rangle\,\frac{e^{-V_m}}{a_m}d\gamma=\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi(x)\rangle\, e^{-V}d\gamma. \tag{5.3.40}$$

From (5.3.40), it is not hard to see that

$$\lim_{m\to+\infty}\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi_m(x)\rangle\, e^{-V}d\gamma=\int_{\mathbb{R}^n}\langle\xi(x),\nabla\Phi(x)\rangle\, e^{-V}d\gamma.$$

Now comparing with (i), we get that $\nabla\Phi(x)=x+Y(x)$ or $Y=\nabla\varphi$.

Theorem 5.3.10. *Let $V\in\mathbb{D}_1^2(\mathbb{R}^n,\gamma)$ and $W\in\mathbb{D}_2^2(\mathbb{R}^n,\gamma)$ satisfying conditions (5.3.32). Then the optimal transport map $x\to x+\nabla\varphi(x)$ which pushes $e^{-V}\gamma$ to $e^{-W}\gamma$ is such that $\varphi\in\mathbb{D}_2^2(\mathbb{R}^n,\gamma)$ and*

$$\begin{aligned}\int_{\mathbb{R}^n}|\nabla V|^2e^{-V}d\gamma-\int_{\mathbb{R}^n}|\nabla W|^2e^{-W}d\gamma+\frac{2}{1-c}\int_{\mathbb{R}^n}||\nabla^2W||_{HS}^2e^{-W}d\gamma\\ \geq 2\mathrm{Ent}_\gamma(e^{-V})-2\mathrm{Ent}_\gamma(e^{-W})+\frac{1-c}{2}\int_{\mathbb{R}^n}||\nabla^2\varphi||_{HS}^2e^{-V}d\gamma.\end{aligned}$$

Proof. Replacing V by V_m in (5.3.38) and note that

$$\underline{\lim}_{m\to+\infty}\int_{\mathbb{R}^n}||\nabla^2\varphi_m||_{HS}^2\,\frac{e^{-V_m}}{a_m}d\gamma\geq\underline{\lim}_{m\to+\infty}\int_{\mathbb{R}^n}||\nabla^2\varphi_m||_{HS}^2\,\frac{e^{-V}}{a_m}d\gamma\geq\int_{\mathbb{R}^n}||\nabla^2\varphi||_{HS}^2\,e^{-V}d\gamma,$$

we get the result by letting $m\to+\infty$ in (5.3.38). It remains to prove that $\varphi\in L^2(e^{-V}\gamma)$. In fact, let Π_0 be the optimal plan induced by $x\to x+\nabla\varphi(x)$. Then (see section 1), under Π_0,

$$\varphi(x)+\psi(y)=|x-y|^2.$$

But we have seen in section 1 that $\psi\in L^2(e^{-W}\gamma)$. Then under Π_0,

$$\varphi(x)^2\leq 2\psi(y)^2+2|x-y|^4.$$

Let Ω be the set of couples (x,y) such that above inequality holds, then $\Pi_0(\Omega)=1$. We have

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\varphi^2\,d\Pi_0=\int_\Omega\varphi^2\,d\Pi_0\leq 2\int_{\mathbb{R}^n}\psi^2d\Pi_0+2\int_{\mathbb{R}^n\times\mathbb{R}^n}|x-y|^4\,d\Pi_0(x,y).$$

It follows that

$$\int_{\mathbb{R}^n}\varphi^2\,e^{-V}d\gamma\leq 2\int_{\mathbb{R}^n}\psi^2\,e^{-W}d\gamma+16\delta_2\int_{\mathbb{R}^n}|x|^4d\gamma(x),$$

which is finite. The proof is complete. □

We conclude this section by the following result.

Theorem 5.3.11. *Let* $V_1, V_2 \in \mathbb{D}_1^2(\mathbb{R}^n, \gamma)$ *and* $W_1, W_2 \in \mathbb{D}_2^2(\mathbb{R}^n, \gamma)$ *satisfying* (5.3.32) *and* (H). *Let* $\nabla\varphi_1, \nabla\varphi_2$ *be the associated optimal transport maps. Then for* $1 \leq p < 2$

$$\begin{aligned} ||\nabla^2\varphi_1 - \nabla^2\varphi_2||^2_{L^p(e^{-V_2}\gamma)} \leq & 2M(\nabla^2\varphi_1, \nabla^2\varphi_2)\Big[3\int_{\mathbb{R}^n}(V_1 - V_2)e^{-V_2}d\gamma \\ & + \frac{2}{1-c}\int_{\mathbb{R}^n}|\nabla(W_1 - W_2)|^2 e^{-W_2}d\gamma\Big], \end{aligned} \tag{5.3.41}$$

where

$$M(\nabla^2\varphi_1, \nabla^2\varphi_2) = \max\Big(\Big\|||I + \nabla^2\varphi_1||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)}, \Big\|||I + \nabla^2\varphi_2||_{op}\Big\|^2_{L^{\frac{2p}{2-p}}(e^{-V_2}\gamma)}\Big).$$

Chapter 6

Monge Problem on infinite dimensional spaces

This chapter is concerned with the existence of optimal transport maps on a Wiener space (X, H, μ). We will discuss the following three situations:

1. The space X, itself is a separable Hilbert space, says, $X = l^2(c)$ introduced in chapter 4, endowed with the Hilbert norm $||.||$. The cost will be $c(x, y) = ||x - y||$. We will follow recent works by Champion and De Pascale [21].

2. The cost on the Wiener space (X, H, μ) will be $c(x, y) = |x - y|_H^2$. In this case, the existence and uniqueness of optimal transport maps have been proved by Feyel and Üstünel. Our contribution is that when the target measure is a logarithmic concave measure, we can construct explicitely optimal transport maps and establish more regularity property.

3. The cost will be $c(x, y) = ||x - y||_{k,\gamma}^p$ considered in Chapter 2, which was proved to be strictly convex.

6.1 On infinite dimensional Hilbert spaces

Let $X = l^2(c)$ which is the space of sequence $x := (x_n)$ such that

$$||x|| = \sum_{n \geq 0} c_n x_n^2 < +\infty,$$

where (c_n) is a sequence of positive real number such that $\sum_{n \geq 0} c_n < +\infty$. Without loss of generality, we assume that

$$\sup_{n \geq 0} c_n \leq 1.$$

The space X supports a Gaussian measure μ, such that the covariance matrice can be expressed by

$$\int_X \langle e_n, x\rangle\langle e_m, x\rangle \, d\mu(x) = \delta_{nm}\, c_n,$$

where (e_n) denotes the canonical basis of $\mathbb{R}^{\mathbb{N}}$ and δ_{nm} is the Kronecker's symbole.

In the approach of Champion and De Pascale, the differentiation theorem for the measure of reference played a key role. Unfortunately, this property is not well established in infinite dimensional spaces. However in the case where c_n decreases very rapidly, J. Tiser proved [56] that such a property holds.

Theorem 6.1.1. *Suppose that for some $\alpha > 5/2$,*

$$\frac{c_{n+1}}{c_n} \le n^{-\alpha}, \quad n \ge 1. \tag{6.1.1}$$

Then

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - f(x)| d\mu = 0 \qquad \text{for } \mu - \text{a.a. x} \in X$$

for any $f \in L^p(X, \mu)$ and $p > 1$.

The set of $x \in X$ such that $\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - f(x)| d\mu = 0$ is called the set of *Lebesgue points* of f and will be denoted by $Leb(f)$. Thus Theorem 6.1.1 says that $\mu(Leb(f)) = 1$. In the case of $f = \mathbb{1}_A$, we will call x a *Lebesgue point of A*.

In what follows, we assume that the measure μ satisfies the condition (6.1.1). The aim of this section is to prove the following theorem.

Theorem 6.1.2. *Let ρ_0 and ρ_1 be probability measures on X, having finite relative entropy with respect to μ. Then the problem*

$$\inf_{T_\# \rho_0 = \rho_1} \int_X \|x - T(x)\| d\rho_0(x) \tag{6.1.2}$$

has at least one solution $T : X \longrightarrow X$.

Remark 6.1.3. *In fact Theorem 6.1.1 is required only to get the Proposition 6.1.10. All other results in this section are available without Lebesgue points.*

The classical way to find a solution of (6.1.2) is to introduce the following Monge-Kantorovich problem:

$$\min_{\Pi \in C(\rho_0,\rho_1)} \int_{X\times X} \|x - y\| d\Pi(x, y), \tag{6.1.3}$$

where $C(\rho_0, \rho_1)$ is the set of *couplings* between ρ_0 and ρ_1. The nonempty set of solutions, says, optimal couplings to (6.1.3) will be denoted by $\mathcal{O}_1(\rho_0, \rho_1)$. Among these optimal couplings, we shall show there is at least one which is carried by a graph of some map T and therefore this map will be a solution to (6.1.2).
With the power 1, the cost $||.||$ is not strictly convex, the set $\mathcal{O}_1(\rho_0, \rho_1)$ does not contain sufficient informations to construct such a map T. Thus we need to introduce a *second variational problem*, with a new cost to minimize over the set of optimal couplings of (6.1.3):

$$\min_{\Pi \in \mathcal{O}_1(\rho_0,\rho_1)} \int_{X\times X} \alpha(x-y) d\Pi(x,y), \tag{6.1.4}$$

with

$$\alpha(x-y) := \sqrt{1+||x-y||^2}.$$

This cost α being strictly convex, will bring in some sense the directions that the optimal coupling should take in order to be concentrated on a graph of some map. We denote by $\mathcal{O}_2(\rho_0, \rho_1)$ the subset of $\mathcal{O}_1(\rho_0, \rho_1)$ of those optimal couplings which minimize (6.1.4). It is easy to see that $\alpha(x-y) \leq 1 + ||x-y||$ so that if (6.1.3) is finite for some coupling then (6.1.4) is also finite, and the set $\mathcal{O}_2(\rho_0, \rho_1)$ is a nonempty (by weak compacity) and a convex subset of $C(\rho_0, \rho_1)$.

We say that a coupling $\Pi \in C(\rho_0, \rho_1)$ satisfies the *convexity property* if the relative entropy is $1-$convex along $\rho_t := ((1-t)P_1 + tP_2)_\# \Pi$, namely

$$\mathrm{Ent}_\mu(\rho_t) \leq (1-t)\mathrm{Ent}_\mu(\rho_0) + tEnt_\mu(\rho_1) - \frac{t(1-t)}{2} W^2_{1,||.||}(\rho_0, \rho_1),$$

holds for any $t \in (0,1)$.
Finally we are interested in the following set:

$$\overline{\mathcal{O}_2}(\rho_0, \rho_1) := \big\{\Pi \in \mathcal{O}_2(\rho_0, \rho_1), \Pi \text{ enjoys the } \textit{convexity property}\big\}.$$

The fact that $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ is non empty is the purpose of Theorem 6.1.6. It will play a key role in our approach since any coupling of $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ will bring us sufficient information to show that it is concentrated on a graph of some measurable map.

Lemma 6.1.4. *If $\Pi \in \mathcal{O}_2(\rho_0, \rho_1)$ then Π is concentrated on some $\sigma-$compact set Γ satisfying:*

$$\forall (x,y),(x',y') \in \Gamma, \quad x \in [x',y'] \Rightarrow (\nabla\alpha(y-x') - \nabla\alpha(y'-x), x-x') \geq 0, \tag{6.1.5}$$

where $[x',y']$ denotes the segment from x' to y'.

Proof. Since Π is an optimal coupling, there is a Borel subset Γ of $X \times X$ which is $||.||-$cyclically monotone. By inner regularity of probability measure, up to remove a Borel set of zero measure, we can take Γ as a $\sigma-$compact subset. According to Proposition 3.2.5, we can find a potential $u : X \longrightarrow X$ such that:

$$\forall (x, y) \in \Gamma, \qquad u(x) - u(y) = \|x - y\|.$$

Note that Π minimizes also

$$\min_{\Pi \in C(\rho_0, \rho_1)} \int_{X \times X} \beta(x, y) d\Pi(x, y),$$

where

$$\beta(x, y) = \begin{cases} \alpha(x - y) \text{ if } u(x) - u(y) = ||x - y||, \\ \quad +\infty \qquad\qquad \text{otherwise.} \end{cases}$$

Let $(x, y), (x', y') \in \Gamma$ such that $x \in [x', y']$. We have then:

$$\begin{aligned} u(x) &= u(y) + \|x - y\|, \\ u(x') &= u(y') + \|x' - y'\|, \end{aligned}$$

and since $x \in [x', y']$, we also have:

$$||x' - y'|| = ||x - x'|| + ||x - y'||.$$

Our potential u is a $1-$Lipschitz map, so:

$$u(x') = u(y') + \|x - x'\| + \|x - y'\| \geq u(x) + \|x - x'\| \geq u(x').$$

This equality leads to:

$$\begin{aligned} u(x') &= u(x) + \|x - x'\| = u(y) + \|x - y\| + \|x - x'\| \\ &\geq u(y) + \|y - x'\| \geq u(x'). \end{aligned}$$

With the previous notation, it turns out that $\beta(x', y) = \alpha(x' - y)$ and $\beta(x, y') = \alpha(x - y')$. Moreover thanks to Proposition 3.2.3, we also know that Π is $\beta-$cyclically monotone hence by symmetry of α:

$$\alpha(y - x) + \alpha(y' - x') \leq \alpha(y' - x) + \alpha(y - x').$$

But by convexity of α, we have:

$$\begin{aligned} \alpha(y - x) - \alpha(y - x') &\geq \nabla\alpha(y - x').(x' - x), \\ \alpha(y' - x) - \alpha(y' - x') &\leq -\nabla\alpha(y' - x).(x - x'). \end{aligned}$$

So combining these inequalities with the $\alpha-$monotonicity we get:

$$(\nabla\alpha(y - x') - \nabla\alpha(y' - x), x - x') \geq 0.$$

□

Remark 6.1.5. As in [21] the only reason to deal with $\sigma-$compact set Γ, is that the projection $P_1(\Gamma)$ is also $\sigma-$compact, and in particular a Borel set.

$\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ is non empty:

We recall that in our case the *Wasserstein distance* is defined as

$$W(\rho_0, \rho_1) := \inf_{\Pi \in C(\rho_0, \rho_1)} \int_{X \times X} \|x - y\| d\Pi(x, y).$$

Theorem 6.1.6. *$\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ is a non empty set.*

Proof. Let $\Pi_\varepsilon \in C(\rho_0, \rho_1)$ be an optimal coupling with respect to

$$c_\varepsilon(x, y) = \|x - y\| + \varepsilon\, \alpha(x - y)$$

given in Proposition 4.3.3. Therefore the inequality (4.3.6) holds for Π_ε. If Π is a limit point of $(\Pi_\varepsilon)_\varepsilon$, then the inequality (4.3.7) holds for Π, namely Π satisfies the *convexity property*. We claim that any cluster point of $(\Pi_\varepsilon)_\varepsilon$ belongs to $\mathcal{O}_2(\rho_0, \rho_1)$. As a consequence, the set $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ will be non empty. Here is a proof to the claim.

Let Π be a limit point of $(\Pi_\varepsilon)_\varepsilon$.

First, $\Pi \in \mathcal{O}_1(\rho_0, \rho_1)$. Indeed if $\Pi_0 \in \mathcal{O}_1(\rho_0, \rho_1)$, for $\varepsilon > 0$:

$$\begin{aligned} \int \|x - y\| d\Pi_\varepsilon &\le \int \|x - y\| d\Pi_\varepsilon + \varepsilon \int \alpha(x - y) d\Pi_\varepsilon \\ &\le \int \|x - y\| d\Pi_0 + \varepsilon \int \alpha(x - y) d\Pi_0. \end{aligned}$$

Letting $\varepsilon \to 0$,

$$\int \|x - y\| d\Pi \le \liminf_{\varepsilon \to 0} \int \|x - y\| d\Pi_\varepsilon \le \int \|x - y\| d\Pi_0.$$

Secondly $\Pi \in \mathcal{O}_2(\rho_0, \rho_1)$. Indeed if $\Pi_0 \in \mathcal{O}_2(\rho_0, \rho_1)$, for $\varepsilon > 0$:

$$\begin{aligned} \int \|x - y\| d\Pi_\varepsilon + \varepsilon \int \alpha(x - y) d\Pi_\varepsilon &\le \int \|x - y\| d\Pi_0 + \varepsilon \int \alpha(x - y) d\Pi_0 \\ &\le \int \|x - y\| d\Pi_\varepsilon + \varepsilon \int \alpha(x - y) d\Pi_0, \end{aligned}$$

the latter inequality is provided by the fact that Π_0 belongs in particular to $\mathcal{O}_1(\rho_0, \rho_1)$. Remove the same terms, dividing by ε and letting $\varepsilon \to 0$,

$$\int \alpha(x - y) d\Pi \le \liminf_{\varepsilon \to 0} \int \alpha(x - y) d\Pi_\varepsilon \le \int \alpha(x - y) d\Pi_0.$$

□

Note also that for Π_1 and Π_2 are two couplings in $C(\rho_0, \rho_1)$ enjoying the *convexity property*, every linear combination $(1-t)\Pi_1 + t\Pi_2$ still enjoys the *convexity property*. As a consequence $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ is a convex set.

Properties of coupling belonging to $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$:

Throughout this part, *Differentiation theorem* 6.1.1 is used many times. We will present results in general framework. We consider $\Pi \in C(\rho_0, \rho_1)$ and $\Gamma \subset X \times X$ a $\sigma-$compact set on which Π is concentrated. For all the sequel we assume that $\rho_0 = f\mu$ (the first measure has a density f w.r.t. μ).
Let us fix a sequence of positive number $(\delta_p)_p$ which tends to 0 when p goes to infinity.

The following Lemma is a reinforcement of the one in [21] (Lemma 3.3).

Lemma 6.1.7. *Let $(y_n)_n$ be a dense sequence in X. Then we can find a Borel subset $D(\Gamma)$ of $X \times X$ on which Π is still concentrated and such that for all $(x, y) \in D(\Gamma)$ and $r > 0$, there exist $n, k \in \mathbb{N}$ satisfying $y \in B(y_n, \frac{1}{k+1}) \subset B(y, r)$, $x \in Leb(f) \cap Leb(f_{n,k})$ and for all $p \in \mathbb{N}$:*

$$\|f_{n,k}|_{B(x,\delta_p)}\|_{L^\infty} > 0,$$

where $f_{n,k}$ is the density of $(P_1)_\# \Pi_{|X \times \bar{B}(y_n, \frac{1}{k+1})}$ with respect to μ.

Proof. Let $\delta = \delta_p > 0$ be fixed. We can find a covering of X with a countable number of balls $(B(x_m^{(p)}, \delta/2))_m$. For any $(n, k) \in \mathbb{N}^2$, we consider $f_{n,k}$ the density of the first marginal of the restriction of Π to $X \times \bar{B}(y_n, \frac{1}{k+1})$ w.r.t. μ. Fix $n, k \in \mathbb{N}$ and consider

$$D_{n,k}(\delta) := \left(\cup_{m \in \mathbb{N}} \{x \in B(x_m^{(p)}, \delta/2),\ \|f_{n,k}|_{B(x,\delta)}\|_{L^\infty} = 0\}\right) \times \bar{B}(y_n, \frac{1}{k+1}).$$

It turns out that

$$\Pi(D_{n,k}(\delta)) \leq \sum_{m \in \mathbb{N}} \int_{B(x_m^{(p)}, \delta/2) \setminus \{\|f_{n,k}|_{B(x,\delta)}\|_{L^\infty} > 0\}} f_{n,k}(x) d\mu(x) = 0.$$

Set $C_{n,k} = X \setminus (Leb(f) \cap Leb(f_{n,k})) \times X$. Then by Theorem [56],

$$\Pi(C_{n,k}) = \rho_0 \left(X \setminus (Leb(f) \cap Leb(f_{n,k}))\right) = 0.$$

Therefore Π is concentrated on the set $D_\delta(\Gamma) := \Gamma \setminus (\cup_{n,k}(D_{n,k}(\delta) \cup C_{n,k}))$. It follows $D(\Gamma) := \cap_p D_{\delta_p}(\Gamma)$ has the desired properties. Indeed for any $\delta_p > 0$ if $(x, y) \in D_{\delta_p}(\Gamma)$, by density we can find $m, n, k \in \mathbb{N}$ such that $x \in B(x_m^{(p)}, \delta_p/2), y \in B(y_n, 1/(k+1)) \subset B(y, r)$. The result follows. □

Notice that the previous result is still true for any coupling, not necessary *optimal*.

Definition 6.1.8. *Let Γ be a $\sigma-$compact subset of $X \times X$. For $y \in X$ and $r > 0$, we define:*

$$\Gamma^{-1}(\bar{B}(y,r)) := P_1\left(\Gamma \cap (H \times \bar{B}(y,r))\right).$$

*An element (x,y) of Γ is called a $\Gamma-$*regular point *if x is a Lebesgue point of $\Gamma^{-1}(\bar{B}(y,r))$ for any $r > 0$.*

It is worth to noting that from the definition 6.1.8, if Π is concentrated on Γ, then for all Borel subset A of X:

$$\Pi(A \times \bar{B}(y,r)) = \Pi\left(A \cap \Gamma^{-1}(\bar{B}(y,r)) \times \bar{B}(y,r)\right).$$

Lemma 6.1.9. *Let $D(\Gamma)$ be the subset constructed in Lemma 6.1.7; then any point in $D(\Gamma)$ is a $\Gamma-$regular point. Namely, for $(x,y) \in D(\Gamma)$,*

$$\lim_{\delta \to 0} \frac{\mu(\Gamma^{-1}(\bar{B}(y,r)) \cap B(x,\delta))}{\mu(B(x,\delta))} = 1.$$

We introduce the following notation:

$$T(\Gamma) = \{(1-t)x + ty,\ (x,y) \in \Gamma\}.$$

Since Γ is $\sigma-$compact, $T(\Gamma)$ is $\sigma-$compact as well.

Proposition 6.1.10. *Let $\rho_0, \rho_1 \in D(\mathrm{Ent}_\mu)$, and $\Pi \in \overline{\mathcal{O}_2}(\rho_0,\rho_1)$ concentrated on a $\sigma-$compact set Γ. Then for all $(x,y_0),(x,y_1)$ belonging to the set $D(\Gamma)$ obtained in the Lemma 6.1.7, with $y_0 \neq y_1$ and for each $r > 0$ small enough such that the closed balls centered at y_0 and y_1 with radius r are disjoint, it holds:*

$$\mu\Big(T\left(\Gamma \cap (B(x,\delta_p) \times B(y_0,r))\right) \cap \Gamma^{-1}(\bar{B}(y_1,r)) \cap B(x,2\delta_p)\Big) > 0,$$

for $p \in \mathbb{N}$ large enough.

Proof. First we remark by Lemma 6.1.9 that

$$\lim_{\delta \to 0} \frac{\mu\Big(\Gamma^{-1}(\bar{B}(y_0,r)) \cap \Gamma^{-1}(\bar{B}(y,r)) \cap B(x,\delta)\Big)}{\mu(B(x,\delta))} = 1. \tag{6.1.6}$$

By Lemma 6.1.7, there exist $n_0, n_1, k \in \mathbb{N}$ such that $B(y_{n_0}, \frac{1}{k+1}) \subset B(y_0,r)$, $B(y_{n_1}, \frac{1}{k+1}) \subset B(y_1,r)$. Since δ_p decreases to 0, we find $p \in \mathbb{N}$ large enough so that $0 < \delta = \delta_p < \|x - y_0\| + r$.

The corresponding densities given by Lemma 6.1.7 are denoted by $f_{n_0,k},\ f_{n_1,k}$. Let us consider the Borel subset (up to a negligible set)

$$G_x := \{z \in B(x,\delta),\ f_{n_0,k}(z) > 0,\ f_{n_1,k}(z) > 0\},$$

which has a positive measure: $\mu(G_x) > 0$. This is due to (6.1.6) and to the fact that x is a Lebesgue point of $f_{n_0,k}$ and of $f_{n_1,k}$.

We notice that:

$$\Pi\left(G_x \times \bar{B}(y_{n_1}, \frac{1}{k+1})\right) = \Pi\left(G_x \cap \Gamma^{-1}(\bar{B}(y_{n_1}, \frac{1}{k+1})) \times \bar{B}(y_{n_1}, \frac{1}{k+1})\right).$$

Hence,

$$\int_{G_x \cap \Gamma^{-1}(\bar{B}(y_{n_1}, \frac{1}{k+1}))} f_{n_1,k} d\mu = \int_{G_x} f_{n_1,k} d\mu > 0.$$

It follows that

$$\mu(G_x \cap \Gamma^{-1}(\bar{B}(y_1, r))) \geq \mu\left(G_x \cap \Gamma^{-1}(\bar{B}(y_{n_1}, \frac{1}{k_1+1}))\right) > 0. \tag{6.1.7}$$

Let

$$A(\delta) := B(x, 2\delta) \cap \Gamma^{-1}(\bar{B}(y_1, r)) \cap T\left(\Gamma \cap (B(x,\delta) \times B(y_0, r))\right).$$

Consider the set $A_x := G_x \times \bar{B}(y_{n_0}, \frac{1}{k+1})$, and denote by Π_{A_x} the restriction of Π on A_x. We fix from now $t \in (0, \frac{\delta}{\|x-y_0\|+r})$ so that: if $z \in B(x,\delta)$ and $w \in B(y_0, r)$ then $(1-t)z + tw \in B(x, 2\delta)$. Indeed

$$\begin{aligned} ||(1-t)z + tw - x|| &\leq (1-t)||z-x|| + t||w-x|| \\ &\leq ||z-x|| + t(||w-y_0|| + ||y_0 - x||) \\ &< \delta + \delta = 2\delta. \end{aligned}$$

Therefore if we define $\rho_t^{A_x} := ((1-t)P_1 + tP_2)_\# \Pi_{A_x}$, firstly we have:

$$(P_1)_\# \Pi_{A_x}(G_x) \leq (P_1)_\# \Pi_{A_x}(B(x,\delta)) \leq \rho_t^{A_x}(B(x, 2\delta))$$

and

$$(P_1)_\# \Pi_{A_x}(G_x \cap \Gamma^{-1}(\bar{B}(y_1, r))) \leq \rho_t^{A_x}(B(x, 2\delta) \cap \Gamma^{-1}(\bar{B}(y_1, r))).$$

Secondly thanks to (6.1.7):

$$\begin{aligned} (P_1)_\# \Pi_{A_x}(G_x \cap \Gamma^{-1}(\bar{B}(y_1, r))) &= \Pi\left(G_x \cap \Gamma^{-1}(\bar{B}(y_1, r)) \times \bar{B}(y_{n_0}, \frac{1}{k+1})\right) \\ &= \int_{G_x \cap \Gamma^{-1}(\bar{B}(y_1, r))} f_{n_0,k} d\mu > 0. \end{aligned}$$

And we deduce

$$\rho_t^{A_x}(B(x,2\delta)\cap\Gamma^{-1}(\bar{B}(y_1,r)))>0. \tag{6.1.8}$$

On the other hand, notice that $\rho_t^{A_x}$ is concentrated on $T(\Gamma\cap(B(x,\delta)\times B(y_0,r))$ hence:

$$\begin{aligned}&\rho_t^{A_x}(B(x,2\delta)\cap\Gamma^{-1}(\bar{B}(y_1,r)))\\=&\rho_t^{A_x}\left(B(x,2\delta)\cap T\left(\Gamma\cap(B(x,\delta)\times B(y_0,r))\right)\cap\Gamma^{-1}(\bar{B}(y_1,r))\right).\end{aligned}$$

Combining this fact with (6.1.8), we get:

$$\rho_t^{A_x}(A(\delta))>0.$$

Now remark that $\rho_t^{A_x}(A(\delta))\le\rho_t(A(\delta))$. By convexity inequality, ρ_t is absolutely continuous w.r.t. μ. Hence it implies $\mu(A(\delta))>0$. □

Proof of Theorem 6.1.2.
In fact, it remains to prove that

Theorem 6.1.11. *Any element of $\overline{\mathcal{O}_2}(\rho_0,\rho_1)$ is induced by a map T. Moreover $\overline{\mathcal{O}_2}(\rho_0,\rho_1)$ is reduced to one element.*

Proof. Let $\Pi\in\overline{\mathcal{O}_2}(\rho_0,\rho_1)$. In particular $\Pi\in\mathcal{O}_2(\rho_0,\rho_1)$ and is concentrated on a $\sigma-$compact set Γ satisfying (6.1.5). Furthermore Lemma 6.1.7 provides us a $\sigma-$compact set $D(\Gamma)$ on which Π is still concentrated. We claim that $D(\Gamma)$ is contained in a graph of some Borel map. Let (x_0,y_0) and (x_0,y_1) in $D(\Gamma)$ and suppose that $y_0\neq y_1$. We can also assume $x_0\neq y_0$. By strict convexity of α, we have:

$$((y_1-x_0)-(y_0-x_0),\nabla\alpha(y_1-x_0)-\nabla\alpha(y_0-x_0))>0.$$

Hence either $(y_1-x_0,\nabla\alpha(y_1-x_0)-\nabla\alpha(y_0-x_0))$ or $(y_0-x_0,\nabla\alpha(y_0-x_0)-\nabla\alpha(y_1-x_0))$ is positive. So without loss of generality we assume that:

$$(\nabla\alpha(y_1-x_0)-\nabla\alpha(y_0-x_0),y_0-x_0)<0.$$

By expression

$$(\nabla\alpha(x),y)=\frac{(x,y)}{\sqrt{1+||x||^2}},$$

we see that there exists $r>0$ small enough so that for all $x,x'\in B(x_0,r)$ and for all $y'\in B(y_0,r),\ y\in B(y_1,r)$:

$$(\nabla\alpha(y-x')-\nabla\alpha(y'-x),y'-x)<0. \tag{6.1.9}$$

$r > 0$ can be chosen such that the balls $\bar{B}(y_0, r)$ and $\bar{B}(y_1, r)$ are disjoint. Applying Proposition 6.1.10 to $((x_0, y_0), (x_0, y_1))$ we get:

$$\mu\left(T\left(\Gamma \cap (B(x_0, \delta_p) \times B(y_0, r))\right) \cap \Gamma^{-1}(\bar{B}(y_1, r)) \cap B(x_0, 2\delta_p)\right) > 0,$$

for $p \in \mathbb{N}$ large enough. As a consequence we can find a $\delta = \delta_p \in (0, r/2)$ small enough in such a way that there exist $(x', y') \in \Gamma \cap (B(x_0, \delta) \times B(y_0, r))$ and $x \in [x', y'] \cap B(x_0, 2\delta)$ and y such that:

$$(x, y) \in \Gamma \cap \left(([x', y'] \cap B(x_0, 2\delta)) \times B(y_1, r)\right).$$

Since $x \in [x', y']$, we have $x - x' = \frac{|x-x'|}{|y'-x|}(y' - x)$. So by (6.1.5), we have:

$$(\nabla\alpha(y - x') - \nabla\alpha(y' - x), x - x') = \frac{|x - x'|}{|y' - x|}(\nabla\alpha(y - x') - \nabla\alpha(y' - x), y' - x) \geq 0,$$

which contradicts (6.1.9). Therefore $y_1 = y_0$ and Π is supported by the graph of a map T.

Uniqueness of $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$. Let Π_1 and Π_2 in $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$, supported respectively by T_1 and T_2. By convexity of $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$,

$$\Pi = \frac{\Pi_1 + \Pi_2}{2} \in \overline{\mathcal{O}_2}(\rho_0, \rho_1).$$

Therefore Π will be supported by a map T. Let $\varphi, \psi : X \to \mathbb{R}$ be bounded continuous functions, we have

$$\int_{X \times X} \varphi(x)\psi(y)\, d\Pi(x, y) = \frac{1}{2}\Big[\int_{X \times X} \varphi(x)\psi(y)\, d\Pi_1(x, y) + \int_{X \times X} \varphi(x)\psi(y)\, d\Pi_2(x, y)\Big],$$

which yields

$$\int_X \varphi(x)\psi(T(x))\, d\rho_0(x) = \int_X \varphi(x)\frac{1}{2}\Big(\psi(T_1(x)) + \psi(T_2(x))\Big)\, d\rho_0(x).$$

It follows that for ρ-a.e x,

$$\delta_{T(x)} = \frac{1}{2}(\delta_{T_1(x)} + \delta_{T_2(x)}).$$

Therefore $T = T_1 = T_2$. □

Let us make some comments.

We have proved that $\overline{\mathcal{O}_2}(\rho_0, \rho_1)$ is reduced to one element. However we do not know if $\mathcal{O}_2(\rho_0, \rho_1)$ has a unique element.

In [21], the authors do not require the absolute continuity of ρ_t because the Lebesgue measure is doubling and invariant by translations. Thanks to that they can obtain good bounds for ρ_t (see Proposition 2.2 in [21]).

6.1.1 Stability of optimal maps

Let $c_\varepsilon(x,y) := ||x-y|| + \varepsilon\alpha(x-y)$ and $c(x,y) := ||x-y||$. Since c_ε is strictly convex and differentiable, by the recent work of Champion and De Pascale [22], there is a unique optimal coupling Π_ε of (P_ε) and in addition Π_ε is carried by a graph T_ε. Thanks to the Proposition 4.3.3, the unicity yields that Π_ε satisfies the *convexity property*

$$\mathrm{Ent}_\mu(\rho_t) \leq (1-t)\mathrm{Ent}_\mu(\rho_0) + t\mathrm{Ent}_\mu(\rho_1) - \frac{t(1-t)}{2(1+\varepsilon)^2}\left(\mathcal{W}_{\varepsilon,||.||}(\rho_0,\rho_1) - \varepsilon\right)^2,$$

for any $t \in [0,1]$ and $\rho_t := (T_t)_\#\Pi_\varepsilon$. As in the proof of Theorem 6.1.6, and by Theorem 6.1.11, $(\Pi_\varepsilon)_\varepsilon$ converges weakly to a unique optimal coupling Π for c, satisfying the *convexity property*:

$$\mathrm{Ent}_\mu(\rho_t) \leq (1-t)\mathrm{Ent}_\mu(\rho_0) + t\mathrm{Ent}_\mu(\rho_1) - \frac{t(1-t)}{2}\mathcal{W}_{||.||}(\rho_0,\rho_1)^2.$$

Moreover Π is carried by some graph T. We have the following stability result.

Proposition 6.1.12. *$(T_\varepsilon)_\varepsilon$ converges to T in probability, namely:*

$$\rho_0\left(\{x \in X,\ ||T_\varepsilon(x) - T(x)|| > \eta\}\right) \overset{\varepsilon\to 0}{\longrightarrow} 0, \qquad \forall \eta > 0.$$

The proof of this Proposition lies in the use of **Lusin's theorem**, which is true in our case because of the inner regularity of Gaussian measure μ: there exists a sequence of compact sets $K_n \subset X$ such that

$$\mu\left(\cup_{n\geq 1} K_n\right) = 1.$$

Proof. Let $\delta > 0$ be fixed. We can find a compact subset $\tilde{K} \subset X$ such that $\rho_0(\tilde{K}^c) < \delta/2$. By Lusin's Theorem, there is a compact subset $K \subset \tilde{K}$ such that $\rho_0(\tilde{K}\backslash K) < \delta/2$ and on which T is continuous. We consider for $\eta > 0$,

$$A_\eta := \left\{(x,y) \in K \times X,\ ||T(x) - y|| \geq \eta\right\}.$$

Since Π is concentrated on the graph of T, we have $\Pi(A_\eta) = 0$ for any $\eta > 0$. As Π_ε converges weakly to Π and A_η is closed, we have

$$\begin{aligned} 0 = \Pi(A_\eta) &\geq \limsup_{\varepsilon\to 0} \Pi_\varepsilon(A_\eta) \\ &= \limsup_{\varepsilon\to 0} \rho_0\left(x \in K,\ ||T(x) - T_\varepsilon(x)|| \geq \eta\right) \\ &\geq \limsup_{\varepsilon\to 0} \rho_0\left(x \in H,\ ||T(x) - T_\varepsilon(x)|| \geq \eta\right) - \delta. \end{aligned}$$

Letting $\delta \to 0$ yields the result. □

6.2 On the Wiener space with the quadratic cost

Let (X, H, μ) be an abstract Wiener space. In this section, we will consider

$$c(x, y) = d_H(x, y)^2,$$

where

$$d_H(x, y) = \begin{cases} |x-y|_H & \text{if} \quad x - y \in H; \\ +\infty & \text{otherwise.} \end{cases}$$

For $\nu_1, \nu_2 \in \mathcal{P}(X)$, we consider the following Wasserstein distance

$$W_2^2(\nu_1, \nu_2) = \inf\Big\{ \int_{X\times X} d_H(x, y)^2 \, \Pi(dx, dy);\ \Pi \in C(\nu_1, \nu_2) \Big\},$$

where $C(\nu_1, \nu_2)$ denotes the totality of probability measures on the product space $X \times X$, having ν_1, ν_2 as marginal laws. Throughout this section, the notion of *optimal coupling* will refer to the previous Wasserstein distance (w.r.t. d_H^2).
Note that $W_2(\nu_1, \nu_2)$ could take value $+\infty$. By Talagrand's inequality (see section 5.1), $W_2^2(\mu, f\mu) \le 2\mathrm{Ent}_\mu(f)$, we have

$$W_2(f\mu, g\mu) \le \sqrt{2}\Big(\sqrt{\mathrm{Ent}_\mu(f)} + \sqrt{\mathrm{Ent}_\mu(g)} \Big), \qquad (6.2.1)$$

which is finite, if the measures $f\mu$ and $g\mu$ have finite entropy. In this situation, it was proven in [37] that there is a unique map $\xi : X \to H$ such that $x \to x + \xi(x)$ pushes $f\mu$ to $g\mu$ and $W_2(f\mu, g\mu)^2 = \int_X |\xi|_H^2 \, f d\mu$. However for a general source measure $f\mu$, the construction in [37] is not explicit. In this section, we will give an explicit construction.

More precisely, the strategy is to use finite dimensional approximation, as explained in Chapter 2. Once you deal with measures in finite dimensional spaces, the Cameron-Martin norm is nothing but the Euclidian norm, so the Brenier's theorem (see Chapter 3) is available. It provides us an optimal transport map, being a gradient of some convex function. According to suitable assumptions on the densities, it turns out that the optimal map belongs to a Sobolev space. This latter fact yields the strong convergence of the optimal maps (up to a subsequence) to get some map on the Wiener space. It remains to verify that this limit map is the optimal one.

Let $V : X \to \mathbb{R}$ be a measurable function such that e^{-V} is bounded. Consider

$$\mathcal{E}_V(F, F) = \int_X ||\nabla F||_{H\otimes K}^2 \, e^{-V} d\mu, \quad F \in \mathrm{Cylin}(X, K). \qquad (6.2.2)$$

It is well-known that if

$$\int_X |\nabla V|^2 \, e^{-V} d\mu < +\infty, \tag{6.2.3}$$

then the quadratic form (6.2.2) is closable over Cylin(X, K).

Now let $W : X \to \mathbb{R}$ be a measurable function such that the Poincaré Inequality holds true:

$$\int_X (f - \mathbb{E}_W(f))^2 \, e^{-W} d\mu \leq \int_X |\nabla f|^2 \, e^{-W} d\mu, \tag{6.2.4}$$

where $\mathbb{E}_W$ denotes the integral with respect to the measure $e^{-W}\mu$.

We will denote by $\mathbb{D}^p_k(X, K; e^{-V}\mu)$ the closure of Cylin(X, K) with respect to the norm defined in (2.1.9) replacing μ by $e^{-V}\mu$.

Theorem 6.2.1. *Let $V : X \to \mathbb{R}$ satisfies (6.2.3) and $W \in \mathbb{D}^2_2(X)$ satisfies (6.2.4) and such that*

$$\int_X e^{-V} \, d\mu = \int_X e^{-W} \, d\mu = 1.$$

Then there is a $\psi \in \mathbb{D}^2_1(X, e^{-W}\mu)$ such that $x \to S(x) = x + \nabla\psi(x)$ is the optimal transport map which pushes $e^{-W}\mu$ to $e^{-V}\mu$; moreover the inverse map of S is given by $x \to x + \eta(x)$ with $\eta \in L^2(X, H; e^{-V}\mu)$.

Proof. Let $\{e_n;\ n \geq 1\} \subset X^*$ be an orthonormal basis of H and set

$$H_n = \text{span}\{e_1, \ldots, e_n\}$$

the vector space spanned by $e_1, \ldots, e_n$, endowed with the induced norm of H. Let γ_n be the standard Gaussian measure on H_n. Denote

$$\pi_n(x) = \sum_{j=1}^{n} e_j(x) \, e_j.$$

Then π_n sends the Wiener measure μ to γ_n. Let $\mathcal{F}_n$ be the sub σ-field on X generated by π_n, and $\mathbb{E}(\ |\mathcal{F}_n)$ be the conditional expectation with respect to μ and to $\mathcal{F}_n$. Then we can write down

$$\mathbb{E}(e^{-W}|\mathcal{F}_n) = e^{-W_n} \circ \pi_n, \quad \mathbb{E}(e^{-V}|\mathcal{F}_n) = e^{-V_n} \circ \pi_n. \tag{6.2.5}$$

Note that for any $f \in L^1(H_n, \gamma_n)$,

$$\int_X f \circ \pi_n e^{-W} \, d\mu = \int_X f \circ \pi_n \, \mathbb{E}(e^{-W}|\mathcal{F}_n) \, d\mu = \int_{H_n} f e^{-W_n} d\gamma_n.$$

Applying (6.2.4) to $f \circ \pi_n$ yields

$$\int_{H_n} \Big(f - \int_{H_n} f e^{-W_n} d\gamma_n \Big)^2 e^{-W_n} d\gamma_n \leq \int_{H_n} |\nabla f|^2 e^{-W_n} d\gamma_n, \quad f \in C_b^1(H_n). \tag{6.2.6}$$

By Kantorovich dual representation 3.2.4, we have

$$W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n) = \sup_{(\psi,\varphi)\in\Phi_c} J(\psi,\varphi),$$

where

$$\Phi_c := \big\{ (\psi,\varphi) \in L^1(e^{-W_n}\gamma_n) \times L^1(e^{-V_n}\gamma_n); \varphi(y) - \psi(x) \leq |x-y|_{H_n}^2 \big\},$$

and

$$J(\psi,\varphi) := -\int_{H_n} \psi(x) e^{-W_n(x)} d\gamma_n(x) + \int_{H_n} \varphi(y)\, e^{-V_n(y)} d\gamma_n(y).$$

We know there exists a couple of functions (ψ_n, φ_n) in Φ_c, which can be chosen to be concave, such that $W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n) = J(\psi_n, \varphi_n)$. Now we prove the sequence $\{W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n)\}_{n\geq 1}$ is increasing, and converges to $W_2^2(e^{-W}\mu, e^{-V}\mu)$.
Let $q_n : W \times W \longrightarrow H_n \times H_n$ be defined as $q_n(x,y) = (\pi_n(x), \pi_n(y))$. If $\Pi_0 \in C(e^{-W}\mu, e^{-V}\mu)$ is an optimal coupling, then $(q_n)_\# \Pi_0$ is a coupling between $e^{-W_n}\gamma_n$, and $e^{-V_n}\gamma_n$, therefore we have:

$$\begin{aligned} W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n) &\leq \int_{H_n \times H_n} |x-y|^2 d(q_n)_\# \Pi_0(x,y) \\ &\leq \int_{W\times W} |x-y|_H^2 d\Pi_0(x,y) = W_2^2(e^{-W}\mu, e^{-V}\mu). \end{aligned}$$

Hence $\sup_{n\geq 1} W_2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n) \leq W_2(e^{-W}\mu, e^{-V}\mu)$.
Now consider a sequence of optimal couplings $(\Pi_0^n)_{n\geq 1}$ between the corresponding marginals $e^{-W_n}\gamma_n$ and $e^{-V_n}\gamma_n$. It is straightforward to see that the sequence $(W_2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n))_n$ is non decreasing, since for $m \leq n$, it holds $(q_m)_\# \Pi_0^n \in C(e^{-W_m}\gamma_m, e^{-V_m}\gamma_n)$.
By the previous work we can extract a weak cluster point Π_0 of the sequence. Because the function d_H is lower semi-continuous, we have:

$$\begin{aligned} \int_{X\times X} |x-y|_H^2 d\Pi_0(x,y) &\leq \liminf_n \int_{X\times X} |x-y|_H^2 d\Pi_0^n(x,y) \\ &\leq \sup_n \int_{X\times X} |x-y|_H^2 d\Pi_0^n(x,y) \\ &\leq W_2^2(e^{-W}\mu, e^{-V}\mu). \end{aligned}$$

As a consequence we get the result:

$$\lim_n W_2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n) = W_2(e^{-W}\mu, e^{-V}\mu).$$

Recall that $\Pi_0^n \in C(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n)$ is an optimal coupling, that is,

$$\int_{H_n\times H_n} |x-y|_{H_n}^2\, d\Pi_0^n(x,y) = W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n).$$

Then it holds true,

$$|x-y|_{H_n}^2 \geq \varphi_n(y) - \psi_n(x), \quad (x,y)\in H_n\times H_n, \tag{6.2.7}$$

and under Π_0^n:

$$|x-y|_{H_n}^2 = \varphi_n(y) - \psi_n(x). \tag{6.2.8}$$

Combining (6.2.7) and (6.2.8), Π_0^n is supported by the graph of $x \to x - \frac{1}{2}\nabla\psi_n(x)$ so that

$$\frac{1}{4}\int_{H_n} |\nabla\psi_n|^2\, e^{-W_n} d\gamma_n = W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n).$$

Now by (6.2.6), changing ψ_n to $\psi_n - \int_{H_n}\psi_n e^{-W_n}d\gamma_n$, then $\psi_n \in \mathbb{D}_1^2(e^{-W_n}\gamma_n)$ and

$$||\psi_n||_{\mathbb{D}_1^2(e^{-W_n}\gamma_n)}^2 \leq 2\int_{H_n} |\nabla\psi_n|^2\, e^{-W_n} d\gamma_n.$$

According to (6.2.1), we get that $\sup_{n\geq 1} ||\psi_n||_{\mathbb{D}_1^2(e^{-W_n}\gamma_n)}^2 < +\infty$. Now consider $\tilde{\psi}_n = \psi_n \circ \pi_n$, $\tilde{\varphi}_n = \varphi_n \circ \pi_n$. Then

$$\sup_{n\geq 1} ||\tilde{\psi}_n||_{\mathbb{D}_1^2(e^{-W}\mu)} < +\infty. \tag{6.2.9}$$

As in [36], define $F_n(x,y) = d_H(x,y)^2 + \tilde{\psi}_n(x) - \tilde{\varphi}_n(y)$, which is non negative according to (6.2.7). Let Π_0 be an optimal coupling between $e^{-W}\mu$ and $e^{-V}\mu$. We have

$$\begin{aligned}\int_{X\times X} F_n(x,y)\Pi_0(dx,dy) &= W_2^2(e^{-W}\mu, e^{-V}\mu) + \int_X \tilde{\psi}_n(x) e^{-W} d\mu - \int_X \tilde{\varphi}_n(y)\, e^{-V} d\mu \\ &= W_2^2(e^{-W}\mu, e^{-V}\mu) + \int_{H_n} \psi_n(x) e^{-W_n} d\gamma_n - \int_{H_n} \varphi_n(y)\, e^{-V_n} d\gamma_n \\ &= W_2^2(e^{-W}\mu, e^{-V}\mu) - W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n)\end{aligned} \tag{6.2.10}$$

which tends to 0 as $n \to +\infty$. Now returning to (6.2.9), by Banach-Saks theorem, up to a subsequence, the Cesaro mean $\frac{1}{n}\sum_{j=1}^n \tilde{\psi}_j$ converges to $\hat{\psi}$ in $D_1^2(e^{-W}\mu)$. Therefore

$$\frac{1}{n}\sum_{j=1}^n \tilde{\varphi}_n(y) = d_H^2(x,y) + \frac{1}{n}\sum_{j=1}^n \tilde{\psi}_j(x) - \frac{1}{n}\sum_{j=1}^n F_j(x,y)$$

which converges in L^1 to $\hat{\varphi}(y) = d_H^2(x,y) + \hat{\psi}(x)$. Now define

$$\psi = \varliminf_{n\to+\infty} \frac{1}{n}\sum_{j=1}^n \tilde{\psi}_j, \quad \varphi = \varliminf_{n\to+\infty} \frac{1}{n}\sum_{j=1}^n \tilde{\varphi}_j.$$

Then $\psi = \hat{\psi}$ for $e^{-W}\mu$ almost all, $\varphi = \hat{\varphi}$ for $e^{-V}\mu$ almost all, and by (6.2.7), it holds that

$$\varphi(y) - \psi(x) \le d_H^2(x,y), \quad (x,y) \in X \times X. \tag{6.2.11}$$

Also by above construction, under Π_0

$$\varphi(y) - \psi(x) = d_H^2(x,y). \tag{6.2.12}$$

Denote by Θ_0 the subset of (x,y) satisfying (6.2.12). On the other hand, the fact that $\psi \in \mathbb{D}_1^2(e^{-W}\mu)$ implies that for any $h \in H$, there is a full measure subset $\Omega_h \subset X$ such that for $x \in \Omega_h$, there is a sequence $\varepsilon_j \downarrow 0$ such that

$$\langle \nabla\psi(x), h\rangle_H = \lim_{j\to+\infty} \frac{\psi(x+\varepsilon_j h) - \psi(x)}{\varepsilon_j}.$$

Let D be a countable dense subset of H. Then there exists a full measure subset Ω such that for each $x \in \Omega$, for any $h \in D$, there is a sequence $\varepsilon_j \downarrow 0$ such that

$$\langle \nabla\psi(x), h\rangle_H = \lim_{j\to+\infty} \frac{\psi(x+\varepsilon_j h) - \psi(x)}{\varepsilon_j}.$$

Set $\Theta = (\Omega \times X) \cap \Theta_0$. Then $\Pi_0(\Theta) = 1$. For each couple $(x,y) \in \Theta$, we have $\varphi(y) - \psi(x) = d_H^2(x,y)$ and $\varphi(y) - \psi(x+\varepsilon_j h) \le d_H^2(x+\varepsilon_j h, y)$. Because $x - y \in H$ Π_0-a.a. it follows that

$$\psi(x+\varepsilon_j h) - \psi(x) \ge 2\varepsilon_j \langle h, x-y\rangle_H + \varepsilon_j^2 |h|_H^2.$$

Therefore $\langle \nabla\psi(x), h\rangle_H \ge 2\langle x-y, h\rangle_H$ for any $h \in D$. From which we deduce that

$$y = x - \frac{1}{2}\nabla\psi(x), \tag{6.2.13}$$

and Π_0 is supported by the graph of $x \to S(x) = x - \frac{1}{2}\nabla\psi(x)$. Replacing $-\frac{1}{2}\psi$ by ψ, we get the statement of the first part of the theorem. For the second part, we refer to section 4 in [37]. □

For the use of Chapter 7, we emphaze that the above constructed whole sequence

$$\tilde{\varphi}_n \to \varphi \text{ in } L^1(e^{-V}\mu). \tag{6.2.14}$$

In fact, if $\tilde{\psi}$ is another cluster point of $\{\tilde{\psi}_n; n \geq 1\}$ for the weak topology of $\mathbb{D}_1^2(e^{-W}\mu)$, then under the optimal plan Π_0, the relation (6.2.13) holds for $\tilde{\psi}$. Therefore $\nabla\psi = \nabla\tilde{\psi}$ almost everywhere for $e^{-W}\mu$; it follows that $\psi = \tilde{\psi}$, since $\int_X \psi e^{-W} d\mu = \int_X \tilde{\psi}\, e^{-W} d\mu = 0$. Now note that

$$\begin{aligned}\int_X |\nabla\tilde{\psi}_n|_H^2 e^{-W} d\mu &= \int_{H_n} |\nabla\psi_n|_{H_n}^2 e^{-W_n} d\gamma_n = W_2^2(e^{-W_n}\gamma_n, e^{-V_n}\gamma_n)\\ &\to W_2^2(e^{-W}\mu, e^{-V}\mu) = \int_X |\nabla\psi|_H^2 e^{-W} d\mu.\end{aligned}$$

Combining these two points, we see that $\tilde{\psi}_n$ converges to ψ in $\mathbb{D}_1^2(e^{-W}\mu)$. By (6.2.10), the sequence $\tilde{\varphi}_n$ converges to φ in $L^1(e^{-V}\mu)$. □

Let us make a few comments about the assumption of W. A sufficient condition for that (6.2.4) holds is when $W \in \mathbb{D}_2^2(X)$ safisfies

$$\nabla^2 W \geq -c\,\mathrm{Id}, \quad c \in [0, 1). \tag{6.2.15}$$

Indeed thanks to the Proposition 2.3.3, (6.2.15) implies the following logarithmic Sobolev inequality

$$(1-c)\int_X \frac{|f|}{||f||_{L^2(e^{-W}\mu)}}\, e^{-W} d\mu \leq \int_X |\nabla f|^2\, e^{-W} d\mu, \quad f \in \mathrm{Cylin}(X). \tag{6.2.16}$$

It is also known (see for example [61]) that (6.2.16) is stronger than Poincaré inequality

$$(1-c)\int_X (f - \mathbb{E}_W(f))^2\, e^{-W} d\mu \leq \int_X |\nabla f|^2\, e^{-W} d\mu, \tag{6.2.17}$$

where $\mathbb{E}_W$ denotes the integral with respect to the measure $e^{-W}\mu$.

6.3 On the Wiener space with a Sobolev type norm

Let X be the classical Wiener space. Recall the pseudo-distance $\|.\|_{k,\gamma}$ is defined as:

$$\|w\|_{k,\gamma} := \left(\int_0^1 \int_0^1 \frac{(w(t) - w(s))^{2k}}{|t-s|^{1+2k\gamma}} dtds\right)^{1/2k}.$$

Here the notion of *optimal coupling* will be refer to this cost, namely minimizers Π of

$$\int_{X\times X} \|x-y\|_{k,\gamma}^p d\Pi(x,y), \tag{6.3.1}$$

where $p \geq 1$. with p is a constant greater than 1. We consider $\hat{X} := \{x \in X;\ \|x\|_{k,\gamma} < \infty\}$. For a sake of simplicity, we still denote $X = \hat{X}$ and all measures below will be Borel with respect to the induced topology.

In Chapter 2, we have seen that $\|.\|_{k,\gamma}$ is a strictly convex and differentiable (Lemma 2.2.1) norm.

Among many methods to solve the Monge Problem, there is a direct one: it is related to the existence of Kantorovich potentials (see Proposition 3.2.4) and to solve y in function of x through the following system ($c(x,y) := \|x-y\|_{k,\gamma}^p$):

$$\begin{cases} \phi^c(y) - \phi(x) = c(x,y) & \Pi - \text{almost everywhere}, \\ \phi^c(y) - \phi(x) \leq c(x,y) & \text{everywhere}. \end{cases}$$

As it is explained in Villani's book [58], this system can be solved directly when the cost c and the potential ϕ are differentiable, as soon as $\nabla_x c(x,.)$ is injective, namely c satisfies Twist condition. It is the case when $p > 1$. But the method fails when p equals to 1. In the latter case we can focus on another strategy, developped in a recent paper of Cavalletti [19]. The author solves the Monge Problem in an abstract Wiener space where the cost is the Cameron-Martin norm (without any power). It turns out that the classical Wiener space endowed with the norm $\|.\|_{k,\gamma}$ enjoys similar properties, that we can employ here.

6.3.1 $c(x,y) = \|x-y\|_{k,\gamma}^p$ when $p > 1$

When $p > 1$, the cost $c(x,y) = \|x-y\|_{k,\gamma}^p$ is a strictly convex function. Since c is differentiably we get the injectivity of $\nabla_x c(x,.)$. Compared with the next section we lose the $H-$Lipschitz property of $c-$convex functions. Indeed for any $H-$Lipschitz function φ, we write:

$$\begin{aligned} |\varphi(x) - \varphi(y)| &\leq |\|x-\xi\|_{k,\gamma}^p - \|y-\xi\|_{k,\gamma}^p| \\ &\leq \|x-y\|_{k,\gamma} M_\xi, \end{aligned}$$

where the constant M_ξ depends on ξ and is not necessarly bounded. However we will see that in this case $c-$convex functions (hence potentials) are *locally* $H-$Lipschitz. Since differentiability is a local property, we should apply the Rademacher theorem.

We follow Fathi and Figalli in [33] to obtain that $c-$convex functions are locally Lipschitz with respect to $\|.\|_{k,\gamma}$. The key argument is that the sup of a family of uniformly $\|.\|_{k,\gamma}-$Lipschitz functions, is also $\|.\|_{k,\gamma}-$Lipschitz.
The interest of the following proof is that the method is direct: one does not need to pass by finite dimensional approximations.

Theorem 6.3.1. *Let ρ_0 and ρ_1 be two probability measures on X, and such that the first one is absolutely continuous with respect to the Wiener measure μ. Assume (6.3.1) is finite for some coupling $\Pi \in C(\rho_0, \rho_1)$.*
Then there exists a unique optimal coupling between ρ_0 and ρ_1 relatively to the cost c. Moreoever it is concentrated on a graph of some Borel map $T : X \longrightarrow X$ unique up to a set of zero measure for μ.

Proof. Let $\Pi_0 \in C(\rho_0, \rho_1)$ be an optimal coupling for c. We shall show that Π_0 is concentrated on a graph of some Borel map. It is well known (Proposition 3.2.4) that under the assumption of the theorem, since Π_0 is concentrated on a $\sigma-$compact Γ (by inner regularity) set which is $c-$cyclically monotone, there is a $c-$convex map $\varphi : X \longrightarrow \mathbb{R}$ (so-called Kantorovich potential) such that

$$\varphi^c(y) - \varphi(x) = \|x - y\|^p_{k,\gamma} \quad \forall (x, y) \in \Gamma.$$

Moreover from the definition of $c-$convexity, we also have

$$\varphi^c(y) - \varphi(x) \le \|x - y\|^p_{k,\gamma} \quad \forall (x, y) \in X \times X. \tag{6.3.2}$$

Since φ^c is finite everywhere, if we consider subsets $W_n := \{\varphi^c \le n\}$ for $n \in \mathbb{N}$ then:

$$W_n \subset W_{n+1} \text{ and } \bigcup_{n\in\mathbb{N}} W_n = X.$$

Our cost $c(., y) = \|. - y\|^p_{k,\gamma}$ is locally $\|.\|_{k,\gamma}-$Lispchitz locally uniformly in y, that is, for any $R > 0$, there is a constant $L_R > 0$ such that

$$|c(z_1, y) - c(z_2, y)| \le L_R \, ||z_1 - z_2||_{k,\gamma} \quad \text{for } z_1, z_2, y \in B(0, R),$$

where $B(0, R)$ is the ball of radius R for the norm $||.||_{k,\gamma}$. Hence for each $y \in X$ there exists a neighborhood E_y of y such that $(\|. - z\|^p_{k,\gamma})_{z\in E_y}$ is a uniform family of locally $\|.\|_{k,\gamma}-$Lipschitz functions, the local Lipschitz constant being independent of $z \in E_y$. Moreover by separability, we can find a sequence $(y_l)_{l\in\mathbb{N}}$ of elements of X such that:

$$\bigcup_{l\in\mathbb{N}} E_{y_l} = X.$$

Now consider increasing subsets of X:

$$V_n := W_n \bigcap (\bigcup_{l=1}^{n} E_{y_l}).$$

We can define maps approximating φ as follow:

$$\begin{aligned} \varphi_n : X &\longrightarrow X \\ x &\longmapsto \sup_{y \in V_n} \left(\varphi^c(y) - \|x-y\|^p_{k,\gamma} \right). \end{aligned}$$

Notice that

$$\varphi_n(x) = \max_{l=1,\dots,n} \sup_{y \in W_n \cap E_{y_l}} \left(\varphi^c(y) - \|x-y\|^p_{k,\gamma} \right).$$

But since $\varphi^c \leq n$ on W_n, φ_n is also bounded from above by n. Therefore the sequence $(\varphi^c(y) - \|. - y\|^p_{k,\gamma})_{y \in W_n \cap E_{y_l}}$ is uniformly locally $\|.\|_{k,\gamma}-$Lipschitz and bounded from above. Finally φ_n being a maximum of uniformly locally $\|.\|_{k,\gamma}-$Lipschitz functions, is locally $\|.\|_{k,\gamma}-$Lispchitz as well. We can extend φ_n to a $\|.\|_{k,\gamma}-$ Lipschitz function everywhere on X still denoted by φ_n. By (2.2.2), we get:

$$|\varphi_n(w+h) - \varphi_n(w)| \leq C\|h\|_{k,\gamma} \leq 2C|h|_H \quad \forall w \in X, \ \forall h \in H.$$

In other words φ_n is a $H-$Lipschitz function. Thanks to Rademacher theorem on the Wiener space (see [27]), there exists a Borel subset F_n of X with full $\mu-$(hence ρ_0-)measure such that for all $x \in F_n$, φ_n is differentiable at x along all directions in H. Then for each $x \in F := \cap_n F_n$ (which has also full ρ_0-measure), each φ_n is differentiable at x.

By the increasing of $(V_n)_n$, it is clear that $\varphi_n \leq \varphi_{n+1} \leq \varphi$ everywhere on X. Moreover with same arguments as in [33], if $C_n := P_1(\Gamma \cap (X \times V_n))$, then $\varphi_{|C_n} = \varphi_{n|C_n} = \varphi_{l|C_n}$ for all $l \geq n$ and all $n \in \mathbb{N}$. Fix $x \in C_n \cap F$. By definition of C_n it exists $y_x \in V_n$ such as:

$$\begin{aligned} \varphi^c(y_x) - \varphi_n(x) &= \|x - y_x\|^p_{k,\gamma}, \\ \text{or } \varphi^c(y_x) - \varphi(x) &= \|x - y_x\|^p_{k,\gamma}. \end{aligned}$$

Subtracting (6.3.2) with (x', y_x) to the previous equality, we get for all $x' \in X$ and $h \in H$:

$$\varphi(x') - \varphi(x) \geq \|x - y_x\|^p_{k,\gamma} - \|x' - y_x\|^p_{k,\gamma}.$$

Taking $x' = x + \varepsilon h$ with $\varepsilon > 0$, $h \in H$, dividing by ε and letting ε tend to 0, we get

$$\langle \nabla\varphi(x), h\rangle_H \geq -\langle \nabla_x c(x, y_x), h\rangle_H.$$

By linearity in h:

$$\nabla\varphi(x) - \nabla_x c(x, y_x) = 0. \tag{6.3.3}$$

Indeed $c(., y_x)$ is differentiable at x thanks to Lemma 2.2.1. The strict convexity of $c(x, y) = \|x - y\|_{k,\gamma}^p$ yields $\nabla_x c(x, .)$ is injective and (6.3.3) gives:

$$y_x = (\nabla_x c(x, .))^{-1}(\nabla\varphi(x)) =: T(x),$$

where $(\nabla_x c(x, .))^{-1}$ is the inverse of the map $y \longmapsto \nabla_x c(x, y)$. Notice here that T is uniquely determined. We deduce that $\Gamma \cap (X \times V_n)$ is the graph of the map T over $C_n \cap F$ for all $n \in \mathbb{N}$. But $(C_n)_n$ and $(V_n)_n$ are increasing and such that $\bigcup_n V_n = X$. Therefore Γ is a graph over $P_1(\Gamma) \cap F$ with $P_1(\Gamma) = \bigcup_n C_n$.
We can extend T onto a measurable map over X as it is explained in [33]. We obtain Γ is included in the graph of a measurable map T, unique up to a set of ρ_0-measure. In other words $\Pi_0 = (id \times T)_{\#}\rho_0$.
We have proved that any optimal coupling is carried by a graph of some map. So if $\Pi_1, \Pi_2 \in C(\rho_0, \rho_1)$ are optimal for $\|.\|_{k,\gamma}$ then any convex combination of Π_1 and Π_2 is also optimal. Take $\Pi := \frac{1}{2}(\Pi_1 + \Pi_2)$ be an optimal coupling between ρ_0 and ρ_1: there exists some measurable map T such that $\Pi = (Id \times T)_{\#}\rho_0$. Let f be the density of Π_1 with respect to Π. Then for any continuous bounded functions φ we have:

$$\begin{aligned}\int_X \varphi(x)d\rho_0(x) &= \int_{X\times X} \varphi(x)d\Pi_1(x, y)\\ &= \int_{X\times X} \varphi(x)f(x, y)d\Pi(x, y)\\ &= \int_X \varphi(x)f(x, T(x))d\rho_0(x).\end{aligned}$$

This yields $f(x, T(x)) = 1$ ρ_0-a.e., hence $f = 1$ $\Pi-$a.e. It leads to $\Pi = \Pi_1$ and finally $\Pi_2 = \Pi_1 = (Id \times T)_{\#}\rho_0$. ☐

6.3.2 $c(x, y) = \|x - y\|_{k,\gamma}$

When $p = 1$, $c(x, y) = \|x - y\|_{k,\gamma}$. Hence if a map φ is $c-$convex then it is $1-$Lipschitz, hence $H-$Lipschitz. Indeed:

$$|\varphi(x + h) - \varphi(x)| \leq \|h\|_{k,\gamma} \leq C_{k,\gamma}|h|_H, \quad \forall h \in H \;\; \forall x \in X.$$

Therefore we can use Rademacher theorem [27] on the Wiener space, to differentiate any $H-$Lipschitz functions. But the difficulty in this case is that the cost, being a norm, is *not strictly convex*, so we lose the injectivity of the map $y \longmapsto \nabla_x c(x, y)$. The method used in the first section requires the differentiation theorem for the Wiener measure, which is not available.

We will follow the method of [19] developed by Bianchini and Cavalletti in [10]. The method uses a *selection theorem*. By strict convexity of our norm $\|.\|_{k,\gamma}$ proved in Lemma 2.2.1, $(X, \|.\|_{k,\gamma})$ is a geodesic non branching space.

We will not develop fully the method but ony briefly indicate the different steps :

1. reduce the initial Monge-Kantorovich Problem to the one-dimensional Monge-Kantorovich Problem along distinct geodesics : this is possible since the space is non-branching.
2. verify that the conditional measures provided by disintegration of both measures ρ_0 and ρ_1 on each geodesic have no atom: this is possible thanks to properties of Gaussian measure. The aim is to get one optimal map on each geodesic.
3. piece obtained maps together to get a transport map for the initial Monge Problem by a general selection theorem.

We refer to [19] and [10] for more details. In our case, the cost $\|.\|_{k,\gamma}$ is smooth enough (continuity) to guarantee the existence of a Kantorovich potential φ (Proposition 3.2.4) such that there is a $\sigma-$compact subset Γ on which any optimal coupling Π is concentrated and

$$\Gamma := \{(x, y) \in X \times X;\ \varphi^c(y) - \varphi(x) = \|x - y\|_{k,\gamma}\}.$$

From now, let us consider an *optimal* (relative to the cost $c(x, y) = \|x - y\|_{k,\gamma}$) coupling Π_0 between two probability measures ρ_0 and ρ_1 on X, both absolutely continuous with respect to the Wiener measure μ. Let $\pi_n : X \to V_n$ be the finite dimensional projection, where V_n is a space of functions piecewisely linear, described in Chapter 2. Denote by $\rho_0^n := (\pi_n)_\#\rho_0$ and $\rho_1^n := (\pi_n)_\#\rho_1$, which are absolutely continuous with respect to the Gaussian measure γ_n on V_n. Since the restriction of $||.||_{k,\gamma}$ on V_n is differentiable out of 0, by a result due to Caffarelli, M. Feldman, and R.J. McCann [16], there is an optimal map $T^n : V_n \to V_n$ such that $\Pi_0^n := (id \times T^n)_\#\rho_0^n$ is the unique optimal couplage between ρ_0^n and ρ_1^n. In other words, Π_0^n is concentrated on some Borel set $\Gamma_n \subset Graph(T^n)$.

The following result shows that the method of [19] really works well.

Proposition 6.3.2. *Assume that there exists $M > 0$ such that densities f_0 and f_1 of respectively ρ_0^n and ρ_1^n are bounded by M. Then the following estimate holds true for all Borel subset $A \subset V_n$:*

$$\gamma_n(T_{n,t}(A)) \geq \frac{1}{M}\rho_0^n(A) \qquad \forall t \in [0, 1],$$

where $T_{n,t} := (1-t)Id + tT^n$.

We will follow the proof of [19]. The only difference is to consider Monge maps for the cost induced by $\|.\|_{k,\gamma}^p$ with ($p > 1$), instead of $|.|_H^p$. Indeed costs $\|.\|_{k,\gamma}^p$ satisfy conditions of Proposition 3.4.4, so that the associated optimal maps T_p^n are approximately differentiable.

Proof. Fix $p > 1$. Since ρ_0^n is absolutely continuous w.r.t. $\gamma_n := (\pi_n)_\# \mu$, by Proposition 3.4.2, the Monge Problem

$$\inf_{T\#\rho_0^n = \rho_1^n} \int_X \|x - T(x)\|_{k,\gamma}^p d\rho_0^n(x),$$

admits a unique solution T_p. Besides by Proposition 3.4.4, T_p is approximately differentiable ρ_0^n-a.s, and by Lemma 3.4.5,

$$f_0^n(x)\frac{1}{\sqrt{2\pi}}e^{-|x|^2/2} = f_1^n(T_p(x))|det(\tilde{\nabla}T_p(x))e^{-\frac{1}{2}|T_p(x)|^2/2}.$$

Besides $|det(\tilde{\nabla}T_p(x))| > 0$ and $f_1^n(T_p(x)) > 0$ for ρ_0^n-a.e. $x \in \mathbb{R}^n$. Hence we can write ρ_0^n-a.s.

$$|det(\tilde{\nabla}T_p(x))| = \frac{f_0^n(x)}{f_1^n(T_p(x))}\exp\left\{-\frac{1}{2}(|x|^2 - |T_p(x)|^2)\right\}.$$

Now consider $T_{p,t} := (1-t)Id + tT_p$. By the same arguments that in the proof of Proposition 4.2.1 and by the concavity of $t \longmapsto det((1-t)Id + tD)^{1/n}$, it holds

$$\log\left(det(\tilde{\nabla}T_{p,t}(x))^{1/n}\right) \geq t\log\left(det(\tilde{\nabla}T_p(x))^{1/n}\right).$$

Therefore:

$$det(\tilde{\nabla}T_{p,t}(x)) \geq |det(\tilde{\nabla}T_p(x))|^t = \left(\frac{f_0^n(x)}{f_1^n(T_p(x))}\right)^t \exp\left\{-\frac{t}{2}(|x|^2 - |T_p(x)|^2)\right\}.$$

Following [19], for any $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\begin{aligned}
\gamma_n(T_{p,t}(A)) &= \int_A det(\tilde{\nabla}T_{p,t}(x))\exp\left\{-\frac{1}{2}|T_{p,t}(x)|^2\right\}dx \\
&= \int_A det(\tilde{\nabla}T_{p,t}(x))\exp\left\{-\frac{1}{2}\left(|T_{p,t}(x)|^2 - |x|^2\right)\right\}d\gamma_n(x) \\
&\geq \int_A \left(\frac{f_0^n(x)}{f_1^n(T_p(x))}\right)^t \exp\left\{\frac{1}{2}\|x - T_p(x)\|^2(t-t^2)\right\}d\gamma_n(x) \\
&\geq \frac{1}{M^t}\int_A f_0^n(x)^t d\gamma_n(x) = \frac{1}{M^t}\int_A f_0^n(x)^{t-1}d\rho_0^n(x) \geq \frac{1}{M}\rho_0^n(A).
\end{aligned}$$

Since $(Id \times T_p)_\# \rho_0^n$ converges weakly to $(Id \times T^n)_\# \rho_0^n$, letting $p \to 1$, proceeding as in [19], or in [8], we obtain

$$\gamma_n(T_{n,t}(A)) \geq \frac{1}{M} \rho_0^n(A).$$

□

Let $T_t(x, y) = (1 - t)x + ty$. Then the above result can be reformulated by

$$\gamma_n(T_t(\Gamma_n \cap (A \times V_n))) \geq M \rho_0^n(A).$$

Coming back to the Wiener space, we have the following result:

Proposition 6.3.3. *Assume that the density of ρ_o and ρ_1 with respect to μ are bounded by $M > 0$; then for any compact subset $A \subset X$, we have:*

$$\mu(T_t(\Gamma \cap A \times X)) \geq M \rho_0(A).$$

The proof, given again in [19], holds true in a quite general setting, provided the cost is at least lower semi-continuous. Again following [19] step by step, we get the following result.

Theorem 6.3.4. *Let ρ_0 and ρ_1 be two probability measures on X of finite entropy. Then there exists an optimal coupling between ρ_0 and ρ_1 which is concentrated on a graph of some Borel map $T : X \longrightarrow X$.*

Note that by Young inequality

$$||x||_{k,\gamma}^2 f_0(x) \leq e^{\alpha ||x||_{k,\gamma}^2} + \frac{f_0(x)}{\alpha} \log(\frac{f_0(x)}{\alpha}),$$

we get

$$\int_X ||x||_{k,\gamma}^2 f_0(x) \, d\mu(x) \leq \int_X e^{\alpha ||x||_{k,\gamma}^2} \, d\mu(x) + \text{Ent}_\mu(\rho_0/\alpha),$$

which is finite if $\text{Ent}_\mu(\rho_0) < +\infty$, since by Fernique's theorem

$$\int_X e^{\alpha ||x||_{k,\gamma}^2} \, d\mu(x) < +\infty$$

for α small enough. Therefore any probability measure in $D(Ent_\mu)$ has finite second moment with respect to $||.||_{k,\gamma}$.

□

Chapter 7

Monge-Ampère equation on Wiener spaces

Let ρ_0 and ρ_1 be two probability measures on $\mathbb{R}^n$. Throughout all this part, when we talk about *optimal* map, we always refer to optimality with respect to the cost being the *square of the Euclidian norm*, that is:

$$c(x,y) = |x-y|^2.$$

If ρ_0 is absolutely continuous with respect to the Lebesgue measure, Brenier's theorem gives us the (unique) optimal transport map $T = \nabla\Phi$ which is the gradient of some convex function Φ. In addition we have the characterization of the optimal map, namely if $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex and is such that $(\nabla\Phi)_{\#}\rho_0 = \rho_1$, then $T := \nabla\Phi$ is necessarly the optimal map between ρ_0 and ρ_1, that is minimizing the quantity

$$\int_{\mathbb{R}^n} |x - T(x)|^2 d\rho_0(x),$$

among all maps $S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ such that $S_{\#}\rho_0 = \rho_1$.

When both ρ_0 and ρ_1 are absolutely continuous, with respective densities say f_0 and f_1, the *preserving mass* condition $T_{\#}\rho_0 = \rho_1$ is equivalent (at least formally) to the fully nonlinear partial derivative equation:

$$f_0(x) = f_1(T(x))|det(\nabla T(x))| \quad \text{a.s.}$$

This is the so called *Monge-Ampère equation*. It corresponds to the change of variables formula, and the result was proved first by McCann in [50].
Thanks to the characterization of the optimal map (see Brenier's Theorem in Chapter 3), any convex solution $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ of

$$f_0(x) = f_1(\nabla\Phi(x))det(\nabla^2\Phi(x)), \tag{7.0.1}$$

induces the optimal map, letting $T := \nabla\Phi$. Conversely the optimal map $T = \nabla\Phi$ is such that Φ solves (7.0.1).

The regularity of solutions of Monge-Ampère equation has been intensively studied: in $\mathbb{R}^n$ we can cite Caffarelli around 90's ([15]), and more recently De Philippis and Figalli ([26] or [25]), in the Wiener space by Feyel and Üstünel ([37]), Bogachev and Kolesnikov ([13] and [45]). Therefore it relies to the regularity of the optimal transport maps. Our purpose is to extend results of those latter, and construct a strong solution to Monge-Ampère equation on an abstract Wiener space.

In order to pass on the Wiener space, we consider measures absolutely continuous with respect to the standard Gaussian measure, that we will be denote by γ in $\mathbb{R}^n$ for all the sequel. So let be the optimal

$$\nabla\Phi_{\#} : e^{-V}\gamma \longrightarrow e^{-W}\gamma.$$

The corresponding Monge-Ampère equation becomes

$$e^{-V(x)-\frac{|x|^2}{2}} = e^{-W(\nabla\Phi(x))-\frac{|\nabla\Phi(x)|^2}{2}} det(\nabla^2\Phi(x)).$$

Because the determinant makes no sense in infinite dimension, we deal with det_2 the *FredholmCarleman determinant* defined by:

$$det_2(I+K) := \prod_{i=1}^{\infty}(1+k_i)e^{-k_i},$$

for any K a symmetric HilbertSchmidt operator with eigenvalues k_i.

Now let (X, H, μ) be an abstract Wiener space and $e^{-V}\mu,\ e^{-W}\mu \in \mathcal{P}(X)$ two probability measures absolutely continuous with respect to the Wiener measure μ. Our main result is the following (see Theorem 7.2.1):

Theorem. *If $V \in \mathbb{D}_1^2(X)$ and $W \in \mathbb{D}_2^2(X)$ satisfy*

$$0 < \delta_1 \leq e^{-V} \leq \delta_2, \quad e^{-W} \leq \delta_2, \quad \nabla^2 W \geq -cId, \quad c \in [0,1),$$

then there exists a function $\varphi \in \mathbb{D}_2^2(X)$ such that $x \to x + \nabla\varphi(x)$ pushes $e^{-V}\mu$ to $e^{-W}\mu$ and solves the Monge-Ampère equation

$$e^{-V} = e^{-W(T)}e^{\mathcal{L}\varphi-\frac{1}{2}|\nabla\varphi|^2}\, det_2(\mathrm{Id}_{H\otimes H} + \nabla^2\varphi),$$

where $T(x) = x + \nabla\varphi(x)$, and $\mathcal{L}$ is the Ornstein-Uhlenbeck operator.

It includes two special cases:

- One studied in [37] where the source measure is the Wiener measure and the target measure is $H-$log concave:

$$e^{-V}\mu = \mu \quad \text{and} \quad W \text{ is } H \text{ convex.}$$

- Another one in [13] where the source measure has its Fisher's information finite, and the target measure is the Wiener measure:

$$\int_X |\nabla V|^2 e^{-V} d\mu < \infty \quad \text{and} \quad e^{-W}\mu = \mu.$$

We can not tell from the previous situation if T is the optimal map. The assumptions are in fact too weak. Nevertheless we can reinforce them to get the optimal map. This is the aim of Theorem 7.1.6.
Besides, we prove that the map S constructed in Section 6.2, admits an inverse map T which is $T(x) = x + \nabla\varphi(x)$ with $\varphi \in \mathbb{D}_2^2(X)$ (see Theorem 7.2.2).

To this end, thanks to dimension free inequalities obtained in Chapter 5 Section 5.3, we get new results in finite dimension. More specifically we obtain the following result (Theorem 7.1.2) which will be a key ingredient for our purpose:

Theorem. *If $V \in \mathbb{D}_1^2(\mathbb{R}^n, \gamma)$ and $W \in \mathbb{D}_2^2(\mathbb{R}^n, \gamma)$ satisfy*

$$e^{-V} \leq \delta_2, \quad e^{-W} \leq \delta_2, \quad \nabla^2 W \geq -cId, \quad c \in [0, 1),$$

then $\mathcal{L}\varphi$ exists in $L^1(\mathbb{R}^n, e^{-V}d\gamma)$ and the optimal map $\nabla\Phi(x) = x+\nabla\varphi(x)$ between $e^{-V}\gamma$ and $e^{-W}\gamma$ solves the Monge-Ampère equation

$$e^{-V} = e^{-W(\nabla\Phi)} e^{\mathcal{L}\varphi - \frac{1}{2}|\nabla\varphi|^2} \det{}_2(\mathrm{Id} + \nabla^2\varphi).$$

Let's begin with finite dimension case.

7.1 Monge-Ampère equations in finite dimension

Let $e^{-V}\gamma, e^{-W}\gamma \in \mathcal{P}(\mathbb{R}^n)$. The main assumptions made in this section are the following:

$$e^{-V} \leq \delta_2, \quad e^{-W} \leq \delta_2, \quad \nabla^2 W \geq -cId, \quad c \in [0, 1). \tag{7.1.1}$$

Besides we sometimes assume

$$(H) \qquad 0 < \delta_1 \leq e^{-V}.$$

With the condition (H) we get a first result (Theorem 7.1.1), using the same techniques as in Chapter 5, Section 5.3. For the sequel we would like to remove the condition (H). It will be possible thanks to the Theorem 5.3.10, which provides us a dimension free inequality.

Theorem 7.1.1. *Let* $V \in \mathbb{D}_1^2(\mathbb{R}^n, \gamma)$ *and* $W \in \mathbb{D}_2^2(\mathbb{R}^n, \gamma)$ *satisfying conditions* (7.1.1) *and* (H). *Then the optimal transport map* $x \to x + \nabla\varphi(x)$ *from* $e^{-V}\gamma$ *to* $e^{-W}\gamma$ *solves the following Monge-Ampère equation*

$$e^{-V} = e^{-W(\nabla\Phi)} e^{\mathcal{L}\varphi - \frac{1}{2}|\nabla\varphi|^2} \det{}_2(\mathrm{Id} + \nabla^2\varphi), \tag{7.1.2}$$

where $\nabla\Phi(x) = x + \nabla\varphi(x)$.

Proof. Let V_m, W_m be the approximating sequences considered in Chapter 4, Section 1.2. that are:

$$V_m = \chi_m\, P_{\frac{1}{m}} V + \log \int_{\mathbb{R}^n} e^{-\chi_m\, P_{\frac{1}{m}} V} d\gamma\,, \quad W_m = P_{\frac{1}{n}} W + \log \int_{\mathbb{R}^n} e^{-P_{\frac{1}{n}} W}\, d\gamma,$$

where $P_{\frac{1}{m}}$ is the Ornstein-Uhlenbeck semi group at time $\frac{1}{m}$, $\chi_m \in C_c^\infty(\mathbb{R}^n)$ is a smooth function with compact support satisfying usual conditions: $0 \le \chi_m \le 1$ and

$$\chi_m(x) = 1 \text{ if } |x| \le m, \quad \chi_m(x) = 0 \text{ if } |x| \ge m+2, \quad \sup_{m \ge 1} ||\nabla\chi_m||_\infty \le 1.$$

Then

$$e^{-V_m} = e^{-W_m(\nabla\Phi_m)} e^{\mathcal{L}\varphi_m - \frac{1}{2}|\nabla\varphi_m|^2} \det{}_2(\mathrm{Id} + \nabla^2\varphi_m), \tag{7.1.3}$$

where $\nabla\Phi_m(x) = x + \nabla\varphi_m(x)$ is the optimal mal pushing $e^{-V_m}\gamma$ forward to $e^{-W_m}\gamma$. In order to pass to the limit in (7.1.3), we have to prove the convergence of $\mathcal{L}\varphi_m$ to $\mathcal{L}\varphi$, and $W_m(\nabla\Phi_m)$ to $W(\nabla\Phi)$. By (5.3.35)-(5.3.37), we see that for any $1 < p < 2$, up to a subsequence

$$\lim_{m \to +\infty} ||\varphi_m - \varphi||_{\mathbb{D}_2^p(\gamma)} = 0.$$

Now by Meyer inequality for Gaussian measure (see [48]),

$$\int_{\mathbb{R}^n} |\mathcal{L}\varphi_m - \mathcal{L}\varphi|^p\, d\gamma \le C_p\, ||\varphi_m - \varphi||^p_{\mathbb{D}_2^p(\gamma)}.$$

Therefore for a subsequence, $\mathcal{L}\varphi_m \to \mathcal{L}\varphi$ almost all. Now

$$\int_{\mathbb{R}^n} |W_m(\nabla\Phi_m) - W(\nabla\Phi)|\, d\gamma \le \int_{\mathbb{R}^n} |W_m(\nabla\Phi_m) - W(\nabla\Phi_m)|\, d\gamma + \int_{\mathbb{R}^n} |W(\nabla\Phi_m) - W(\nabla\Phi)|\, d\gamma. \tag{7.1.4}$$

By condition (H), the first term of the right hand side of (7.1.4) is less than

$$\frac{1}{\delta_1} \int_{\mathbb{R}^n} |W_m(\nabla\Phi_m) - W(\nabla\Phi_m)|\, e^{-V_m} d\gamma = \frac{1}{\delta_1} \int_{\mathbb{R}^n} |W_m - W|\, e^{-W_m} d\gamma \to 0,$$

as $m \to +\infty$. For estimating the second term, let $\varepsilon > 0$, choose $\hat{W} \in C_b(\mathbb{R}^n)$ such that

$$||W - \hat{W}||_{L^1(\gamma)} \le \varepsilon.$$

We have

$$\begin{aligned}\int_{\mathbb{R}^n} |W(\nabla\Phi_m) - W(\nabla\Phi)|\, d\gamma &\le \frac{1}{\delta_1}\int_{\mathbb{R}^n} |W - \hat{W}|(\nabla\Phi_m)\, e^{-V_m} d\gamma \\ &\quad + \int_{\mathbb{R}^n} |\hat{W}(\nabla\Phi_m) - \hat{W}(\nabla\Phi)|\, d\gamma + \frac{1}{\delta_1}\int_{\mathbb{R}^n} |W - \hat{W}|(\nabla\Phi)\, e^{-V} d\gamma \\ &\le \frac{2\delta_2}{\delta_1}||W - \hat{W}||_{L^1(\gamma)} + \int_{\mathbb{R}^n} |\hat{W}(\nabla\Phi_m) - \hat{W}(\nabla\Phi)|\, d\gamma.\end{aligned}$$

It follows that

$$\lim_{m\to+\infty} \int_{\mathbb{R}^n} |W(\nabla\Phi_m) - W(\nabla\Phi)|\, d\gamma = 0.$$

So, combining this with (7.1.4), up to a subsequence, $W_m(\nabla\Phi_m) \to W(\nabla\Phi)$ almost all. The proof of (7.1.2) is complete. □

In what follows, we will drop the condition (H).

Theorem 7.1.2. *Let $V \in \mathbb{D}_1^2(\mathbb{R}^n, \gamma)$ and $W \in \mathbb{D}_2^2(\mathbb{R}^n, \gamma)$ satisfying conditions* (7.1.1)*. Then $\mathcal{L}\varphi$ exists in $L^1(\mathbb{R}^n, e^{-V} d\gamma)$ and*

$$e^{-V} = e^{-W(\nabla\Phi)} e^{\mathcal{L}\varphi - \frac{1}{2}|\nabla\varphi|^2} \mathrm{det}_2(\mathrm{Id} + \nabla^2\varphi),$$

where $\nabla\Phi(x) = x + \nabla\varphi(x)$.

Proof. Consider $V_m = V \wedge m$ for $m \ge 1$; then $V_p \le V_m$ if $p \le m$. Set $a_m = \int_{\mathbb{R}^n} e^{-V_m}\, d\gamma$, which goes to 1 as $m \to +\infty$. Without loss of generality, we assume that $\frac{1}{2} \le a_m \le 2$. Let $x \to x + \varphi_m(x)$ be the optimal map from $\frac{e^{-V_m}}{a_m} d\gamma$ to $e^{-W} d\gamma$. By Theorem 5.3.10,

$$\begin{aligned}&\int_{\mathbb{R}^n} ||\mathrm{Id} + \nabla^2\varphi_m||_{op}^2\, \frac{e^{-V_m}}{a_m}\, d\gamma \\ &\le 2\Big(1 + \frac{2}{1-c}\int_{\mathbb{R}^n} |\nabla V_m|^2 \frac{e^{-V_m}}{a_m} d\gamma + (\frac{2}{1-c})^2 \int_{\mathbb{R}^n} ||\nabla^2 W||_{HS}^2 e^{-W} d\gamma\Big),\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}^n} ||\mathrm{Id}+\nabla^2\varphi_p||_{op}^2\,\frac{e^{-V_m}}{a_m}\,d\gamma &\leq 2\int_{\mathbb{R}^n}\Bigl(1+||\nabla^2\varphi_p||_{HS}^2\Bigr)\,\frac{e^{-V_p}}{a_p}\,e^{V_p-V_m}\,\frac{a_p}{a_m}\,d\gamma\\ &\leq 8\int_{\mathbb{R}^n}\Bigl(1+||\nabla^2\varphi_p||_{HS}^2\Bigr)\,\frac{e^{-V_p}}{a_p}\,d\gamma\\ &\leq 8\Bigl(1+\frac{2}{1-c}\int_{\mathbb{R}^n}|\nabla V_p|^2\frac{e^{-V_p}}{a_p}d\gamma+(\frac{2}{1-c})^2\int_{\mathbb{R}^n}||\nabla^2 W||_{HS}^2 e^{-W}d\gamma\Bigr).\end{aligned}$$

Therefore according to Thorem 5.3.11, it exists a constant $C>0$ independent of m, such that

$$\frac{1}{a_m}\int_{\mathbb{R}^n}||\nabla^2\varphi_m-\nabla^2\varphi_p||_{HS}\,e^{-V}d\gamma \leq C\int_{\mathbb{R}^n}|V_m-V_p|\,\frac{e^{-V_m}}{a_m}d\gamma \leq 2C\delta_2||V_m-V_p||_{L^2(\gamma)}.$$

It follows that $\{\nabla^2\varphi_m;\ m\geq 1\}$ is a Cauchy sequence in $L^1(e^{-V}d\gamma)$. Up to subsequence, $\nabla^2\varphi_m$ converges to $\nabla^2\varphi$ almost all. On the other hand, by Theorem 5.3.1,

$$\int_{\mathbb{R}^n}|\nabla\varphi_m-\nabla\varphi_p|^2\,\frac{e^{-V_m}}{a_m}\,d\gamma \leq \frac{4}{1-c}\int_{\mathbb{R}^n}|V_m-V_p+\log a_m-\log a_p|\,\frac{e^{-V_m}}{a_m}\,d\gamma,$$

which tends to 0 as $p,m\to+\infty$. Therefore up to a subsequence, $\nabla\varphi_m$ converges to $\nabla\varphi$ almost all.

Now using Theorem 7.1.1, we have

$$\frac{e^{-V_m}}{a_m} = e^{-W(\nabla\Phi_m)}e^{\mathcal{L}\varphi_m-\frac{1}{2}|\nabla\varphi_m|^2}\mathrm{det}_2(\mathrm{Id}+\nabla^2\varphi_m), \tag{7.1.5}$$

where $\nabla\Phi_m(x)=x+\nabla\varphi_m(x)$. As what did in the last part of the proof to Theorem 7.1.1, we have

$$\lim_{m\to\infty}\int_{\mathbb{R}^n}|e^{-W(\nabla\Phi_m)}-e^{-W(\nabla\Phi)}|\,e^{-V}d\gamma=0. \tag{7.1.6}$$

Therefore for a subsequence, we proved that each term except $\mathcal{L}\varphi_m$ in (7.1.5) converges almost all; it follows

$$\text{up to a subsequence, } \mathcal{L}\varphi_m \text{ converges to a function } F \text{ almost all.} \tag{7.1.7}$$

The fact that $F\in L^1(\mathbb{R}^n, e^{-V}d\gamma)$ comes from the relation

$$F=-V+W(\nabla\Phi)+\frac{1}{2}|\nabla\varphi|^2-\log\mathrm{det}_2(\mathrm{Id}+\nabla^2\varphi).$$

Now it remains to prove that $\mathcal{L}\varphi$ exists in $L^1(\mathbb{R}^n, e^{-V}d\gamma)$ and $F=\mathcal{L}\varphi$. The difficulty is that we have no more the control in $L^2(e^{-V}d\gamma)$ of $\mathcal{L}\varphi_m$ by $\nabla^2\varphi_m$. We will proceed as in [13].

Lemma 7.1.3. *Assume that* $e^{-V} \geq \delta_1 > 0$. *Then there exists a constant* K *independent of* δ_1 *such that for any* $f \in \mathbb{D}_2^2(\mathbb{R}^n, e^{-V}d\gamma)$,

$$\int_{\mathbb{R}^n} (\mathcal{L}f)^2 e^{-|\nabla f|^2}\, e^{-V} d\gamma \leq K \Big(1 + \int_{\mathbb{R}^n} |\nabla^2 f|^2\, e^{-V} d\gamma + \int_{\mathbb{R}^n} |\nabla V|^2\, e^{-V} d\gamma\Big). \quad (7.1.8)$$

Proof. Any $f \in \mathbb{D}_2^2(\mathbb{R}^n, e^{-V}d\gamma)$ is also in $\mathbb{D}_2^2(\mathbb{R}^n, d\gamma)$; then $\mathcal{L}f$ exists in $L^2(\mathbb{R}^n, e^{-V}d\gamma)$, and we can approximate f by functions in C^2 bounded with bounded derivatives up to order 2. For the moment, assume that f is in the latter class. So

$$\int_{\mathbb{R}^n} (\mathcal{L}f)^2 e^{-|\nabla f|^2}\, e^{-V} d\gamma = -\int_{\mathbb{R}^n} \langle \nabla f,\, \nabla(\mathcal{L}f e^{-|\nabla f|^2} e^{-V})\rangle\, d\gamma. \quad (7.1.9)$$

We have

$$\begin{aligned} \langle \nabla f,\, \nabla(\mathcal{L}f e^{-|\nabla f|^2} e^{-V})\rangle &= \langle \nabla f, \nabla \mathcal{L}f\rangle\, e^{-|\nabla f|^2} e^{-V} \\ &\quad - 2\langle \nabla f \otimes \nabla f, \nabla^2 f\rangle e^{-V} \mathcal{L}f e^{-|\nabla f|^2} - \langle \nabla f, \nabla V\rangle \mathcal{L}f e^{-|\nabla f|^2} e^{-V}. \end{aligned} \quad (7.1.10)$$

By Cauchy-Schwarz inequality,

$$\begin{aligned} &\int_{\mathbb{R}^n} \langle \nabla f \otimes \nabla f, \nabla^2 f\rangle e^{-V} \mathcal{L}f e^{-|\nabla f|^2} d\gamma \\ &\leq \Big(\int_{\mathbb{R}^n} \langle \nabla f \otimes \nabla f, \nabla^2 f\rangle^2 e^{-|\nabla f|^2} e^{-V} d\gamma\Big)^{1/2} \Big(\int_{\mathbb{R}^n} (\mathcal{L}f)^2 e^{-|\nabla f|^2} e^{-V} d\gamma\Big)^{1/2}. \end{aligned}$$

In the same way, we treat the last term in (7.1.10). Set $A = \int_{\mathbb{R}^n} \langle \nabla f,\, \nabla \mathcal{L}f\rangle e^{-|\nabla f|^2} e^{-V} d\gamma$,

$$B = 2\Big(\int_{\mathbb{R}^n} \langle \nabla f \otimes \nabla f, \nabla^2 f\rangle^2 e^{-|\nabla f|^2} e^{-V} d\gamma\Big)^{1/2} + \Big(\int_{\mathbb{R}^n} \langle \nabla f, \nabla V\rangle^2 e^{-|\nabla f|^2} e^{-V} d\gamma\Big)^{1/2},$$

and $Y = \Big(\int_{\mathbb{R}^n} (\mathcal{L}f)^2 e^{-|\nabla f|^2} e^{-V} d\gamma\Big)^{1/2}$. Then combining (7.1.9), (7.1.10) and par above computation, we get

$$Y^2 \leq -A + BY. \quad (7.1.11)$$

It follows that the discriminant of $P(\lambda) = \lambda^2 - B\lambda + A$ is non negative and $P(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)$. The relation (7.1.11) implies that Y is between two roots of P. In particular,

$$Y \leq (B + \sqrt{B^2 - 4A})/2. \quad (7.1.12)$$

It is obvious that for a numerical constant $K_1 > 0$,

$$B^2 \leq K_1 \Big(\int_{\mathbb{R}^n} |\nabla^2 f|^2\, e^{-V} d\gamma + \int_{\mathbb{R}^n} |\nabla V|^2\, e^{-V} d\gamma\Big).$$

For estimating the term A, we use the commutation formula for Gaussian measures (Proposition 2.1.5),

$$\nabla\mathcal{L}f = \mathcal{L}\nabla f - \nabla f,$$

so that we get

$$|A| \leq K_1\Big(1 + \int_{\mathbb{R}^n} |\nabla^2 f|^2\, e^{-V} d\gamma + \int_{\mathbb{R}^n} |\nabla V|^2\, e^{-V} d\gamma\Big).$$

Now the relation (7.1.12) yields (7.1.8). □

Applying (7.1.8) to φ_m, we have

$$\sup_{m\geq 1} \int_{\mathbb{R}^n} (\mathcal{L}\varphi_m)^2 e^{-|\nabla\varphi_m|^2}\, e^{-V} d\gamma < +\infty.$$

Therefore the family $\{\mathcal{L}\varphi_m\, e^{-|\nabla\varphi_m|^2/2}\}$ is uniformly integrable with respect to $e^{-V}d\gamma$. Then for any $\xi \in C_b^1(\mathbb{R}^n)$,

$$\lim_{m\to+\infty} \int_{\mathbb{R}^n} \mathcal{L}\varphi_m e^{-|\nabla\varphi_m|^2/2} \xi\, e^{-V} d\gamma = \int_{\mathbb{R}^n} F e^{-|\nabla\varphi|^2/2} \xi\, e^{-V} d\gamma. \qquad (7.1.13)$$

But

$$\int_{\mathbb{R}^n} \mathcal{L}\varphi_m e^{-|\nabla\varphi_m|^2/2} \xi\, e^{-V} d\gamma = \int_{\mathbb{R}^n} \langle \nabla\varphi_m \otimes \nabla\varphi_m,\, \nabla^2\varphi_m\rangle e^{-|\nabla\varphi_m|^2/2}\, \xi e^{-V} d\gamma - \int_{\mathbb{R}^n} \langle \varphi_m, \nabla(\xi e^{-V})\rangle e^{-|\nabla\varphi_m|^2/2} d\gamma,$$

which converges to $\int_{\mathbb{R}^n} \langle \nabla\varphi \otimes \nabla\varphi,\, \nabla^2\varphi\rangle e^{-|\nabla\varphi|^2/2}\, \xi e^{-V} d\gamma - \int_{\mathbb{R}^n} \langle \varphi, \nabla(\xi e^{-V})\rangle e^{-|\nabla\varphi|^2/2} d\gamma$. So we get

$$\int_{\mathbb{R}^n} (F - \langle\nabla\varphi, \nabla V\rangle) e^{-|\nabla\varphi|^2/2} \xi\, e^{-V} d\gamma = -\int_{\mathbb{R}^n} \langle \nabla\varphi,\, \nabla(\xi e^{-|\nabla\varphi|^2/2})\rangle\, e^{-V} d\gamma. \qquad (7.1.14)$$

Note that the generator $\mathcal{L}_V$ associated to the Dirichlet form $\mathcal{E}_V(f,f) = \int_{\mathbb{R}^n} |\nabla f|^2\, e^{-V} d\gamma$ admits the expression $\mathcal{L}_V(f) = \mathcal{L}(f) - \langle \nabla f, \nabla V\rangle$. Therefore the relation (7.1.14) tells us that $F = \mathcal{L}\varphi$. □

7.2 Monge-Ampère equations on the Wiener space

We return now to the situation in Theorem 6.2.1. Let $V \in \mathbb{D}_1^2(X)$ and $W \in \mathbb{D}_2^2(X)$ such that $\int_X e^{-V} d\mu = \int_X e^{-W} d\mu = 1$. Assume that

$$e^{-V} \leq \delta_2, \quad e^{-W} \leq \delta_2, \quad \nabla^2 W \geq -cId, \quad c \in [0,1). \qquad (7.2.1)$$

Let $\{e_n; n \geq 1\} \subset X^*$ be an orthonormal basis of H and H_n the subspace spanned by $\{e_1, \ldots, e_n\}$. As in section 1, denote $\pi_n(x) = \sum_{j=1}^{n} e_j(x)e_j$ and $\mathcal{F}_n$ the sub σ-field generated by π_n. In the sequel, we will see that the manner to regularize the density functions e^{-V} and e^{-W} has impacts on final results.

Set

$$\mathbb{E}(e^{-V}|\mathcal{F}_n) = e^{-V_n} \circ \pi_n,\ \mathbb{E}(W|\mathcal{F}_n) = W_n \circ \pi_n. \tag{7.2.2}$$

It is obvious that $\nabla^2 W_n \geq -c\,\mathrm{Id}_{H_n \otimes H_n}$. Applying Theorem 5.3.10, there is a $\varphi_n \in \mathbb{D}_2^2(H_n, \gamma_n)$ such that $x \to x + \nabla\varphi_n(x)$ is the optimal transport map which pushes $e^{-V_n}\gamma_n$ to $e^{-W_n}\gamma_n$. Let $\tilde{\varphi}_n = \varphi_n \circ \pi_n$. We have

$$\begin{aligned} &\frac{1-c}{2}\int_{H_n} ||\nabla^2\varphi_n||_{HS}^2 e^{-V_n} d\gamma_n \\ &\leq \int_{H_n} |\nabla V_n|^2 e^{-V_n} d\gamma_n + \frac{2}{1-c}\int_{H_n} ||\nabla^2 W_n||_{HS}^2 e^{-W_n} d\gamma_n. \end{aligned} \tag{7.2.3}$$

By Cauchy-Schwarz inequality for conditional expectation,

$$|\nabla\mathbb{E}(e^{-V}|\mathcal{F}_n)|_{H_n}^2 \leq \mathbb{E}(|\nabla V|_H^2 e^{-V}|\mathcal{F}_n)\,\mathbb{E}(e^{-V}|\mathcal{F}_n)$$

which implies that $\int_{H_n} |\nabla V_n|^2 e^{-V_n} d\gamma_n \leq \int_X |\nabla V|^2 e^{-V} d\mu$. So (7.2.3) yields

$$\frac{1-c}{2}\int_X ||\nabla^2\tilde{\varphi}_n||_{HS}^2 e^{-V} d\mu \leq \int_X |\nabla V|^2 e^{-V} d\mu + \frac{2\delta_2}{1-c}\int_X ||\nabla^2 W||_{HS}^2 d\mu. \tag{7.2.4}$$

Let n, m be two integers such that $n > m$, and $\pi_m^n : H_n \to H_m$ the orthogonal projection. Then $I_{H_n} + \nabla(\varphi_m \circ \pi_m^n)$ pushes $e^{-V_m} \circ \pi_m^n \gamma_n$ to $e^{-W_m} \circ \pi_m^n\, \gamma_n$. In fact, for any bounded continuous function $f : H_n \to \mathbb{R}$,

$$\begin{aligned} &\int_{H_n} f\big(x + \pi_m^n(\nabla\varphi_m) \circ \pi_m^n(x)\big) e^{-V_m} \circ \pi_m^n d\gamma_n \\ &= \int_{H_m^\perp} \Big[\int_{H_m} f(z' + z + \pi_m^n(\nabla\varphi_m)(z)) e^{-V_m}(z) d\gamma_m(z)\Big] d\hat{\gamma}(z'), \end{aligned}$$

where $H_n = H_m \oplus H_m^\perp$ and $\gamma_n = \gamma_m \otimes \hat{\gamma}$. Note that $\pi_m^n(\nabla\varphi_m) = \nabla\varphi_m$; then the last term in above equality yields

$$\int_{H_m^\perp} \Big[\int_{H_m} f(z' + y) e^{-W_m}(y) d\gamma_m(y)\Big] d\hat{\gamma}(z') = \int_{H_n} f(x) e^{-W_m} \circ \pi_m^n(x) d\gamma_n(x).$$

Now by (5.3.16),

$$||\nabla\varphi_n - \nabla(\varphi_m \circ \pi^n_m)||^2_{L^2(e^{-V_n}\gamma_n)}$$
$$\leq \frac{4}{1-c}\int (V_n - V_m \circ \pi^n_m)e^{-V_n} d\gamma_n + \frac{4}{(1-c)^2}\int_{H_n} |\nabla W_n - \nabla(W_m \circ \pi^n_m)|^2\, e^{-W_n}\, d\gamma_n,$$

or

$$||\nabla\tilde{\varphi}_n - \nabla\tilde{\varphi}_m||^2_{L^2(e^{-V}\mu)}$$
$$\leq \frac{4}{1-c}\int_X (V_n \circ \pi_n - V_m \circ \pi_m)e^{-V} d\mu + \frac{4\delta_2}{(1-c)^2}\int_X |\nabla\mathbb{E}(W|\mathcal{F}_n) - \nabla\mathbb{E}(W|\mathcal{F}_m)|^2\, d\mu. \tag{7.2.5}$$

Now in order to control the sequence of functions $\tilde{\varphi}_n$, we suppose that

$$e^{-V} \geq \delta_1 > 0. \tag{7.2.6}$$

Under (7.2.6), it is clear that

$$\int_X (V_n \circ \pi_n - V_m \circ \pi_m)e^{-V} d\mu \to 0, \text{ as } n, m \to +\infty.$$

Now replacing $\tilde{\varphi}_n$ by $\tilde{\varphi}_n - \int_X \tilde{\varphi}_n\, d\mu$ and according to Poincaré inequality, and by (7.2.5), we see that $\tilde{\varphi}_n$ converges in $\mathbb{D}^2_1(X)$ to a function φ. On the other hand, by (7.2.4), $\tilde{\varphi}_n$ converges to a function $\hat{\varphi} \in \mathbb{D}^2_2(X)$ weakly. By uniqueness of limits, we see in fact that $\varphi \in \mathbb{D}^2_2(X)$. Now we proceed as in Section 7.1, we have

$$\lim_{n\to+\infty}\int_X ||\nabla^2\tilde{\varphi}_n - \nabla^2\varphi||_{HS}\, d\mu = 0. \tag{7.2.7}$$

Combining (7.2.7) and (7.2.4), up to a subsequence, for any $1 < p < 2$,

$$\lim_{n\to+\infty}\int_X ||\nabla^2\tilde{\varphi}_n - \nabla^2\varphi||^p_{HS}\, d\mu = 0. \tag{7.2.8}$$

By Meyer inequality ([48]),

$$\lim_{n\to+\infty}\int_X ||\mathcal{L}\tilde{\varphi}_n - \mathcal{L}\varphi||^p_{HS}\, d\mu = 0. \tag{7.2.9}$$

So everything goes well under the supplementary condition (7.2.6). We finally get

Theorem 7.2.1. *Under conditions* (7.2.1) *and* (7.2.6)*, there exists a function* $\varphi \in \mathbb{D}^2_2(X)$ *such that* $x \to x + \nabla\varphi(x)$ *pushes* $e^{-V}\mu$ *to* $e^{-W}\mu$ *and solves the Monge-Ampère equation*

$$e^{-V} = e^{-W(T)}e^{\mathcal{L}\varphi - \frac{1}{2}|\nabla\varphi|^2}\, det_2(\mathrm{Id}_{H\otimes H} + \nabla^2\varphi),$$

where $T(x) = x + \nabla\varphi(x)$.

Remark: The regularization of W used in (7.2.2) does not allows to prove that

$$W_2^2(e^{-V_n}\gamma_n, e^{-W_n}\gamma_n)$$

converges to $W_2^2(e^{-V}\mu, e^{-W}\mu)$ contrary to section 1; we do not know if the map T constructed in Theorem 7.2.1 is the optimal transport : which is due to the singularity of the cost function d_H in contrast to finite dimensional case (see subsection 3.1).

Theorem 7.2.2. *Assume all conditions in Theorem 7.2.1 and that W_n defined by*

$$\mathbb{E}(e^{-W}|\mathcal{F}_n) = e^{-W_n} \circ \pi_n,$$

belongs to $\mathbb{D}_2^2(H_n)$ for all $n \geq 1$. Then there is a function $\varphi \in \mathbb{D}_2^2(X)$ such that $x \to T(x) = x + \nabla\varphi(x)$ is the optimal transport map which pushes $e^{-V}\mu$ to $e^{-W}\mu$ and T is the inverse map of S in Theorem 6.2.1.

Proof. By Proposition 5.1 in [35], W_n satisfies the condition (7.1.1). So we can repeat the arguments as above, but the difference is that in actual case, $W_2^2(e^{-V_n}\gamma_n, e^{-W_n}\gamma_n)$ converges to $W_2^2(e^{-V}\mu, e^{-W}\mu)$. Using notations in the proof of Theorem 6.2.1, $x \to x - \frac{1}{2}\nabla\varphi_n(x)$ is the optimal transport map, which pushes $e^{-V_n}\gamma_n$ to $e^{-W_n}\gamma_n$. So that

$$W_2^2(e^{-V}\mu, e^{-W}\mu) = \frac{1}{4}\int_X |\nabla\varphi|_H^2\, e^{-V} d\mu,$$

that means that $x \to T(x) = x - \frac{1}{2}\nabla\varphi(x)$ is the optimal transport map which pushes $e^{-V}\mu$ to $e^{-W}\mu$. To see that T is the inverse map of S in Theorem 6.2.1, we use (6.2.14), which implies that under the optimal plan Γ_0,

$$-2\psi(x) + \varphi(y) = d_H(x,y)^2,$$

since we have replaced $-\frac{1}{2}\psi$ by ψ at the end of the proof of Theorem 6.2.1. Again, because $\varphi \in \mathbb{D}_2^2(X)$, we can differentiate φ, so that under Γ_0,

$$x = y - \frac{1}{2}\nabla\varphi(y).$$

Therefore $\eta \in L^2(X, H, e^{-V}\mu)$ is given by $\eta = -\frac{1}{2}\nabla\varphi$ with $\varphi \in \mathbb{D}_2^2(X)$. □

Examples: (i) If $W \in \mathbb{D}_2^2(X)$ satisfies $\int_X |\nabla W|^4\, d\mu < +\infty$ and $0 < \delta_1 \leq e^{-W} \leq \delta_2$ then condition in Theorem 7.2.2 holds. □

(ii) For an orthonormal basis $\{e_n; n \geq 1\}$ of H, define $W(x) = \sum_{n\geq 1} \lambda_n e_n(x)^2$, where $\lambda_n > -1/2$ and $\sum_{n\geq 1} |\lambda_n| < +\infty$. We have,

$$\mathbb{E}(e^{-W}|\mathcal{F}_n) = e^{-\sum_{k=1}^n \lambda_k e_k(x)^2} \prod_{k>n} \mathbb{E}(e^{-\lambda_k e_k(x)^2}) = \alpha_n e^{-\sum_{k=1}^n \lambda_k e_k(x)^2},$$

where $\alpha_n = \prod_{k>n} \frac{1}{\sqrt{1+2\lambda_k}}$. So condition in Theorem 7.2.2 holds. □

Notations:

- (X, d) Polish space
- $\mathcal{P}(X)$ the set of Borel probability measures on X
- $\mathcal{P}_p(X)$ the subset of $\mathcal{P}(X)$ of measures with finite $p-$th moment order
- (X, H, μ) an abstract Wiener space, with Wiener measure μ
- $d_H(w, w')$ the pseudo-distance between w and $w' \in X$, induced by the norm $|.|_H$
- $T_\# \rho_0 := \rho_0 \circ T^{-1}$ the push-forward measure
- $C(\rho_0, \rho_1)$ the set of couplings between two probability measures ρ_0 and ρ_1
- $C_0(\rho_0, \rho_1)$ the set of optimal couplings (relatively to a cost)
- Π_0 optimal coupling between two probability measures (w.r.t. a given cost)
- $\mathbb{D}_2^p(X)$ Sobolev space over X
- $W_{p,c}(\rho_0, \rho_1)$ the $p-$Wasserstein distance between ρ_0 and ρ_1 w.r.t. c
- $Ent_\mu(\rho)$ relative entropy of ρ with respect to μ
- $\pi_n : X \longrightarrow V_n$ orthogonal projections onto $n-$dimensional space
- $P_i : X \times X \longrightarrow X$, the projection onto the $i - th$ component $(i = 1, 2)$
- $T_t : X \times X \longrightarrow X$, $T_t(x, y) := (1 - t)x + ty$ for $t \in [0, 1]$
- $(\rho_t)_{0 \leq t \leq 1}$ McCann's interpolation between ρ_0 and ρ_1
- γ_n the standard Gaussian measure on $\mathbb{R}^n$
- $|.|_q$ the $q-$norm in $\mathbb{R}^n$
- $\nabla\Phi(x) = x + \nabla\varphi(x)$ the Brenier's map

Bibliography

[1] S. Aida and T. Zhang. On the Small Time Asymptotics of Diffusion Processes on Path Groups. *Potential Analysis*, 16:67–78, 2002.

[2] H. Airault and P. Malliavin. Integration geometrique sur l'espace de Wiener. *Bulletin des Sciences Mathematiques*, 112:3–52, 1988.

[3] L. Ambrosio. Optimal transport maps in Monge-Kantorovich problem. *Proceedings of the International Congress of Mathematicians, Vol. III*, pages 131–140, 2002.

[4] L. Ambrosio. Lecture notes on optimal transport problems. *Mathematical Aspects of Evolving Interfaces*, pages 1–52, 2003.

[5] L. Ambrosio and N. Gigli. A user's guide to optimal transport. 2011.

[6] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics, 2008.

[7] L. Ambrosio, B. Kirchheim, and A. Pratelli. Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal*, 125:207–241, 2004.

[8] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. *Lecture Notes in Mathematics*, 1813:123–160, 2003.

[9] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *Journal of the European Mathematical Society*, 9:85–121, 2007.

[10] S. Bianchini and F. Cavalletti. The monge problem for distance cost in geodesic spaces. *Submitted Paper*, 2009.

[11] S. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Jounal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.

[12] V.I. Bogachev. *Gaussian measures.* 1998.

[13] V.I. Bogachev and A.V. Kolesnikov. Sobolev regularity for the Monge-Ampere equation in the Wiener space. *arXiv:1110.1822*.

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

[15] L. Caffarelli. The regularity of mappings with a convex potential. *American Mathematical Society*, 5(1), 1992.

[16] L. Caffarelli, M. Feldman, and R.J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *Journal of the American Mathematical Society*, 15:1–26, 2002.

[17] L. Caravenna. A proof of Monge problem in $\mathbb{R}^n$ by stability. *Rend. Istit. Mat. Univ. Trieste*, 43:31–52, 2011.

[18] L. Caravenna. A proof of Sudakov theorem with strictly convex norms. *Math. Zeitschift*, 268:371–407, 2011.

[19] F. Cavalletti. The Monge Problem in Wiener space. *Calculus of Variations*, 45:101–124, 2011.

[20] T. Champion and L. De Pascale. The Monge problem for strictly convex norms in $\mathbb{R}^d$. *J. Eur. Math. Soc.*, 12:1355–1369, 2010.

[21] T. Champion and L. De Pascale. The Monge problem in $\mathbb{R}^d$. *Duke Mathematical Journal*, 157(3):551–572, 2010.

[22] T. Champion and L. De Pascale. On the twist condition and $c-$monotone transport plans. *submitted*, 2012.

[23] D. Cordero-Erausquin. Sur le transport de mesures périodiques. *C.R. Acadmie des Sciences*, 329:199–202, 1999.

[24] Bakry D. and M. Emery. Diffusion hypercontractivities. *Sém. de Probab. XIX, Lect. Notes in Math.*, 1123:77–206, 1985.

[25] G. De Philippis and A. Figalli. Sobolev regularity for Monge-Ampere type equations. *arXiv:1211.2341*.

[26] G. De Philippis and A. Figalli. $W^{2,1}$ regularity for solutions of the Monge-Ampere equation. *arXiv:1111.7207*.

[27] O. Enchev and W. Stroock. Rademacher's theorem for wiener functionals. *The Annals of Probability*, 21(1):25–33, 1993.

[28] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Memoirs of the American Mathematical Society*, 137:653, 1999.

[29] S. Fang. *Introduction to Malliavin Calculus.* Mathematics Series for Graduate Students, 2003.

[30] S. Fang and V. Nolot. Gaussian estimates on sobolev spaces. *arXiv:1207.4907*.

[31] S. Fang and J. Shao. Optimal transport maps for Monge-Kantorovich problem on loop groups. *Journal of Functional Analysis*, 248:225–257, 2007.

[32] S. Fang, J. Shao, and K-T. Sturm. Wasserstein space over the Wiener space. *Probab. Theory Related Fields*, 146(3):535–565, 2010.

[33] A. Fathi and F. Figalli. Optimal transportation on non-compact manifolds. *Israel J. Math.*, 175:1–59, 2010.

[34] M. Feldman and R.J. McCann. Monge's transport problem on a Riemannian manifold. *Transactions of the American Mathematical Society*, 354:1667–1697, 2002.

[35] D. Feyel and A.S. Ustünel. The notion of convexity and concavity on wiener space. *Journal of Functional Analysis*, 176:400–428, 2000.

[36] D. Feyel and A.S. Ustünel. Monge-Kantorovitch measure transportation and Monge-Ampere equation on Wiener space. *Probab. Theory Related Fields*, 128:347–385, 2004.

[37] D. Feyel and A.S. Ustünel. Solution of the Monge-Ampère equation on Wiener space for general log-concave measures. pages 29–55, 2006.

[38] F. Figalli. The monge problem on non-compact manifolds. *The Mathematical Journal of the University of Padova*, 117:147–166, 2007.

[39] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[40] W. Gangbo and V. Oliker. Existence of optimal maps in the reflector-type problems. *ESAIM Control Optim. Calc. Var.*, 13:93–106, 2007.

[41] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $P_2(M)$. *Meth. Appl. of Anal.*, 2011.

[42] N. Gigli. Optimal maps in non branching spaces with Ricci curvature bounded from below. *Geometric and Functional analysis*, 22(4):990–999, 2012.

[43] L. Gross. Abstract Wiener spaces. *Berkeley Symp. Math. Stat. Probab.*, 2:31–41, 1965.

[44] M. KassMann. Harnack inequalities: an introduction. *Boundary Value Problems*, 2007.

[45] A.V. Kolesniko. On Sobolev regularity of mass transport and transportation inequalities. *arXiv:1007.1103*.

[46] A.V. Kolesnikov. Convexity inequalities and optimal transport of infinite-dimensional measures. *Mathematiques pures et appliquees*, 83(11):1373–1404, 2004.

[47] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, 169:903–991, 2009.

[48] P. Malliavin. *Intégration et analyse de Fourier. Probabilités et analyse gaussienne*. Maitrise de mathématiques pures, 1997.

[49] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–323, 1995.

[50] R.J. McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128:153–179, 1997.

[51] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

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

[53] D. Preiss. Gaussian measures and the density theorem. *Commentationes Mathematicae Universitatis Carolinae*, 22(1):181–193, 1981.

[54] J. Shao. Harnack and HWI inequalities on infinite-dimensional spaces. *Acta Mathematica Sinica-english*, 27(6):1195–1204, 2011.

[55] K.T. Sturm. On the Geometry of Metric Measure Spaces I. *Acta Math.*, 196:65–131, 2006.

[56] J. Tiser. Differentiation theorem for Gaussian measures on Hilbert space. *Transactions of the American Mathematical Society*, 308(2):655–666, 1988.

[57] N. S. Trudinger and X. J. Wang. On the Monge mass transfer problem. *Calculus of Variations and Partial Differential Equations*, 13:19–31, 2001.

[58] C. Villani. *Optimal transport, old and new*. Grundlehren der mathematischen Wissenschaften, 2009.

[59] C. Villani. Regularity of optimal transport and cut-locus: from non smooth analysis to geometry to smooth analysis. *Discrete and continuous dynamical systems*, 30:559–571, 2011.

[60] M-K. von Renesse and K-T. Sturm. Transport inequalities, gradient estimates, entropy and ricci curvature. *Pure and Applied Mathematics*, 58(7):923–940, 2005.

[61] F.-Y. Wang. *Functional inequalities, Markov semigroups and spectral theory*. 2005.