



HAL
open science

Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles

Luigia Ripani

► **To cite this version:**

Luigia Ripani. Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles. Algèbres quantiques [math.QA]. Université de Lyon, 2017. Français. NNT : 2017LYSE1274 . tel-01718846

HAL Id: tel-01718846

<https://theses.hal.science/tel-01718846>

Submitted on 27 Feb 2018

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

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

Université Claude Bernard Lyon 1
École doctorale **InfoMath**, ED 512
Spécialité : **Mathématiques**
N. d'ordre 2017LYSE1274

Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles

Thèse de Doctorat

présentée publiquement le 6 décembre 2017 par

Luigia RIPANI

après examen par:

M. Guillaume Carlier Université Paris Dauphine
M. Max von-Renesse Université de Leipzig

devant le jury composé par:

M. Guillaume Carlier	Université Paris Dauphine	Rapporteur
M. Ivan Gentil	Université Claude Bernard Lyon 1	Directeur de thèse
M. Christian Léonard	Université Paris Nanterre	Co-directeur de thèse
Mme Marjolaine Puel	Université de Nice Sophia-Antipolis	Examinatrice
M. Max von-Renesse	Université de Leipzig	Rapporteur
Mme Sylvie Roelly	Université de Potsdam	Examinatrice

Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles



Luigia RIPANI

Thèse de doctorat

Remerciements

Pendant ces dernières années j'ai eu la chance d'apprendre beaucoup et de beaucoup recevoir. C'est donc le moment d'exprimer ma gratitude.

Tout d'abord je souhaite remercier les membres de mon jury. Merci à Guillaume Carlier et Max von-Rennesse d'avoir accepté d'être les rapporteurs de ce manuscrit. Je vous remercie de m'avoir fait part de vos commentaires détaillés et encourageants. Merci aussi à Marjolaine Puel et Sylvie Roelly pour votre disponibilité et l'intérêt que vous avez porté à mes travaux.

Je voudrais ensuite remercier plus particulièrement mes directeurs de recherche, sans qui cette thèse n'aurait jamais eu lieu.

Christian, merci d'avoir pris le temps de me transmettre votre connaissance et votre passion pour le problème de Schrödinger. Je vous remercie également pour vos digressions autour des probas, mais pas seulement, et pour la patience avec laquelle vous avez toujours répondu à mes questions. Votre approche pointilleuse des mathématiques et votre enthousiasme au tableau ont été et resteront pour moi un modèle.

Ivan, sans vous ce beau chapitre lyonnais de ma vie n'aurait pas eu lieu. Echanger avec vous m'a énormément apporté autant sur le plan scientifique que sur le plan humain. Je vous remercie de votre confiance et de votre engagement dans mon projet qui m'ont inspirée et motivée pendant ces années. Vous avez été disponible au quotidien, toujours prêt à écouter mes questions et mes doutes, à me remonter le moral, à me rappeler de profiter de mes week-ends, mais aussi, parfois, de payer mes impôts. Avec vous, je ne me suis jamais sentie bête, au contraire, en sortant de votre bureau je me suis toujours sentie rassurée par votre attitude humble et décontractée.

Je vous remercie aussi de m'avoir introduite dans la famille de l'ANR Stab qui m'a fait découvrir une manière passionnée et amusante de faire de la recherche.

Je profite de l'occasion pour remercier Nicoletta Cancrini qui a été le pont entre ma vie universitaire à l'Aquila et l'aventure doctorale à Lyon. Merci de m'avoir introduite à la Théorie de Barky-Emery et de m'avoir encouragée sur ce chemin. Merci également à l'incroyable Chiara Simeoni, professeure de mathématiques autant que de vie.

Cette thèse a été possible grâce au financement du Labex Milyon pour lequel je suis très reconnaissante. Elle n'aurait pas été la même sans les excellentes conditions de travail du laboratoire ICJ. Je remercie tout spécialement Aurélie, Carine, Céline, Vincent et Laurent qui m'ont gâtée de leur disponibilité, leur professionnalisme et leur bonne humeur.

Je remercie aussi tous les membres de l'ICJ, notamment l'équipe d'analyse et tous mes collègues doctorants et doctorantes. Vous avez rendu ces années au laboratoire très intéressantes et agréables. Un mot en particulier pour Maxime H. et sa Matlab-expertise, Maxime P. et sa zen attitude, et Antoine pour les moments conviviaux au 101bis.

Dans les couloirs de ce laboratoire j'ai aussi rencontré ceux qui sont aujourd'hui parmi mes amis les plus proches.

À Niccolò, merci de m'avoir passé le relais en partageant ton expérience d' "immigré" et en étant solidaire de mes premiers pas d'italienne à Lyon.

Je remercie Quentin avec qui j'ai eu la chance de partager ce bureau pendant deux ans. Quand je ne parlais pas encore français, tu étais déjà là: merci pour tous les mots que tu m'as appris, les mails que tu as relus, les commandes latex que tu m'as enseignées, et les innombrables pauses café et bières ici et là qu'on a partagées! Désormais dans le monde des "adultes" avec Pauline, je vous souhaite à tous les deux simplement le meilleur.

Enfin, Michele, ou plutôt Carlo. Avec toi chaque repas dans la cantine se transforme en dîner au restaurant, avec les confidences, les blagues, les plats partagés ensemble. Tu as apporté avec toi un peu de "chez moi", avec tes "passa a prendere un caffè a casa", nos références télé années '90, et notre côté terrone.

Un mot chaleureux aussi pour tous mes autres amis d'ici ou d'ailleurs, qui m'entourent de bonheur et de soutien. Coline et Rob, mes despa-amis Lisa et Davide, Giovanni, Pierre et Vanessa, Biagio, Anna Claudia, Piotr, Sebastian, Benoit. Sans oublier Federica et Claudia et leur présence constante par les longs face-time, les whatsapp audios et les retrouvailles périodiques. Vous avez rendu beaucoup moins pénible le manque de soleil, de mer et de bonne pizza.

Je tiens à remercier aussi Luigi et Roberta avec qui on a pratiquement préparé tous les examens depuis la licence. Sans vous, je n'aurais pas eu autant de plaisir à étudier, comprendre, aimer et détester les maths.

Un grand merci à ma famille qui m'a toujours soutenu et m'a assuré un environnement serein tout au long de mes études malgré toutes les péripéties rencontrées.

Je remercie en particulier la mamma, Mia, il nonno Renato, Alice, Walter et la petite Bibi de m'accueillir comme "la fille prodigue" à chaque fois que je rentre en Italie.

Je voudrais enfin remercier Guillaume. Notre histoire a commencé en même temps que la thèse et depuis son début, au quotidien, tu es mon premier conseiller et supporteur. Cette expérience de thèse, vécue aussi ensemble, a démontré que nos différences font notre force. Malgré le fait qu'on vienne de mondes opposés, tu as su partager ma passion et t'intéresser aux mathématiques. Merci pour tous les podcasts de vulgarisation scientifique téléchargés, les BDs et les livres lus pour te rapprocher encore plus de mon univers.

Your *cocole* is the new *purée de patates*.

Contents

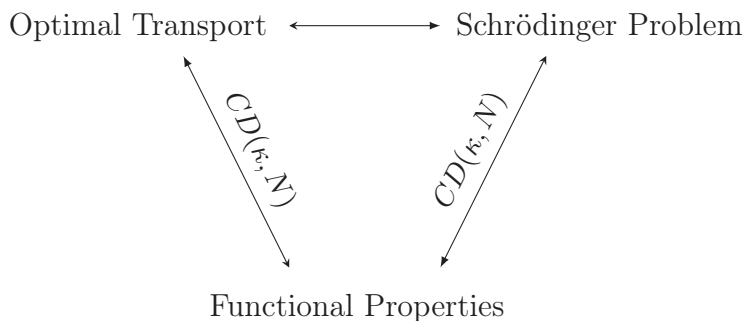
Introduction en français	7
Introduction	11
1 Preliminary tools	27
1.1 Actions of diffusion processes	27
1.1.1 Relative entropy	27
1.1.2 Kinematics of diffusion processes	30
1.1.3 Entropy of diffusion processes	37
1.1.4 Markov semigroups and Γ_2 -calculus	40
1.2 Optimal transport	46
1.2.1 The Monge-Kantorovich problem	47
1.2.2 McCann interpolations	49
1.2.3 Kantorovich and Benamou-Brenier formulations	53
2 The Schrödinger problem	57
2.1 Solutions of the Schrödinger problem	57
2.2 Kinematics of entropic bridges	62
2.3 Entropic cost	70
2.4 Limit for $\varepsilon \rightarrow 0$	73
2.5 Entropy along entropic interpolation	79
2.5.1 Proofs of the preceding results.	81
2.6 Physical motivation and chronology	86
2.6.1 Brief history	89
3 Analogies between OT and SP	91
3.1 Introduction	91
3.2 The reference path measure	95
3.3 Entropic cost and entropic interpolations	97
3.4 Kantorovich dual equality for the entropic cost	100
3.5 Benamou-Brenier formulation of the entropic cost	103
3.6 Contraction with respect to the entropic cost	106
3.7 Examples	110
3.7.1 Wasserstein cost	111
3.7.2 Schrödinger cost	111
3.7.3 Numerical examples	116

4	Properties under $CD(0, n)$	119
4.1	Introduction	119
4.2	Setting	122
4.3	Main results	127
5	EVI under $CD(\kappa, \infty)$	139
5.1	Setting	139
5.2	Main result	141
5.3	Proofs	141
	5.3.1 First method	141
	5.3.2 Second method	145
6	Functional inequalities when $\varepsilon = 1$	153
6.1	Concentration inequality	153
6.2	Prékopa-Leindler inequality	158
	Conclusions and perspectives	165
	Bibliography	166

Introduction en français

Dans cette thèse, nous étudions l'interaction entre les champs classiques du transport optimal et des semigroupes de Markov avec le problème moins connu de minimisation entropique de Schrödinger. Comme resultat des analogies et relations entre ces theories, nous allons deriver des inegalités fonctionnelles.

Cela peut être résumé dans le schema suivant.



Au cours des 20 dernières années, la théorie du transport optimal s'est révélée être un outil efficace pour étudier le comportement asymptotique dans le cas des équations de diffusion, pour prouver des inégalités fonctionnelles et pour étendre des propriétés géométriques dans des espaces extrêmement généraux comme des espaces métriques mesurés, etc. La condition de courbure-dimension de la théorie Bakry-Emery apparaît comme la pierre angulaire de ces applications. Il suffit de penser au cas le plus simple et le plus important de la distance quadratique de Wasserstein W_2 : la contraction du flux de chaleur en W_2 caractérise les bornes inférieures uniformes pour la courbure de Ricci; l'inégalité de Talagrand du transport, comparant W_2 à l'entropie relative est impliquée et implique, par l'inégalité HWI, l'inégalité log-Sobolev; les géodésiques de McCann dans l'espace de Wasserstein $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ permettent de prouver des propriétés fonctionnelles importantes comme la convexité, et des inégalités fonctionnelles standards telles que l'isopérymétrie, des propriétés de concentration de mesure, l'inégalité de Prékopa-Leindler et ainsi de suite. Néanmoins, le manque de régularité des plans minimisation nécessite des arguments d'analyse non lisse.

Le problème de Schrödinger est un problème de minimisation de l'entropie avec des contraintes marginales et un processus de référence fixes. À partir de la théorie des grandes déviations, lorsque le processus de référence est le mouvement Brownien, sa valeur minimale \mathcal{A} converge vers W_2 lorsque la température est nulle. Les interpolations entropiques, solutions du problème de Schrödinger, sont caractérisées en termes de semigroupes de Markov, ce qui implique naturellement les calculs Γ_2 et la condition de courbure-dimension. Datant des années 1930 et négligé pendant des décennies, le problème de Schrödinger connaît depuis ces dernières années une popularité croissante dans différents domaines, grâce à sa relation avec le transport optimal, à la régularité de ses solutions, et à d'autres propriétés performantes dans des calculs numériques.

Le but de ce travail est double. D’abord, nous étudions certaines analogies entre le problème de Schrödinger et le transport optimal fournissant de nouvelles preuves de la formulation duale de Kantorovich et de celle, dynamique, de Benamou-Brenier pour le coût entropique \mathcal{A} . Puis, en tant qu’application de ces connexions, nous dérivons certaines propriétés et inégalités fonctionnelles sous des conditions de courbure-dimension. En particulier, nous prouvons la concavité de l’entropie exponentielle le long des interpolations entropiques sous la condition de courbure-dimension $CD(0, n)$ et la régularité du coût entropique le long du flot de la chaleur. Nous donnons également différentes preuves de l’inégalité variationnelle évolutionnaire pour \mathcal{A} et de la contraction du flux de la chaleur en \mathcal{A} , en retrouvant comme cas limite, les résultats classiques en W_2 , sous $CD(\kappa, \infty)$ et $CD(0, n)$. Enfin, nous proposons une preuve simple de la propriété de concentration gaussienne via le problème de Schrödinger comme alternative aux arguments classiques tel que l’argument de Marton basé sur le transport optimal.

Organisation de la thèse

Dans le chapitre 1 nous rappelons toutes les notions préliminaires et les résultats connus nécessaires à la compréhension du reste des chapitres. En particulier, la section 1.1 est divisée en quatre parties. La première est entièrement consacrée à l’entropie relative. Nous rappelons la définition classique des mesures de probabilité, l’extension aux mesures non bornées et des propriétés importantes telles que la formule variationnelle (1.2) et la propriété additive (1.6). Dans la deuxième partie, nous étudions les propriétés cinématiques des processus de diffusion de Markov. Après avoir rappelé les définitions de base, nous introduisons la notion (et la notation) de mesures de chemin sur l’espace de trajectoires continues (1.10) comme solution au problème de martingale associé à une EDS du type (1.7). Puis nous rappelons les notions de vitesses stochastiques de Nelson, 1.1.9, et donnons leur interprétation physique à l’aide des équations de transport et de Fokker-Plank. À ce stade, nous présentons des exemples que nous utiliserons tout au long du manuscrit, comme le mouvement Brownien, le processus d’Ornstein-Uhlenbeck et le semi-groupe de Kolmogorov. La troisième partie contient les principaux résultats de la théorie de Girsanov dans le théorème 1.1.14 et l’idée de la preuve. Le théorème de Girsanov jouera un rôle crucial dans la section 3.5 et au chapitre 6. Dans la quatrième partie, nous nous concentrons sur les semigroupes de Markov et le critère de Bakry-Emery via l’opérateur Γ_2 . Dans la section 1.2 nous résumons les résultats très basiques du transport optimal dans le contexte euclidien avec une attention particulière au cas quadratique de la distance de Wasserstein.

Dans le chapitre 2 nous présentons le protagoniste principal du manuscrit, le problème de Schrödinger: dans ce chapitre, nous passons sur beaucoup de détails, pour lesquels nous renvoyons à la littérature et nous renonçons aux cas plus généraux où ces résultats sont valables pour des arguments plus directs et explicites afin d’avoir un chapitre de taille raisonnable. La section 2.1 concerne la définition du problème et l’existence de résultats uniques. Dans la section 2.2 nous nous intéressons à la solution, c’est à dire, les ponts et les interpolations entropiques, et nous analysons quelques propriétés cinématiques. Puis dans la section 2.3 nous nous intéressons à la valeur optimale du problème de Schrödinger et nous proposons des moyens alternatifs pour définir le coût entropique, analogue à la distance Wasserstein. Dans les sections 2.4 et 2.5 nous étudions respectivement la connexion du problème de Schrödinger au transport optimal et le comportement de l’entropie

relative le long des interpolations entropiques, via le critère Γ_2 . Enfin, dans la section 2.6, nous donnons un aperçu de la motivation physique statistique de la question originale adressée par Schrödinger et une revue historique de la littérature sur ce sujet.

Le chapitre 3 traite des analogies entre le problème de Schrödinger et le transport optimal (OT). Nous donnons une nouvelle preuve de la double formulation de Kantorovich pour le coût entropique dans la section 3.4 et une formulation à la Benamou-Brenier dans la section 3.5. Comme application nous démontrons à la section 3.6 des inégalités de contractions pour le coût entropique qui, à la limite, donnent les résultats classiques dans la distance de Wasserstein (12). Enfin, dans la dernière section 3.7 nous donnons des exemples explicites d'interpolations entropiques et de McCann dans deux cas simples où des calculs explicites sont possibles et deux exemples numériques simples.

Dans le chapitre 4 nous montrons la propriété de concavité (26) et la propriété de régularité (30) sous la condition $CD(0, N)$. Nous dérivons les inégalités EVI et de contraction dimensionnelles [eqref eq-intro-contr-2n](#) pour le coût entropique.

Dans le chapitre 5, nous prouvons sous le $CD(\kappa, \infty)$, l'inégalité EVI via deux méthodes alternatives.

Dans le chapitre 6 nous utilisons une version simple du problème de Schrödinger avec $\varepsilon = 1$ fixé pour donner une preuve rapide de l'inégalité de concentration dans la section 6.1 et de l'inégalité de Prékopa-Leindler dans la section 6.2, via une méthode de couplage et le théorème de Girsanov.

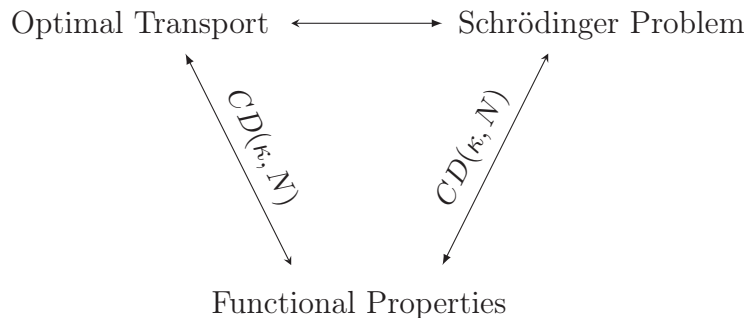
Les chapitres 1 et 2 ne contiennent aucun nouveau résultat, contrairement au reste du manuscrit. Les chapitres 3 et 4 sont des versions légèrement modifiées de, respectivement, [GLR17] et [Rip17] dans lesquelles nous avons ajouté quelques commentaires et utilisons des notations cohérentes avec le reste du manuscrit. Nous avons décidé de les retranscrire telles qu'elles apparaissent dans l'organisation de l'article (résumé, introduction, cadre, résultats ...) pour garder leur autonomie par rapport au reste du manuscrit, et laisser le lecteur libre de lire la partie qui l'intéresse, au prix de définitions parfois redondantes. Nous nous excusons pour ces quelques répétitions. Le chapitre 6 est l'objet d'un nouvel article en préparation avec Gentil et Léonard.

Introduction

In this work we will study the interplay between the classical fields of optimal transport and Markov semigroups with the less explored Schrödinger entropic minimization problem. As a result of the analogies and relations between these theories we will derive some functional inequalities.

The introduction is organized in three parts. In a first part we recover the basic notions of the optimal transport theory and some of the connections with curvature dimension conditions. In a second part we introduce the Schrödinger problem and present some of the new results of this manuscript about the analogies with optimal transport. The third part contains our contributions to the applications of these analogies to functional properties and inequalities via the Bakry-Emery theory.

This can be summarized in the following scheme.



I. Optimal transport and Bakry-Emery theory

We start by giving an informal picture of the very definition of the Monge-Kantorovich problem, its dynamic counterpart and few important and well known properties on \mathbb{R}^n . We will discuss these topics in more details later at Section 1.2.

Optimal Transport

The fathers of the optimal transport theory are Monge (1781) [Mon81] and Kantorovich (1942) [Kan42a]. The idea is to transport in the most economical way some mass between two prescribed distributions. Let us choose for simplicity, as state space the Euclidean space \mathbb{R}^n . We denote by $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ the prescribed configurations and $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ the cost function. Then the Monge-Kantorovich problem consists in,

$$\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) \pi(dx dy); \pi \in \Pi(\mu_0, \mu_1) \right\}, \quad (1)$$

where $\Pi(\mu_0, \mu_1)$ is the set of all probability measures on the product space $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ_0, μ_1 namely, all $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that

$$\pi_0(dx) := \pi(dx \times \mathbb{R}^n) = \mu_0(dx), \quad \pi_1(dy) := \pi(\mathbb{R}^n \times dy) = \mu_1(dy). \quad (2)$$

The Monge-Kantorovich problem admits also a dynamical formulation, proposed by McCann in [McC97], taking into account the whole trajectory of the system and not only the initial and final position. It can be written as,

$$\inf \left\{ \int_{\mathbb{R}^n} C[(\mathbf{T}_t)_{0 \leq t \leq 1}] d\mu_0; \mathbf{T}_0 = Id, \mathbf{T}_1 \# \mu_0 = \mu_1 \right\}, \quad (3)$$

where C is a non-negative function on the space of the continuous trajectories on \mathbb{R}^n , $\Omega := C([0, 1], \mathbb{R}^n)$. In order to be equivalent to its static version (1), the cost function $C : \Omega \rightarrow \mathbb{R}^+$ in (3) has to satisfy the relation

$$c(x, y) = \inf \{ C[(z_t)_{0 \leq t \leq 1}]; z_0 = x, z_1 = y \}.$$

For instance in the quadratic case, $C[(z_t)] = \int_0^1 |\dot{z}_t|^2 / 2 dt$.

A keystone result is due to Brenier who, in the 1980s, characterized the optimal map, in the quadratic case $c(x, y) = |x - y|^2$. Brenier's Theorem establishes that whenever $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, that is the set of probability measures with second order moment finite, and μ_0 is absolutely continuous with respect to the Lebesgue measure, then the optimal plan is given by

$$\pi = (Id \times \nabla \varphi) \# \mu_0 \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$$

where $\nabla \varphi$ is the gradient of a convex function and is called *Brenier's map*. With the push-forward notation we have that $\nabla \varphi \# \mu_0 = \mu_1$, which means that for all regular enough real function h on \mathbb{R}^n ,

$$\int h d\mu_1 = \int h(\nabla \varphi) d\mu_0.$$

On the other hand, the solution of the quadratic dynamical problem is called *McCann* or *displacement interpolation* and is characterized by the formula

$$\mu_t^{Mc} := \mathbf{T}_t \# \mu_0 = [(1 - t)Id + t\nabla \varphi] \# \mu_0, \quad \forall 0 \leq t \leq 1.$$

where $\nabla \varphi$ is the Brenier map.

In the quadratic case, the optimal value of the Monge-Kantorovich problem gives rise to the square of a distance in the space $\mathcal{P}_2(\mathbb{R}^n) = \{\mu \in \mathcal{P}(\mathbb{R}^n), \int |x|^2 d\mu < \infty\}$, that is called the *Wasserstein distance* or *cost*, and is denoted,

$$W_2^2(\mu_0, \mu_1) = \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \pi(dxdy); \pi \in \Pi(\mu_0, \mu_1) \right\} \quad (4)$$

The *Wasserstein space* $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ is geodesic. This means that for any couple of probability measures $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, there exists a path $(\mu_t^{Mc})_{t \in [0, 1]}$ in $\mathcal{P}_2(\mathbb{R}^n)$ such that for any $s, t \in [0, 1]$,

$$W_2(\mu_s^{Mc}, \mu_t^{Mc}) = |t - s| W_2(\mu_0, \mu_1).$$

Such a path is a constant speed geodesic in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, see [AGS08, Ch. 7]. Formally, the optimal transport dynamics is driven by the transport equation,

$$\partial_t \mu_t + \nabla \cdot (\mu_t \nabla \theta_t) = 0$$

for all $t > 0$, to be meant in a weak sense, and initial condition $\mu_{t=0} = \mu_0$. Here θ_t is the solution of the Hamilton-Jacobi equation,

$$\partial_t \theta_t + \frac{|\nabla \theta_t|^2}{2} = 0 \tag{5}$$

with initial condition $\theta_0(x) = \varphi(x) - |x|^2/2$. One issue of optimal transport maps is that they are in general not smooth.

The Wasserstein distance enjoys two alternative formulations useful when dealing with applications of optimal transport. The first one is the dual Kantorovich [Kan42b] formulation. It allows us to write

$$W_2^2(\mu_0, \mu_1) = \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int \mathcal{Q}_1 \psi d\mu_0 - \int \psi d\mu_1 \right\} \tag{6}$$

where \mathcal{Q}_1 is the Hopf-Lax formula at time $t = 1$, $\mathcal{Q}_t \psi(x) = \inf_{y \in \mathbb{R}^n} \{ \psi(y) + |y - x|^2/t \}$ which is the viscosity solution of the Hamilton Jacobi equation (5). The second one is the Eulerian Benamou-Brenier formulation [BB00]. It states that for any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu, v)} \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 d\nu_t dt, \tag{7}$$

where the infimum runs over all paths $(\nu_t, v_t)_{t \in [0,1]}$ where $\nu_t \in \mathcal{P}(\mathbb{R}^n)$ and $v_t(x) \in \mathbb{R}^n$ are such that ν_t is absolutely continuous with respect to time in the weak sense of [AGS08, Ch. 1] for all $0 \leq t \leq 1$, $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and

$$\partial_t \nu_t + \nabla \cdot (\nu_t v_t) = 0, \quad 0 \leq t \leq 1.$$

Markov semigroups and Γ_2 -calculus

Optimal transport matched very soon another well known theory, the one of Markov semigroups and the Bakry-Emery theory, that we recall here briefly.

Given a Markov process $(X_t)_{t \geq 0}$ the associated Markov semigroup is defined for any real suitable measurable function f and any $t \geq 0$ as

$$T_t f(x) := E(f(X_t) | X_0 = x) = \int f(y) p_t(x, dy).$$

where $p_t(x, dy)$ is the transition probability kernel of the process $(X_t)_{t \geq 0}$. To any Markov semigroup we associate an *infinitesimal generator*

$$Lf = \lim_{t \rightarrow 0^+} \frac{T_t f - f}{t},$$

whenever the limit makes sense, and a *reversing* measure $m \in \mathcal{M}_+(\mathbb{R}^n)$ satisfying

$$\int f T_t g dm = \int g T_t f dm,$$

whenever the integral exists. Hence for any suitable couple of functions $f, g \in L^2(m)$ we define the symmetric non-negative and bi-linear *carré du champ* operator as,

$$\Gamma(f, g) = \frac{1}{2}[L(fg) - fLg - gLf].$$

In 1985 in the founding paper [BE85], Bakry and Emery introduced a notion of curvature-dimension bound via the *iterated carré du champ* operator

$$\Gamma_2(f, g) = \frac{1}{2}[\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(Lf, g)].$$

We say that L satisfies the Bakry-Emery condition $CD(\kappa, N)$ for some $\kappa \in \mathbb{R}$ and $N \in [1, \infty)$, if for all suitable function $f \in L^2(m)$

$$\Gamma_2(f) \geq \kappa\Gamma(f) + \frac{1}{N}(Lf)^2. \quad (8)$$

Roughly speaking the interpretation of κ as a curvature and N as a dimension can be seen in a Riemannian manifold via Bochner's formula. For instance consider $L = \Delta_g$ on a smooth complete connected Riemannian manifold (M, g) . Then the Bochner identity is

$$\Gamma_2(f) = \frac{1}{2}L(|\nabla f|^2) - \nabla f \cdot \nabla Lf = \text{Ric}(\nabla f, \nabla f) + \|\nabla^2 f\|^2$$

where Ric is the Ricci curvature and ∇^2 is the Hessian. Hence we can deduce that the $CD(\kappa, N)$ condition is satisfied if and only if the Ricci curvature is bounded from below by κ and the dimension of the manifold is bounded from above by N . This curvature-dimension criterion was first introduced by Bakry and Emery in [BE85] as a sufficient condition for a class of functional inequalities such as hypercontractivity, the log-Sobolev inequality or the Poincaré inequality etc.

Optimal Transport and Curvature-Dimension

The connections between optimal transport theory and curvature-dimension conditions are nowadays innumerable. Far from being exhaustive, we mention here some of the main founding results of the last twenty years.

Displacement convexity. In the pioneering 1997 paper [McC97] McCann introduced the concept of *displacement convexity* for functionals of the type,

$$\mathcal{F}(\mu) = \int F(\rho) dm \quad (9)$$

where $\mu \in \mathcal{P}(\mathbb{R}^n)$, ρ is the density $\rho = d\mu/dm$ and m is the equilibrium measure. We say that a functional \mathcal{F} is κ -geodesically (or displacement) convex in the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ for some $\kappa \in \mathbb{R}$, if for all McCann interpolation $(\mu_t^{Mc})_{0 \leq t \leq 1}$ and all $0 \leq t \leq 1$,

$$\mathcal{F}(\mu_t^{Mc}) \leq (1-t)\mathcal{F}(\mu_0) + t\mathcal{F}(\mu_1) - \frac{\kappa}{2}t(1-t)W_2^2(\mu_0, \mu_1). \quad (10)$$

Transport and HWI inequalities. In the 2000 paper [OV00] Otto and Villani established a link between the transport Talagrand inequality and the log-Sobolev inequality. More precisely, let us introduce the two principal information-theoretical functionals: the *relative entropy* defined for any couple of probability measures as

$$H(\mu|\nu) = \int \log \frac{d\mu}{d\nu} d\mu$$

whenever $\mu \ll \nu$ and $+\infty$ otherwise. While the *Fisher information*,

$$I(\mu|\nu) = \int \left| \nabla \log \frac{d\mu}{d\nu} \right|^2 d\mu.$$

Otto and Villani proved that any measure $m = e^{-V} dx$ with V regular enough satisfying the log-Sobolev inequality

$$H(\mu|m) \leq \frac{1}{2C} I(\mu|m)$$

for some $C > 0$ and for any probability measure $\mu \ll m$, then it also satisfies the Talagrand inequality,

$$W_2(\mu, m) \leq \sqrt{\frac{2H(\mu|m)}{C}}. \quad (11)$$

They also proved the reverse implication, through the introduction of a new inequality relating entropy, Fisher's information and Wasserstein distance, hence called *HWI* inequality, which writes for any $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ and $\mu \ll m$, as

$$H(\mu|m) \leq W_2(\mu, m) \sqrt{I(\mu|m)} - \frac{\kappa}{2} W_2^2(\mu, m).$$

Contraction. In 2005 von-Renesse and Sturm [Sv05] proved the equivalence between $CD(\kappa, \infty)$ and contraction in Wasserstein distance. This means that given a diffusion operator L and denoting by $(T_t)_{t \geq 0}$ the associated Markov semigroup with invariant measure $m \in \mathcal{M}_+(\mathbb{R}^n)$, saying that it satisfies the Bakry-Emery condition (8) with $N = \infty$ is equivalent of saying that the Wasserstein distance is a contraction along the flow of $(T_t)_{t \geq 0}$, that is

$$W_2^2(T_t f dm, T_t g dm) \leq e^{-2\kappa t} W_2^2(f dm, g dm). \quad (12)$$

This result was later established in the flat but dimensional case, $CD(0, N)$ in [BGL15] and [Kuw15],

$$W_2^2(T_s f dx, T_t g dx) \leq W_2^2(f dx, g dx) + 2N(\sqrt{t} - \sqrt{s})^2, \quad \forall s, t \geq 0. \quad (13)$$

It has been recently extended to the general $CD(\kappa, N)$ in [EKS15] and improved in [BGG16] and [BGGK16], both in the Euclidean and the Riemannian cases, writing as,

$$W_2^2(T_t f m, T_t g m) \leq W_2^2(f m, g m) - \frac{2}{N} \int_0^t e^{-2\kappa(t-u)} [H(T_u g|m) - H(T_u f|m)]^2 du. \quad (14)$$

Gradient flow. In 1998 a new approach was proposed by Jordan, Kinderlehrer and Otto [JKO98] on the heat flow, solution of the drift-diffusion equation

$$\partial_t \mu_t = \Delta \mu_t + \nabla \cdot (\mu_t \nabla V).$$

They showed that it can be seen as the gradient flow in the Wasserstein space with its Riemannian-like metric, of the relative entropy functional with respect to the invariant measure $m = e^{-V} dx$. A similar result was proved a bit later in [Ott01] for the porous medium equation. This approach opened a new chapter in the study of integral functional such as (9).

Let us recall the basic definition of gradient flow on \mathbb{R}^n . Given the differential equation,

$$\partial_t x_t = -\nabla F(x_t), \quad t \geq 0 \quad (15)$$

with initial condition $x_0 = x$, then we say that $(x_t)_{t \geq 0}$ is the gradient flow of F .

Evolutionary variational inequality The concept of κ -displacement convexity, gradient flows and the contraction property are strictly related. In the textbook [AGS08] the authors Ambrosio Gigli and Savaré, introduced a new inequality leading them together. It is the *Evolutionary variational equation* (EVI) and, roughly speaking, it defines the gradient flows in metric spaces, with the main advantage of avoiding differentiation of non-smooth trajectories. In the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ it writes as,

$$\frac{1}{2} \frac{d^+}{dt} W_2^2(T_t u, v) + \frac{\kappa}{2} W_2^2(T_t u, v) \leq \mathcal{F}(v) - \mathcal{F}(T_t u) \quad (16)$$

for all $t \geq 0$, and any suitable function u and v . Later on it became a classical strategy to derive contraction in Wasserstein distance along the flow $(T_t)_{t \geq 0}$ from EVI. Moreover it was also proved by Daneri and Savaré [DS08] that the EVI inequality with constant $\kappa \in \mathbb{R}$ implies the κ -displacement convexity (10) of the functional \mathcal{F} . For completeness let us recall that displacement κ -convexity (10) of a functional implies the contraction (12) for its gradient flow.

It is easily seen in the Euclidean setting that for x_t, y_t solutions of (15) with initial condition respectively $x_0 = x$ and $y_0 = y$,

$$\begin{aligned} \frac{d}{dt} |x_t - y_t|^2 &= 2 \langle x_t - y_t, \partial_t x_t - \partial_t y_t \rangle \\ &= -2 \langle x_t - y_t, \nabla F(x_t) - \nabla F(y_t) \rangle \\ &\stackrel{*}{\leq} -2\kappa |x_t - y_t|^2 \end{aligned}$$

and Gronwall's lemma yields to the contraction,

$$|x_t - y_t|^2 \leq e^{-2\kappa t} |x - y|^2, \quad \forall t \geq 0.$$

The κ -convexity appears in the marked inequality, in the form of κ -monotonicity of ∇F ,

$$\langle \nabla F(x) - \nabla F(y), x - y \rangle \geq \kappa |x - y|^2.$$

The EVI approach was successively developed in the dimensional case $CD(\kappa, N)$ by Erbar Kuwada and Sturm in [EKS15].

LSV theory The milestone of the links between optimal transport and Ricci curvature is the Lott-Sturm-Villani theory, simultaneously and independently developed in [LV09] and [Stu06a]. They extended the notion of $CD(\kappa, \infty)$ condition to metric measured spaces (\mathcal{X}, d, m) through convexity properties along geodesics between probability measures on \mathcal{X} .

II. Schrödinger problem and Optimal Transport

In this second part we introduce the Schrödinger problem and briefly recall the definitions of the *entropic interpolation* and *cost*, respectively solution and optimal value of the Schrödinger problem. After recalling that the Schrödinger problem is, roughly speaking, a regular approximation of the Monge-Kantorovich problem, we will explore some new analogy.

Schrödinger problem

Introduced in 1932 [Sch32], the Schrödinger problem is, in modern terms, an entropy minimization problem. We fix a *reference path measure* $R \in \mathcal{M}_+(\Omega)$, where $\Omega = C([0, 1], \mathbb{R}^n)$. We denote the joint law of the initial and final position as,

$$R_{01} = (X_0 \times X_1)_{\#} R \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n),$$

where $(X_t)_{0 \leq t \leq 1}$ is the canonical process on Ω , defined by

$$X_t(\omega) = \omega_t \in \mathbb{R}^n, \quad \forall \omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega, \quad 0 \leq t \leq 1.$$

For any couple of prescribed marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, the Schrödinger problem consists in minimizing the entropy relative to R_{01} among all the probability measures $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals μ_0 and μ_1 , that is,

$$\inf\{H(\pi|R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n), \pi_0 = \mu_0, \pi_1 = \mu_1\}, \quad (17)$$

where the marginals π_0 and π_1 are defined in (2). As the Monge-Kantorovich problem, it admits the following dynamical formulation,

$$\inf\{H(P|R); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \quad (18)$$

where we denoted by P_t (for $t = 0, 1$) the time projection at time t of the path measure P , namely

$$P_t := (X_t)_{\#} P = P(X_t \in \cdot) \in \mathcal{P}(\mathbb{R}^n)$$

where $(X_t)_{t \geq 0}$ is the canonical process.

Problems (17) and (18) are equivalent, in the sense that they share the same optimal value, attained respectively by $\hat{\pi}$ and \hat{P} , mutually characterized by the relation,

$$\hat{P}(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{xy}(\cdot) \hat{\pi}(dxdy) \in \mathcal{P}(\Omega)$$

that means that $\hat{P}_{01} = \hat{\pi}$ and that \hat{P} shares the same bridges $\hat{P}^{xy} = R^{xy} := R(\cdot | X_0 = x, X_1 = y)$ of R . In the manuscript we will mainly consider as reference measure, the Markov reversible path measure R associated with generator $L = (\Delta - \nabla V \cdot \nabla)/2$ and reversing measure $\mathbf{m} = e^{-V} dx$, for a regular enough potential V . Under some sufficient assumption on the reference measure and the marginals, which will be specified in Chapter 2 the Schrödinger problem admits a solution. Uniqueness follows directly from the strictly convex nature of the Schrödinger problem.

In the two basic examples, when $V = 0$ and $V = -|x|^2/2$, which correspond respectively to the Brownian motion and the Ornstein-Uhlenbeck process, we will show that these assumptions are satisfied, whereas the assumptions on the marginals reduce to

$$H(\mu|m) < \infty, \quad \mu \in \mathcal{P}_2(\mathbb{R}^n)$$

that is, the marginal measures have finite relative entropy with respect to the equilibrium measure (respectively $m = \mathcal{L}$ and $m = e^{-|x|^2/2}/(2\pi)^{n/2}dx$), and second order moment finite.

The solution of (18) is called *entropic bridge* and is characterized by the formula,

$$\hat{P} = f(X_0)g(X_1)R \in \mathcal{P}(\Omega)$$

where f, g are measurable non negative functions, satisfying the initial and final conditions,

$$\begin{cases} \mu_0 = fT_1g \\ \mu_1 = gT_1f, \end{cases}$$

where $(T_t)_{t \geq 0}$ is the Markov semigroup associated with the reference R . The *entropic interpolation* is defined as the time marginal flow of the entropic bridge, $\mu_t := \hat{P}_t = (X_t)_\# \hat{P} \in \mathcal{P}(\mathbb{R}^n)$ and characterized by the formula,

$$d\mu_t := T_t f T_{1-t} g dm, \forall 0 \leq t \leq 1.$$

We denote $\mu_t = d\mu_t/dx$ the density of the entropic interpolation with respect to the Lebesgue measure, it is a smooth solution of the transport equation,

$$\partial_t \mu_t + \nabla \cdot (\mu_t \nabla \theta_t) = 0,$$

where the vector field

$$\nabla \theta_t = (\nabla \varphi_t + \nabla \psi_t)/2 \tag{19}$$

is called *current velocity* and the functions ψ_t, φ_t are the *Schrödinger potentials* defined by $\varphi_t = \log T_t f$, and $\psi_t = \log T_{1-t} g$.

From the point of view of random process, if R is associated with the SDE (in a sense that will be made clear later at Section 1.1.2)

$$dX_t = -\frac{\nabla V}{2} dt + dW_t, \tag{20}$$

then the entropic bridge, \hat{P} is associated to the SDE,

$$dX_t = \left(-\frac{\nabla V}{2} + \nabla \psi_t \right) dt + dW_t, \tag{21}$$

where ψ_t is a Schrödinger potential.

Provided that a solution $\hat{P} \in \mathcal{P}(\Omega)$ of (18) exists, it attains the optimal value

$$H(\hat{P}|R) = \inf\{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\}.$$

Thus we define the *entropic cost* between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ as,

$$\mathcal{A}(\mu_0, \mu_1) := \inf\{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \frac{1}{2}[H(\mu_0|m) - H(\mu_1|m)]. \tag{22}$$

It is symmetric and non-negative. The necessity of withdrawing the constant term $[H(\mu_0|m) - H(\mu_1|m)]/2$ will be clear at Sections 2.3 and 3.5. As an alternative, we can define the *forward* and *backward costs* as,

$$\vec{\mathcal{A}}(\mu_0, \mu_1) := \inf\{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - H(\mu_0|m)$$

$$\overleftarrow{\mathcal{A}}(\mu_0, \mu_1) := \inf\{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - H(\mu_1|m).$$

They have the drawback of being non symmetric but they admit a useful dual Kantorovich formulation that will be proved at Section 3.4.

We can recover from here the original question addressed by Schrödinger. Roughly speaking, assume that (20) models the position of a particles system with initial law $\mathcal{L}(X_0) = \mu_0$. We observe the system at time, say, $t = 1$ and we see that $\mathcal{L}(X_1) = \mu_1$ is different from what we expect according to the SDE (20). This is a rare event, with a very low probability to happen. It means that an unexpected drift, namely $\nabla\psi_t$, perturbed the motion of the system, that is thus driven by the SDE (21). Moreover the most likely dynamics of the *perturbed* particles system is \hat{P} , that is the one who minimizes the relative entropy with respect to R .

Schrödinger problem vs. Optimal Transport

Besides the fact of being transportation problems between two prescribed configurations, the Monge-Kantorovich and the Schrödinger problems at a first sight don't seem to have much in common. One is deterministic the other is probabilistic, involving a reference random process and the relative entropy functional. Nevertheless several connections and analogies do actually hold between optimal transport and entropy minimization. Let us discuss informally here the main results.

The link between the Schrödinger problem and optimal transport was first established in [Mik04] not actually in terms of minimal entropy but of h -transforms. Considering problem (17) with a scaling factor $\varepsilon > 0$,

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \frac{\varepsilon}{2}[H(\mu_0|m) + H(\mu_1|m)]$$

where R^ε is associated to the generator $L^\varepsilon = \varepsilon\Delta/2$ then,

$$\lim_{\varepsilon \rightarrow 0^+} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{W_2^2(\mu_0, \mu_1)}{2} \tag{23}$$

and

$$\lim_{\varepsilon \rightarrow 0} \hat{P}^\varepsilon = \hat{P}$$

where $\hat{P} \in \mathcal{P}(\Omega)$ is the unique solution of the dynamic Monge-Kantorovich problem with quadratic cost (4). In [Léo12a] Léonard proved that this limit is actually true whenever the reference measure R^ε in (18) and the cost function in (3) satisfy a large deviation principle,

$$R^\varepsilon(A) \underset{\varepsilon \rightarrow 0^+}{\asymp} e^{-\inf_{\omega \in A} C(\omega)/\varepsilon},$$

for any Borel set $A \subset \Omega$. A formal but persuasive argument to prove the convergence (23) is to consider the static versions (17) and (1) respectively with Brownian motion as reference measure, and the quadratic distance as cost function, namely

$$R_{01}^\varepsilon(dxdy) = \frac{e^{-\frac{|x-y|^2}{2\varepsilon}}}{(2\pi\varepsilon)^{n/2}} dxdy, \quad c(x, y) = |x - y|^2.$$

An explicit computation shows that

$$\varepsilon H(\pi|R_{01}^\varepsilon) = \int \frac{|x-y|^2}{2} d\pi + \varepsilon \int \log \frac{d\pi}{dx dy} d\pi + \varepsilon \log(2\pi\varepsilon)^{n/2}.$$

Hence at the limit

$$\lim_{\varepsilon \rightarrow 0^+} \varepsilon H(\pi|R_{01}^\varepsilon) = \int \frac{|x-y|^2}{2} d\pi.$$

Optimizing over all $\pi \in \Pi(\mu_0, \mu_1)$, we obtain

$$\lim_{\varepsilon \rightarrow 0^+} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{W_2^2(\mu_0, \mu_1)}{2}.$$

This is consistent with the fact that the entropic cost and the entropic interpolations approximate the Wasserstein distance and the McCann interpolation when the rescaling parameter ε vanishes. But even for fixed $\varepsilon > 0$ our two main problems enjoy some important analogy. In [MT06] Mikami and Thieullen and later in [GLR17], Gentil Léonard and the author proved with different methods a dual Kantorovich formulation for the Schrödinger problem, analogous to (6). It states that for every $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that a solution of (18) exists, then

$$\mathcal{A}(\mu_0, \mu_1) = \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int \psi d\mu_1 - \int \mathcal{Q}_1 \psi d\mu_0 \right\},$$

where

$$\mathcal{Q}_1 \psi := \log T_1 e^\psi, \quad (24)$$

and $(T_t)_{t \geq 0}$ is the reference semigroup associated with R . Note that $\mathcal{Q}_t \psi$ satisfies the Hamilton-Jacobi-Bellman equation,

$$\partial_t u + \frac{|\nabla u|^2}{2} + Lu = 0$$

with initial condition $u_{t=0} = \psi$, in analogy with the Hopf-Lax semigroup appearing in (6) defining the solution of the Hamilton-Jacobi equation (5).

The second important analogy is the Benamou-Brenier formulation for the entropic cost proved by Chen Georgiu and Pavon in [CGP16] in the special case of Brownian motion with $L = \Delta/2$ as reference measure, and for a general Kolmogorov semigroup with $L = (\Delta - \nabla V \cdot \nabla)/2$ in [GLR17, Section 5]. It states that the entropic cost can be written as a minimal action problem,

$$\mathcal{A}(\mu_0, \mu_1) = \frac{1}{2} \inf \int_0^1 \int_{\mathbb{R}^n} \left(|v_t(z)|^2 + \frac{1}{4} |\nabla \log \rho_t(z)|^2 \right) \mu_t(z) dz dt,$$

where the infimum runs over all the couples $(\mu_t, v_t)_{0 \leq t \leq 1}$ such that μ_t is a probability on \mathbb{R}^n for any $0 \leq t \leq 1$ satisfying the marginals constraints $\mu_{t=0} = \mu_0$, $\mu_{t=1} = \mu_1$ and such that

$$\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0.$$

Unlike the Benamou-Brenier formulation (7) where we minimize the classical kinetic action, here we minimize the sum on the kinetic plus the osmotic action (that we will define at Section 2.3) which takes into account the osmotic pressure due to the diffusion phenomenon.

III. Applications

From this comparative study of the Schrödinger and the optimal transport problem, a natural question arises: can we develop some new connection between the Schrödinger setting and the Bakry-Emery theory, in analogy to the several connections between optimal transport and curvature dimension, recalled in the first part of the introduction? The answer is yes. We mention here the main results that will be proved in the manuscript as examples of these connections.

Schrödinger problem and Curvature-Dimension

The setting of the Schrödinger problem can be studied under curvature-dimension conditions. Recalling that entropic interpolations are (smooth) approximations of the McCann interpolations, and recalling the important role that plays convexity of the entropy functional along geodesics the natural question that arises is: what is the behavior of the relative entropy with respect to the equilibrium measure m , along entropic interpolations? The surprising answer is that, despite in the Schrödinger setting we lose the geometric properties of geodesics, entropic interpolations preserve the convexity of the entropy. It follows by direct Γ_2 -computations, that the first and second derivative of the entropy along entropic interpolations, can be written in terms of Γ and Γ_2 operators. Denote $(\mu_t)_{0 \leq t \leq 1}$ the entropic interpolation between two probability measures μ_0, μ_1 such that a solution of (18) exists, then

$$\begin{aligned} \frac{d}{dt}H(\mu_t|m) &= \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t = 2 \int \Gamma(\log \rho_t, \theta_t) d\mu_t \\ \frac{d^2}{dt^2}H(\mu_t|m) &= 2 \int \Gamma_2\psi_t + \Gamma_2\varphi_t d\mu_t = \int (4\Gamma_2\theta_t + \Gamma_2 \log \rho_t) d\mu_t \end{aligned} \quad (25)$$

where Γ_2 is the iterated carré du champs operator associated to the generator L of the reference measure R , m is the reversing measure of R , θ_t appears in the current velocity (19) and $\rho_t = d\mu_t/dm$ is the density of the entropic interpolation with respect to the reversing measure m . Applying the Bakry-Emery condition (8) with $\kappa = 0$ and infinite dimension to (25), we obtain immediately the convexity of the relative entropy along entropic interpolation. This was first shown by Léonard in [Léo17]. This result has been extended by the author in [Rip17] under the dimensional condition $CD(0, N)$. It is shown that,

$$[0, 1] \ni s \mapsto \exp\left(-\frac{1}{n}H(\mu_s|\mathcal{L})\right), \quad (26)$$

is concave. In the trivial example in which the entropic interpolation coincides with the heat flow, $\mu_t = T_t f$ for some smooth probability density f , this result recovers Costa's Theorem [Cos85] that is one of the starting points of the Bakry-Emery-Ledoux theory. The concavity property of the exponential entropy (26) holds also along displacement interpolations and is a crucial result in [EKS15]. The gain of dealing with entropic interpolations, is that, being smooth, computations are not formal as in the optimal transport setting but rigorous.

Application: functional inequalities

Optimal transport is an efficient tool to prove functional and geometric inequalities under appropriate curvature-dimension conditions (see for instance [BGL14, Ch. 9] or [Vil09,

Part III]). The analogies and connections to optimal transport, together with the regularity of trajectories and the characterization formula in terms of Markov semigroups, suggest that the Schrödinger problem might also be a useful tool to prove functional inequalities under curvature-dimension bounds.

We propose in this manuscript some new proof of standard functional inequalities, through the notions of entropic interpolations and entropic cost.

- In [GLR17, Sec. 6] we prove a contraction inequality with respect to the entropic cost, along the semigroup $(T_t)_{t \geq 0}$ associated with the reference measure R . In particular we prove that, for R^ε associated with generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ with $\varepsilon > 0$,

(a) If V satisfies $\text{Hess}(V) \geq \kappa \text{Id}$, for some $\kappa \in \mathbb{R}$, then for any $t \geq 0$,

$$\mathcal{A}_b^\varepsilon(T_{\alpha_t(b)} u \mathbf{m}, T_t v \mathbf{m}) \leq \mathcal{A}_{\beta_t(b)}^\varepsilon(u \mathbf{m}, v \mathbf{m}) + \varepsilon[H(T_{\alpha_t(b)} u \mathbf{m} | \mathbf{m}) - H(u \mathbf{m} | \mathbf{m})], \quad (27)$$

where u and v are probability densities with respect to the reversing measure $\mathbf{m} = e^{-V} dx$, and

$$\alpha_t(b) = t + \frac{1}{\kappa} \log \left(\frac{e^{-\varepsilon \kappa b}}{1 + e^{\kappa t} (e^{-\varepsilon \kappa b} - 1)} \right) \quad \beta_t(b) = -\frac{1}{\kappa \varepsilon} \log(1 + e^{\kappa t} (e^{-\varepsilon \kappa b} - 1))$$

where: if $\kappa \leq 0$, $b \in (0, \infty)$ and if $\kappa > 0$, $b \in (0, -\log(1 - e^{-\kappa t})/(\kappa \varepsilon))$.

(b) If $V = 0$ then for any $t \geq 0$,

$$\mathcal{A}^\varepsilon(T_t u \mathcal{L}, T_s v \mathcal{L}) \leq \mathcal{A}^\varepsilon(u \mathcal{L}, v \mathcal{L}) + \frac{n}{2} (\sqrt{t} - \sqrt{s})^2 + \varepsilon[H(T_t u | \mathcal{L}) - H(u | \mathcal{L})]. \quad (28)$$

In the first case we introduced a generalization of entropic cost $\mathcal{A}_\tau^\varepsilon$ defined as,

$$\varepsilon \mathcal{A}_\tau^\varepsilon(\mu_0, \mu_1) = \varepsilon H(\mu_0 | m) + \sup \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}_\tau^\varepsilon \psi d\mu_0, \quad \psi \in C_b(\mathbb{R}^n) \right\}$$

where for any $\psi \in C_b(\mathbb{R}^n)$,

$$\mathcal{Q}_\tau^\varepsilon \psi = \varepsilon \log T_{\varepsilon \tau}(e^{\psi/\varepsilon}).$$

Taking the limit as ε goes to zero of the contraction inequalities in entropic cost, we recover the classical contraction in Wasserstein distance (12), (13). The dimensional case (28) is successively improved by the author in [Rip17], obtaining,

$$\mathcal{A}(T_t u, T_t v) \leq \mathcal{A}(u, v) - n \int_0^t \sinh^2 \left(\frac{H(T_s u | \mathcal{L}) - H(T_s v | \mathcal{L})}{2n} \right) ds. \quad (29)$$

analogous of (14) for $\kappa = 0$. The proof of (27) and (28) are based on the commutation of the semigroup $(T_t)_{t \geq 0}$ and $(\mathcal{Q}_t)_{t \geq 0}$ defined at (24), and is proved at Lemma 3.6.2. While (29) is derived as a consequence of the EVI inequality for the entropic cost, in Corollary 4.3.8.

- The EVI inequality is proved for the entropic cost under the $CD(\kappa, \infty)$ and $CD(0, N)$ conditions. Taking advantage of the regularity properties of the entropic interpolation, we prove in Theorem 4.3.4 that the function,

$$[0, \infty) \ni t \mapsto \mathcal{A}(u \mathcal{L}, T_t v \mathcal{L}),$$

where \mathcal{A} is the entropic cost (22), is differentiable. In particular, for $t = 0$ it holds,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u\mathcal{L}, T_t v\mathcal{L}) = -\frac{1}{2} \left. \frac{d}{ds} \right|_{s=1} H(\mu_s|\mathcal{L}) \quad (30)$$

where $(\mu_s)_{0 \leq s \leq 1}$ is the entropic interpolation between u and v . Hence the entropic interpolation is just a consequence of the concavity of (26). Under $CD(\kappa, \infty)$ with $\kappa \in \mathbb{R}$ we prove in Chapter 5 that for all $t \geq 0$,

$$\frac{d}{dt} \mathcal{A}^\varepsilon(u\mathbf{m}, T_t v\mathbf{m}) + \frac{\kappa}{2} \mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) - \frac{\varepsilon^2}{8} |\kappa| \int_0^1 I_{\mathbf{m}}(\mu_s) ds \leq \frac{1}{2} [H(u|\mathbf{m}) - H(T_t v|\mathbf{m})]$$

where $(\mu_s)_{0 \leq s \leq 1}$ is the entropic interpolation between $u\mathbf{m}$ and $v\mathbf{m}$, and $I_{\mathbf{m}}(\mu_s)$ is the Fisher's information (5.8) of μ_s with respect to the reversing measure $d\mathbf{m}$. We derive this result by means of two strategies already used one in [DS08] and the other in [BGL15] to derive EVI in Wasserstein distance.

We give a proof of the dimensional EVI for the entropic cost, as an immediate consequence of (30) and the Costa's type property (26),

$$\frac{d}{dt} \mathcal{A}(u\mathcal{L}, T_t v\mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n} [H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right).$$

• Finally we propose an easy proof of the Gaussian concentration inequality at Theorem 6.1.1, saying that for any $0 < p < 1$, $t > 0$, $x \in \mathbb{R}^n$ and any Borel measurable subset $A \subset \mathbb{R}^n$ such that $m^{x,t}(A) \geq p$, it holds

$$m^{x,t}(A^r) \geq 1 - \exp \left\{ -\frac{[r - r(p, \kappa, t)]^2}{2\tau(\kappa, t)} \right\}, \quad \forall r \geq r(p, \kappa, t),$$

where $A^r := \{x \in \mathbb{R}^n; d(x, A) \leq r\}$ is the r -enlargement of the subset A , and

$$\tau(\kappa, t) = \begin{cases} (1 - e^{-\kappa t})/\kappa & \text{if } \kappa \neq 0 \\ t & \text{if } \kappa = 0, \end{cases}$$

and $r(p, \kappa, t) := \sqrt{2\tau(\kappa, t) \log(1/p)}$. The proof is very fast. It is based on a coupling method and on the statistical physics nature of the Schrödinger problem (through Girsanov's Theorem 1.1.14); we don't even need the limit to the optimal transport. As a consequence we derive the Talagrand inequality (11).

• With the same method we obtain the Gaussian Prékopa-Leindler inequality recovering the same result of Lehec [Leh13]. We recall that it writes, for any $T > 0$ as

$$\log \int_{\mathbb{R}^n} e^\theta d\gamma_T \geq (1 - \lambda) \log \int_{\mathbb{R}^n} e^\psi d\gamma_T + \lambda \log \int_{\mathbb{R}^n} e^{\psi'} d\gamma_T,$$

where θ, ψ, ψ' are measurable $[-\infty, \infty)$ -valued functions on \mathbb{R}^n satisfying for any $0 < \lambda < 1$

$$\theta((1 - \lambda)x + \lambda x') \geq (1 - \lambda)\psi(x) + \lambda\psi'(x'), \quad \forall x, x' \in \mathbb{R}^n,$$

and $\gamma_T := \mathcal{N}(0, T\text{Id})$. As an application we derive the Prékopa-Leindler inequality for any log-concave measure m (Corollary 6.2.3) and its geometric counterpart, the Brunn-Minkowski inequality (Corollary 6.2.4).

Structure of the manuscript

The manuscript is organized as follows.

In Chapter 1 we collect all the preliminary notions and known results necessary for understanding the rest of the chapters. In particular Section 1.1 is divided into four parts. The first one is entirely devoted to relative entropy. We recall the classical definition for probability measures, the extension to unbounded measures, and important properties such as the variational formula (1.2) and the additive property (1.6). In the second part we investigate kinematic properties of Markov diffusion processes. After recalling the basic definitions, we introduce the notion (and notation) of path measures on the space of continuous trajectories (1.10) as solution of the martingale problem associated with a SDE of the type (1.7). Then we recall the concept of stochastic velocities due to Nelson, in Definition 1.1.9 and give their physical interpretation by means of the Fokker-Plank and the transport equations. At this point we introduce some example that will refer to all along the manuscript, such as the Brownian motion, the Ornstein-Uhlenbeck process and the Kolmogorov semigroup. The third part contains the main results of the Girsanov theory in Theorem 1.1.14 and a sketch of the proof. Girsanov's Theorem will play a crucial role in Section 3.5 and at Chapter 6. In the fourth part we focus on Markov semigroups, and the Bakry-Emery criterion via the Γ_2 -operator. In Section 1.2 we summarize the very basic results of optimal transport in the Euclidean setting with particular attention to the quadratic case of the Wasserstein distance.

In Chapter 2 we present the protagonist of the manuscript, the Schrödinger problem. In this chapter we avoid many details, for which we refer time to time to the literature, and renounce to the most generality in which these results are valid, in favor of more direct and explicit arguments and of a reasonable size of the chapter. Section 2.1 is about the definition of the problem and existence and uniqueness results. In Section 2.2 we focus on the solution, namely the entropic bridges and entropic interpolations, and analyze some kinematic properties. Then in Section 2.3 we are interested in the optimal value of the Schrödinger problem and propose alternative ways to define the entropic cost, the analogous of the Wasserstein distance. In Section 2.4 and Section 2.5 we investigate respectively the connection of the Schrödinger problem to optimal transport and the behavior of the relative entropy along entropic interpolations, via the Γ_2 criterion. Finally in Section 2.6 we give a flavor of the statistical physics motivation of the original question addressed by Schrödinger and an historical review of the literature on the topic.

Chapter 3 deals with the analogies between the Schrödinger problem and optimal transport (OT). We give a new proof of the Kantorovich dual formulation for the entropic cost in Section 3.4 and a Benamou-Brenier formulation in Section 3.5. As an application we prove at Section 3.6 some contraction results in entropic cost that, at the limit, give the classical contraction in Wasserstein distance (12). Finally in last Section 3.7 we give some explicit examples of entropic and McCann interpolations in two basic cases in which explicit computations are possible, that is between two Gaussian measures when the reference measure is either the heat either the Ornstein-Uhlenbeck semigroup, and two easy numerical examples.

In Chapter 4 we prove the concavity property (26) and the regularity property (30) under the $CD(0, N)$ condition. We derive the dimensional EVI and the dimensional contraction

(29) for the entropic cost.

In Chapter 5 we prove under the $CD(\kappa, \infty)$, the EVI inequality via two alternative methods.

In Chapter 6 we use a genuine Schrödinger problem with fixed $\varepsilon = 1$ to give a fast proof of the Concentration inequality at Section 6.1 and of the Prékopa-Leindler inequality at Section 6.2, via a coupling method and the Girsanov theorem.

Chapters 1 and 2 do not contain any new results, contrary to the rest of the manuscript. Chapters 3 and 4 are slightly modified versions of, respectively, [GLR17] and [Rip17] in which we added few comments and use notations consistent with the rest of the manuscript. We decided to report them in their article organization (abstract, introduction, setting, results...) to keep the autonomy with respect to the rest of the manuscript, and leave the reader free to read the part of interest, at the price of some redundant definition. We apologize for the occasional repetitiveness. Chapter 6 is the object of a new article in preparation with Gentil and Léonard.

Chapter 1

Preliminary tools

This chapter is meant to be a toolbox containing all the necessary instruments for a complete comprehension of the rest of the manuscript. All the results presented here are well known results, object, in more details and more generality, of a rich bibliography of manuals and papers, that we will mention time to time. Thus the only ambition of this chapter is to put together all these different topics, sharing uniform notation and the same final goal. It is organized in two parts: Section 1.1 is devoted to relative entropy and diffusion processes, while Section 1.2 to the optimal transport theory.

1.1 Actions of diffusion processes

This section is organized as follows. First at Section 1.1.1 we introduce the relative entropy that will play a central role all along the manuscript. Then at Section 1.1.2 we recall the basic notions of Markov diffusion processes and focus on kinematics properties. At Section 1.1.3 we present the main results of the Girsanov theory, and finally at Section 1.1.4 we discuss the principal facts of Markov semigroups and Γ_2 -theory.

1.1.1 Relative entropy

The *relative entropy* is a fascinating functional recurring in different fields of physics, mathematics and information theory. We start by giving a mathematical picture, far from being complete but sufficient to understand the rest of the following chapters.

Definition 1.1.1 (Relative entropy) *Let \mathcal{X} be a Polish space and $\mu, \nu \in \mathcal{P}(\mathcal{X})$ two probability measures on \mathcal{X} . The relative entropy is defined as,*

$$H(\mu|\nu) = \begin{cases} \int_{\mathcal{X}} \log \frac{d\mu}{d\nu} d\mu & \text{if } \mu \ll \nu, \\ +\infty & \text{otherwise.} \end{cases} \quad (1.1)$$

Some important properties of the entropy can be directly derived by its definition. Indeed, we can write,

$$H(\mu|\nu) = \int h\left(\frac{d\mu}{d\nu}\right) d\nu$$

where $h(a) = a \log a - a + 1$ for any $a \geq 0$ (with the convention that $0 \log 0 = 0$) and $+\infty$ for any negative a . The function $h(\cdot)$ is non-negative and strictly convex, and

it is straightforward to see that these properties are inherited by the relative entropy functional. Note that $h(a) = 0$ if and only if $a = 1$, which implies that

$$H(\mu|\nu) = 0 \iff \mu = \nu.$$

Proposition 1.1.2 (Variational formula) *Let ν be a probability measure; for any $\mu \in \mathcal{P}(\mathcal{X})$ such that $\mu \ll \nu$, the relative entropy can be written as,*

$$H(\mu|\nu) = \sup \left\{ \int u \, d\mu - \log \int e^u \, d\nu; u \text{ meas.}, \int e^u \, d\mu < \infty, \int |u| \, d\nu < \infty \right\}. \quad (1.2)$$

Proof

◁ We recover here the proof of [GL10, Prop. B1]. It is based on the Fenchel inequality $st \leq t \log t - t + e^s$, valid for all $t > 0$, and $s \in \mathbb{R}$. It can be easily checked that the equality is attained when

$$st = t \log t - t + e^s \iff t = e^s. \quad (1.3)$$

Take this inequality with $t = d\mu/d\nu$ and $s = u(x)$,

$$u \frac{d\mu}{d\nu} \leq \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} - \frac{d\mu}{d\nu} + e^u$$

then integrate it with respect to ν ,

$$\int u \, d\mu \leq H(d\mu/d\nu) - 1 + \int e^u \, d\nu.$$

It makes sense under the hypothesis that $\int e^u \, d\nu < \infty$ and $\int |u| \, d\mu < \infty$. Taking the sup among all the measurable function u satisfying $\int e^u \, d\nu < \infty$ and $\int |u| \, d\mu < \infty$, we obtain

$$H(\mu|\nu) = \sup \left\{ \int u \, d\mu - \int (e^u - 1) \, d\nu; u \text{ meas.}, \int e^u \, d\mu < \infty, \int |u| \, d\nu < \infty \right\}.$$

Observe that the equality is given by the monotone convergence theorem. Indeed, easy computations show that the case of equality, $e^u = d\mu/d\nu$, is the limit of the sequence $u_n = \max\{\log(d\mu/d\nu), u^{-n}\}$. Now, replace u by $u + b$ for some $b \in \mathbb{R}$,

$$\int (u + b) \, d\nu - \int e^{(u+b)} - 1 \, d\mu = \int u \, d\nu - e^b \int e^u \, d\mu + b + 1.$$

The sup with respect to $b \in \mathbb{R}$ is attained by $b = -\log \int e^u \, d\mu$, and it gives

$$\sup_{b \in \mathbb{R}} \left\{ \int u \, d\nu - e^b \int e^u \, d\mu + b + 1 \right\} = \int u \, d\nu - \log \int e^u \, d\mu.$$

Finally,

$$\begin{aligned} H(\mu|\nu) &= \sup \left\{ \int u \, d\mu - \int (e^u - 1) \, d\nu; u \text{ meas.}, \int e^u \, d\mu < \infty, \int |u| \, d\nu < \infty \right\} \\ &= \sup \left\{ \int u + b \, d\mu - \int (e^{u+b} - 1) \, d\nu; b \in \mathbb{R}, u \text{ meas.}, \int e^u \, d\mu < \infty, \int |u| \, d\nu < \infty \right\} \\ &= \sup \left\{ \int u \, d\nu - \log \int e^u \, d\mu; u \text{ meas.}, \int e^u \, d\mu < \infty, \int |u| \, d\nu < \infty \right\}. \end{aligned}$$

This concludes the proof. ▷

Note that it follows from (1.2) by standard approximation arguments that

$$H(\mu|\nu) = \sup \left\{ \int u \, d\mu - \log \int e^u \, d\nu; u \in C_b(\mathcal{X}) \right\}.$$

Relative entropy with respect to unbounded measures

We introduce here a generalization of the relative entropy to unbounded measures, which will be necessary later in the manuscript. In order (1.1) to be meaningful when ν has infinite mass, we need some restriction on the set of definition of probability measures. Let \mathcal{X} be a Polish space, and $\nu \in \mathcal{M}_+(\mathcal{X})$ a σ -finite measure. The σ -finiteness assures the existence of a function $w : \mathcal{X} \mapsto [0, \infty)$, such that

$$z_w := \int_{\mathcal{X}} e^{-w} d\nu < \infty. \quad (1.4)$$

Hence, we can define a probability measure associated to ν , by

$$\nu_w := \frac{e^{-w}}{z_w} \nu \in \mathcal{P}(\mathcal{X})$$

therefore, $\log(d\mu/d\nu) = \log(d\mu/d\nu_w) - \log z_w - w$, so that we can write the relative entropy w.r.t. ν via the relative entropy w.r.t. a probability measure as follows,

$$H(\mu|\nu) = H(\mu|\nu_w) - \log z_w - \int w d\mu. \quad (1.5)$$

This formula is well defined for all $\mu \in \mathcal{P}_w(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \int w d\mu < \infty\}$.

Definition 1.1.3 *Let $\nu \in \mathcal{M}_+(\mathcal{X})$ be a σ -finite measure. Then, for all probability measure $\mu \in \mathcal{P}_w(\mathcal{X})$ the relative entropy is defined as*

$$H(\mu|\nu) = \begin{cases} \int_{\mathcal{X}} \log \frac{d\mu}{d\nu} d\mu & \text{if } \mu \ll \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition 1.1.4 *The relative entropy defined at Definition 1.1.3 satisfies the following properties:*

- (i) *Variational formulation, $H(\mu|\nu) = \sup \{ \int u d\mu - \log \int e^u d\nu; u \in C_b(\mathcal{X}) \}$;*
- (ii) *$H(\cdot|\nu)$ is convex on the space of all signed measures $\mathcal{M}(\Omega)$;*
- (iii) *If \mathcal{X} is a topological space, then $H(\cdot|\nu)$ is lower semi-continuous with respect to the topology $\sigma(\mathcal{M}_w(\mathcal{X}), C_w(\mathcal{X}))$, where $\mathcal{M}_w(\mathcal{X})$ is the space of all signed measures μ on \mathcal{X} such that $\int_{\mathcal{X}} w d|\mu| < \infty$*

Proof

\triangleleft (i) The proof is the same as in the classic case of Proposition 1.1.2.

(ii) + (iii) By (i) we have that $H(\cdot|\nu)$ is the supremum of affine continuous functions $\mu \mapsto \int u d\mu - \log \int e^u d\nu$. Therefore it is convex and lower semi-continuous [Roc97]. \triangleright

A last important property that we mention in this section about relative entropy, is the additive property.

Let \mathcal{X}, \mathcal{Y} be two Polish spaces equipped with their Borel σ -fields. Then, for any measurable function $\phi : \mathcal{X} \mapsto \mathcal{Y}$, and any measure $\mu \in \mathcal{M}_+(\mathcal{X})$ we have the disintegration formula,

$$\mu(\cdot) = \int_{\mathcal{Y}} \mu(\cdot|\phi = \tau) \nu_{\phi}(d\tau)$$

where $\mu_\phi = \phi_\# \mu$. We recall the *push-forward* μ_ϕ of a measure μ on a space \mathcal{X} through a map $\phi : \mathcal{X} \rightarrow \mathcal{Y}$, is a measure on \mathcal{Y} defined for all regular enough function $h : \mathcal{Y} \rightarrow \mathbb{R}$ as

$$\int_{\mathcal{Y}} h d\mu_\phi = \int_{\mathcal{X}} h(\phi) d\mu.$$

Proposition 1.1.5 *The following additive property for the relative entropy is satisfied,*

$$H(\mu|\nu) = H(\mu_\phi|\nu_\phi) + \int_{\mathcal{X}} H(\mu(\cdot|\phi = \tau)|\nu(\cdot|\phi = \tau))\mu_\phi(d\tau). \quad (1.6)$$

Proof

◁ The proof is a consequence of [Léo14a, Thm. 1.6] where it is proved an extension of the conditional expectation to unbounded measures. In particular, it allows us to write,

$$\begin{aligned} H(\mu|\nu) &= \int_{\mathcal{Y}} E_\mu \left[\log \left(\frac{d\mu}{d\nu} \right) \middle| \phi = \tau \right] \nu_\phi(d\tau) \\ &= \int_{\mathcal{Y}} \log \frac{d\mu_\phi}{d\nu_\phi} \mu_\phi(\tau) + \int_{\mathcal{Y}} \left[\int_{\mathcal{X}} \log \frac{d\mu(\cdot|\phi = \tau)}{d\nu(\cdot|\phi = \tau)} \mu(d\tau|\phi = \tau) \right] \mu_\phi(d\tau) \\ &= H(\mu_\phi|\nu_\phi) + \int_{\mathcal{X}} H(\mu(\cdot|\phi = \tau)|\nu(\cdot|\phi = \tau))\mu_\phi(d\tau) \end{aligned}$$

that is the announced result. ▷

1.1.2 Kinematics of diffusion processes

In this section we investigate some kinematic property of diffusion stochastic processes. We consider the stochastic differential equation (SDE)

$$\begin{aligned} dX_t &= b(X_t) dt + \sigma(X_t) dW_t, \\ X_0 &= x, \end{aligned} \quad (1.7)$$

where $(W_t)_{t \geq 0}$ is the standard Brownian motion on \mathbb{R}^n , the *diffusion coefficient* σ is a $n \times n$ matrix and the *drift* b is a vector on \mathbb{R}^n . We assume the coefficients b, σ to be smooth functions. In the next theorem we recall some sufficient conditions for the existence of a unique solution of (1.7), see for instance [IW89].

Theorem 1.1.6 *Let b and σ have linear growth, that is,*

$$\|\sigma(x)\| + |b(x)| \leq D(1 + |x|)$$

for some $D \in \mathbb{R}$ and for all $x \in \mathbb{R}^n$, and assume that b, σ are locally Lipschitz, in the sense that for any compact $K \subset \mathbb{R}^n$, there exists a constant $C_K > 0$ such that,

$$\|\sigma(x) - \sigma(y)\| + |b(x) - b(y)| \leq C_K |x - y|$$

for all $x, y \in K$. Then there exists a unique solution $(X_t^x)_{t \geq 0}$ of (1.7) defined on $[0, \infty)$.

The solution of the SDE (1.7) defines a Markov diffusion process. We recall that given a filtration $(\mathcal{F}_t)_{t \in I}$, a process satisfies the Markov property if for all $s, t \geq 0$,

$$P(X_{t+s} \in A | \mathcal{F}_t) = P(X_{t+s} \in A | X_t), \quad \text{a.s.}$$

for any $A \in \mathcal{B}$. Or equivalently, $E(f(X_{t+s}) | \mathcal{F}_t) = E(f(X_{t+s}) | X_t)$ for any suitable function f . In other words we say that a process is Markov if for all time t the past and the future are conditionally independent given the present,

$$P(X_{[0,t]} \in A, X_{[t,1]} \in B | X_t) = P(X_{[0,t]} \in A | X_t)P(X_{[t,1]} \in B | X_t).$$

In the subsequent chapters we will deal only with Markov processes. On the other hand we recall that a diffusion process, is a process driven by a second order differential operator of the form,

$$Lf(x) = \sum_{i,j=1}^n a_{ij}(x) \partial_{ij}^2 f(x) + \sum_{i=1}^n b_i(x) \partial_i f(x). \quad (1.8)$$

We can verify that this is the case for the solution of (1.7). Indeed, for every real smooth function on \mathbb{R}^n , Itô's formula provides,

$$df(X_t) = \nabla f(X_t) \cdot dX_t + \frac{1}{2} dX_t \cdot \nabla^2 f(X_t) dX_t.$$

Let us compute the terms $dX_t^i dX_t^j$,

$$\begin{aligned} dX_t^i dX_t^j &= (b^i dt + \sigma^i(x) dW_t^i) \cdot (b^j dt + \sigma^j dW_t^j) \\ &= b^i b^j dt^2 - b^i \sigma^j dt dW_t^j - b^j \sigma^i dt dW_t^i + \sigma^i dW_t^i \sigma^j dW_t^j. \end{aligned}$$

By definition of Brownian motion we have

$$E(dW_t^i dW_t^j) = \begin{cases} 0 & \text{if } i \neq j \\ dt & \text{if } i = j \end{cases}$$

and we can neglect the higher order terms $dt dW_t$ and dt^2 , hence by (1.7),

$$df(X_t) = \nabla f(X_t) \cdot (b(X_t) dt + \sigma(X_t) dW_t) + \frac{1}{2} \Delta f(X_t) a(X_t) dt$$

where $a = \sigma \sigma^*$. After collecting the terms and integrating with respect to t , it yields to

$$f(X_t^x) = M_t^f + \int_0^t Lf(X_s^x) ds, \quad \forall t \geq 0, x \in \mathbb{R}^n. \quad (1.9)$$

where L is of the type (1.8), and $(M_t^f)_{t \geq 0}$ is a local martingale (for basic definition and properties of martingales see for instance the reference book [Doo53]). In this case the law of the process $(X_t^x)_{t \geq 0}$ is said to solve the *martingale problem* associated with the operator L .

Path measures

In the subsequent chapters we will refer to the solution of the SDE (1.7) through the notion of path measure, that we shall introduce now.

We denote the space of continuous paths on \mathbb{R}^n as

$$\Omega := C([0, 1], \mathbb{R}^n). \quad (1.10)$$

The canonical process $X = (X_t)_{0 \leq t \leq 1}$ is defined for any $0 \leq t \leq 1$ by,

$$X_t(\omega) = \omega_t \in \mathbb{R}^n, \quad \forall \omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega.$$

For any time t , $X_t : \Omega \rightarrow \mathbb{R}^n$ represents the position at time t of the trajectory $\omega \in \Omega$. We equip Ω with the canonical σ -field generated by $(X_t; 0 \leq t \leq 1)$, denoted by $\sigma(X_{[0,1]})$. Any probability measure on Ω is called a *path measure*.

For any path measure $P \in \mathcal{P}(\Omega)$ and any time $0 \leq t \leq 1$, the probability measure

$$P_t := (X_t)_\# P = P(X_t \in \cdot) \in \mathcal{P}(\mathbb{R}^n)$$

is the law of the position of a particle at time t when its whole random history is described by P . It is called the *t-th marginal* of P .

In the following we will say that $P \in \mathcal{P}(\Omega)$ is associated to the SDE (1.7) or is the law of the SDE (1.7), or even P is associated with the operator L , in the sense that P is solution of the martingale problem associated to L . This means that for every smooth function f and for every $t \geq 0$,

$$f(X_t) - f(X_0) - \int_0^t Lf(X_s) ds = M_t^f$$

is a local- P martingale, that is the same of saying that

$$\lim_{h \rightarrow 0^+} \frac{1}{h} E_P(f(X_{t+h}) - f(X_t)) | X_t = z = Lf(z).$$

In the following chapters we will also use the notation

$$dX_t = b(X_t) dt + \sigma dW_t^P$$

where dW_t^P denote a P -martingale increment ($E_P(dW_t^P | X_t) = 0$) and a P -Brownian increment (with quadratic variation $[W_t^P] := \int_0^t a_s ds$, with $a_s = \sigma_s \sigma_s^*$ P -a.e.).

Given the time-reversal mapping

$$X_t^* := X_{1-t}, \quad 0 \leq t \leq 1, \quad (1.11)$$

we introduce the next important properties for path measures.

Definition 1.1.7 (Reversible and invariant measure) We say that a path measure $R \in \mathcal{P}(\Omega)$ is reversible with reversing measure $m \in \mathcal{M}_+(X)$ if $R_0 = m$ and for all $0 \leq s \leq t \leq 1$, the time reversed $R^* := (X^*)_\# R$ satisfies,

$$R_{[s,t]}^* = R_{[s,t]},$$

that is equivalent to say that R is invariant with respect to the inversion of time.

We say that R is stationary if

$$(X_t)_\# R = m \in \mathcal{P}(\mathbb{R}^n), \quad \forall t \geq 0.$$

Clearly reversibility is stronger than stationarity. Indeed it means that $R_s = R_t$ for all $0 \leq s \leq t \leq 1$ hence $R_t = R_0 = m$ for all $t \in [0, 1]$.

In the following we will mainly focus on the case of (1.7) with a constant diffusion coefficient, that is

$$dX_t = b(X_t) dt + \sigma dW_t, \quad \forall t \geq 0 \quad (1.12)$$

for some $\sigma > 0$.

Examples 1.1.8 *Let us mention some fundamental examples that will occur often in this manuscript.*

1. *The easiest case is the one with no drift, i.e. $b = 0$. In this case $dX_t = \sigma dW_t$, is the Brownian motion and the operator in (1.8) is*

$$Lf = \sigma^2 \Delta f / 2.$$

It is an easy exercise to prove that the reversing measure is $m = \mathcal{L}$, where \mathcal{L} is the Lebesgue measure on \mathbb{R}^n . Note that m is not a probability measure since it has infinite mass on \mathbb{R}^n .

2. *Another typical example is when the drift is $b(x) = -x$. The solution is called Ornstein-Uhlenbeck process, the operator L is given by,*

$$Lf = (\sigma^2 \Delta f - x \cdot \nabla f) / 2.$$

and the reversing measure is the Gaussian measure on \mathbb{R}^n , $dm = e^{-|x|^2/2} / (2\pi)^{n/2} dx$.

3. *In general we will consider the Kolmogorov process, solution of the SDE (1.12) with a drift of the type $b = -\nabla V / 2$, for some regular enough potential V . Then the operator L is*

$$Lf = (\sigma^2 \Delta - \nabla V \cdot \nabla f) / 2,$$

and the reversing measure is $\mathbf{m} = e^{-V} dx$. Typically we will assume V to be C^2 , such that $\int e^{-V} dx = 1$ and κ -convex, that is $\nabla^2 V \geq \kappa \text{Id}$ for some real κ .

Nelson's velocities

We start by recalling the well known link between the SDE (1.12) and the related partial differential equations.

Let $\mu(x, t)$ be the probability density of the diffusion process which satisfies (1.12) so that for any measurable set $B \subset \mathbb{R}^n$,

$$P(X_t \in B) = \int_B \mu(x, t) dx. \quad (1.13)$$

Then it follows by Itô's formula that for any regular and integrable enough test function u and any $t \geq 0$,

$$\begin{aligned} \int_{\mathbb{R}^n} u(z) \mu_t(z) dz &= E_P u(X_0) + E_P \int_0^t (b \cdot \nabla u + \sigma^2 \Delta u / 2)(X_s) ds \\ &= \int_{\mathbb{R}^n} u(z) \mu_0(z) dz + \int_0^t ds \int_{\mathbb{R}^n} (b \cdot \nabla u + \sigma^2 \Delta u / 2)(z) \mu_s(z) dz \end{aligned}$$

and after integration by parts, we obtain

$$\partial_t \mu_t + \nabla \cdot (b \mu_t) - \frac{\sigma^2}{2} \Delta \mu_t = 0. \quad (1.14)$$

that is the Fokker Plank equation [Doo53, Ch. 6].

In order to go further in the study of diffusion kinematics, let us introduce the notions of backward and forward mean velocities, and current and osmotic velocities that we will refer to with the name of Nelson's velocities. We refer to [Nel67] and [Nel88] for a detailed discussion on these notions.

Definition 1.1.9 (Nelson's velocities) *Given a Markov diffusion process $(X_t)_{t \geq 0}$ solution of the SDE (1.12), and let $P \in \mathcal{P}(\Omega)$ be the path measure associated to it. We define for $dtdP_t(dz)$ -almost all $(t, z) \in [0, 1] \times \mathbb{R}^n$, when these quantities are meaningful,*

(i) *the forward velocity as,*

$$\vec{v}_t^P(z) := \lim_{h \rightarrow 0^+} \frac{1}{h} E \left(X_{t+h} - X_t \mid X_t = z \right);$$

(ii) *the backward velocity as*

$$\overleftarrow{v}_t^P(z) := \lim_{h \rightarrow 0^+} \frac{1}{h} E \left(X_{t-h} - X_t \mid X_t = z \right);$$

(iii) *the current and osmotic velocities respectively by*

$$v_t^{cu,P}(z) := \lim_{h \rightarrow 0^+} \frac{1}{2h} E_P(X_{t+h} - X_{t-h} \mid X_t = z) \quad (1.15)$$

$$v_t^{os,P}(z) := \lim_{h \rightarrow 0^+} \frac{1}{2h} E_P(X_{t+h} - 2X_t + X_{t-h} \mid X_t = z). \quad (1.16)$$

By the definition of current and osmotic velocities, it is clear that they are given respectively by the half difference and the average of the forward and backward velocities, that is,

$$v_t^{cu,P} = \frac{\vec{v}_t^P - \overleftarrow{v}_t^P}{2}, \quad v_t^{os,P} = \frac{\vec{v}_t^P + \overleftarrow{v}_t^P}{2}. \quad (1.17)$$

Moreover we see that

$$\vec{v}_t^P = v^{cu,P} + v^{os,P}, \quad \overleftarrow{v}_t^P = -v^{cu,P} + v^{os,P}. \quad (1.18)$$

It is an immediate consequence of relations (1.17), that the Nelson velocities satisfy the parallelogram identity,

$$2(|v^{cu}|^2 + |v^{os}|^2) = |\vec{v}|^2 + |\overleftarrow{v}|^2. \quad (1.19)$$

Remark 1.1.10 *These notions are defined for any $P \in \mathcal{P}(\Omega)$. However, in the general case where P might not be Markov, the forward and backward velocities of P are defined by*

$$\begin{aligned} \vec{v}^P(t, X_{[0,t]}) &:= \lim_{h \rightarrow 0^+} \frac{1}{h} E_P \left(X_{t+h} - X_t \mid X_{[0,t]} \right) \\ \overleftarrow{v}^P(t, X_{[t,1]}) &:= \lim_{h \rightarrow 0^+} \frac{1}{h} E_P \left(X_{t-h} - X_t \mid X_{[t,1]} \right). \end{aligned}$$

In this case the current and osmotic velocities depend on the whole past and future history of the path and it is not true in general that they can be expressed by formulas of the type (1.15), (1.16).

Remark 1.1.11 Note that we defined the backward velocity with opposite sign with respect to the original definition of Nelson; this is actually a more natural choice when dealing with reversed time.

Time reversed

Since we introduced two time directions notions of velocities, it is interesting to see which is the relation with the Nelson velocities of the time reversed process.

Let us consider the time reversed $P^* = (X^*)_{\#}P$, where (X^*) is the time reversal map introduced at (1.11). By Definition 1.1.9, the backward velocity is given by

$$\begin{aligned}\overleftarrow{v}_t^P &= \lim_{h \rightarrow 0^+} \frac{1}{h} E_P(X_{t-h} - X_t | X_t) \\ &= \lim_{h \rightarrow 0^+} \frac{1}{h} E_{P^*}(X_{1-t+h} - X_{1-t} | X_{1-t}) \circ X^* = \overrightarrow{v}_{1-t}^* \end{aligned} \quad (1.20)$$

where the lighter index $*$ stands for P^* . The same relation holds for the forward velocity of P , hence we can deduce that,

$$v_t^{cu,*} = \frac{\overrightarrow{v}_t^* - \overleftarrow{v}_t^*}{2} = \frac{\overleftarrow{v}_{1-t}^P - \overrightarrow{v}_{1-t}^P}{2} = -v_{1-t}^{cu,P}$$

and analogously,

$$v_t^{os,*} = \frac{\overrightarrow{v}_t^* + \overleftarrow{v}_t^*}{2} = \frac{\overleftarrow{v}_{1-t}^P + \overrightarrow{v}_{1-t}^P}{2} = v_{1-t}^{os,P},$$

for $dtdP_t(z)$ -almost all $(t, z) \in [0, 1] \times \mathbb{R}^n$.

Examples 1.1.12 (First part) Let us identify the Nelson velocities in our typical examples.

1. In the deterministic case, i.e. $\sigma = 0$ and $X_t = x_t$ satisfying the ordinary differential equation $x_t = x_0 + \int_0^t b(x_s) ds$, it is straightforward to see that

$$\overrightarrow{v}_t = v_t^{cu} = b(z), \quad \overleftarrow{v}_t = -b(z) \quad v_t^{os} = 0.$$

2. For the general Markov diffusion process (1.12) the forward velocity is $\overrightarrow{v}_t(z) = b(z)$, indeed,

$$\begin{aligned} \frac{1}{h} E(X_{t+h} - X_t | X_t = z) &= \frac{1}{h} E \left(\int_t^{t+h} b(X_s) ds | X_t = z \right) + \sqrt{\sigma} E(W_{t+h} - W_t | X_t = z) \\ &= \frac{1}{h} E \left(\int_t^{t+h} b(X_s) ds | X_t = z \right) \end{aligned}$$

where the last equality derives by the definition of Brownian motion; then taking the limit, we have

$$\overrightarrow{v}_t(z) = \lim_{h \rightarrow 0^+} \frac{1}{h} E(X_{t+h} - X_t | X_t = z) = b(z). \quad (1.21)$$

We want to see now the relations between Nelson's velocities and the density $\mu_t = dP_t/dz$ introduced at (1.13). At the light of (1.21), the Fokker-Plank equation derived at (1.14) can be written in terms of forward velocity as,

$$\partial_t \mu_t + \nabla \cdot (\vec{v}_t^P \mu_t) - \frac{\sigma^2}{2} \Delta \mu_t = 0. \quad (1.22)$$

Writing the Fokker-Plank equation for the time reversed process P^* ,

$$\partial_t \mu_t^* + \nabla \cdot (\vec{v}_t^* \mu_t^*) - \frac{\sigma^2}{2} \Delta \mu_t^* = 0$$

where we denoted $\mu_t^* := dP_t^*/dz = dP_{1-t}/dz = \mu_{1-t}$. Hence $\partial_t \mu_{1-t}^* = -\partial_t \mu_t$, and together with (1.20) the last equation can be written as,

$$\partial_t \mu_t - \nabla \cdot (\overleftarrow{v}_t^P \mu_t) + \frac{\sigma^2}{2} \Delta \mu_t = 0 \quad (1.23)$$

Taking the half sum of (1.22) and (1.23), we have

$$\partial_t \mu_t + \nabla \cdot \left(\frac{\vec{v}_t^P - \overleftarrow{v}_t^P}{2} \mu_t \right) = 0$$

that gives, by (1.17), the *current equation*

$$\partial_t \mu_t + \nabla \cdot (v_t^{cu,P} \mu_t) = 0. \quad (1.24)$$

On the other hand, by taking the difference of (1.22) and (1.23) we obtain,

$$\nabla \cdot [\mu_t (\vec{v}_t^P + \overleftarrow{v}_t^P - \sigma^2 \nabla \log \mu_t)] = 0$$

it means by (1.17) that the quantity $v_t^{os,P} - \sigma^2 \nabla \log \sqrt{\mu_t}$ is divergence free. It can be proved that this quantity is actually zero via a duality formula (appearing in [Nag64], [Nel67, Ch. 13] and [Föl86]) that is,

$$v_t^{os,P} = \frac{\sigma^2}{2} \nabla \log \mu_t = \sigma^2 \nabla \log \sqrt{\mu_t}. \quad (1.25)$$

Examples 1.1.13 (Second part) *Going back to the Examples 1.1.12, we study the case when $b = -\sigma^2 \nabla V/2$ for some C^2 real function V , where σ is the diffusion coefficient. As pointed out in the Examples 1.1.8, (3), the reversing measure is $\mathbf{m} = e^{-V} dz$. Hence, $\mu_t = dP_t/dz = e^{-V}$. According to (1.25), we have,*

$$v_t^{os,P} = \frac{\sigma^2}{2} \nabla \log(e^{-V}) = -\frac{\sigma^2}{2} \nabla V.$$

Moreover, by (1.21) the forward velocity is

$$\vec{v}_t^P = -\frac{\sigma^2}{2} \nabla V$$

and together with the relations (1.17) and (1.18) we can conclude that,

$$v_t^{cu,P} = 0, \quad \overleftarrow{v}_t^P = -\frac{\sigma^2}{2} \nabla V.$$

Summerizing, when $P \in \mathcal{P}(\Omega)$ is associated to the SDE

$$dX_t = -\frac{\sigma^2}{2} \nabla V dt + \sigma dW_t$$

the Nelson velocities are,

$$\vec{v}_t^P = \overleftarrow{v}_t^P = v_t^{os,P} = -\frac{\sigma^2}{2} \nabla V, \quad v_t^{cu,P} = 0.$$

Physical interpretation

From a physical point of view, the diffusion process (1.12) describes the trajectory of many particles suspended in a fluid, under the action of an external (deterministic) force. Think for instance to a ink drop released in water, or to some perfume quickly spreading in the air. The diffusion, that describes the motion of the particles from a region of high concentration to a region of low concentration, is the result of random displacements due to collisions between molecules in the suspension, namely Brownian motions.

Following [Nel67, Ch. 4], at the equilibrium the external force K is balanced by the *osmotic pressure* that is the force due to collisions,

$$K = kT \frac{\nabla \rho}{\rho}$$

where ρ is the probability density (the number of particles per unit of volume divided by the total number of particles), k is the Boltzmann constant and T is the absolute temperature. Without entering in the details, if we divide both sides by $m\beta$ where m is the mass of a particle and β is some constant with the dimension of sec^{-1} , encoding the friction experienced by the particle moving in the fluid, we obtain,

$$\frac{K}{m\beta} = D \frac{\nabla \rho}{\rho}$$

where $D := kT/m\beta$ is the diffusion coefficient derived by Einstein in [Ein05]. Thus, a dimensional analysis shows us that the left hand side is a velocity, and is due to the external force, whereas the right hand side is the velocity due to the osmotic pressure (diffusion) and coincides with (1.25).

On the other hand the meaning of the current velocity is easily derived by the continuity equation (1.24). It consists in the velocity field driving the flux, hence the movement of particles (or in other words the *current*) according to equation (1.24) (see for instance the textbook [Eva98] for a general discussion on continuity equations).

1.1.3 Entropy of diffusion processes

In this section we focus on the Girsanov Theory. It will play a crucial role when we will derive a dynamic Benamou-Brenier formulation for the entropic cost at Section 3.5 and in the application to the concentration and Prékopa-Leindler inequality at Chapter 6.

Let $R \in \mathcal{P}(\Omega)$ be associated to the SDE (1.12), with $\sigma > 0$ and the vector field $b : [0, 1] \times \Omega \rightarrow \mathbb{R}^n$ locally Lipschitz.

Theorem 1.1.14 (Girsanov's theorem) *Let $P \in \mathcal{P}(\Omega)$ be such that $H(P|R) < \infty$. Then there exists a vector field $\beta : [0, 1] \times \Omega \rightarrow \mathbb{R}^n$ defined P -a.e. such that,*

1.

$$\sigma^2 E_P \int_0^1 |\beta_t|^2 dt < \infty;$$

2. P is associated to the SDE

$$dX_t = [b_t(X_t) + \sigma^2 \beta_t(X_t)] dt + \sigma dW_t^P.$$

Assume furthermore that for all $x \in \mathbb{R}^n$, R^x is the only weak solution of (1.12) with initial law δ_x , then

3. The density of P with respect to R is,

$$\frac{dP}{dR} = \frac{dP_0}{dR_0}(X_0) \exp \left\{ \sigma \int_0^1 \beta_t \cdot dW_t^R - \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right\}; \quad (1.26)$$

4. The relative entropy of P with respect to R is given by,

$$H(P|R) = H(P_0|R_0) + \frac{\sigma^2}{2} E_P \int_0^1 |\beta_t|^2 dt.$$

Remark 1.1.15 Note that since σ is a positive and finite constant, it does not play any role in 1. However we chose to include it, to underline the fact that the diffusion coefficient does have a role. Indeed, whenever $\sigma(X_t)$ is not constant, then $M_t^R = \int_0^t \sigma_s dW_s^R$, with quadratic variation,

$$[M^R, M^R]_t = \int_0^t a_s ds \quad R\text{-a.e.}$$

where $a_t := \sigma_t \sigma_t^*(X_t)$. In this situation, 1. becomes,

$$E_P \int_0^1 \beta_t \cdot a_t \beta_t dt < \infty.$$

Sketch of the Proof

◁ We only give a sketch of the proof and refer to [Léo12b, Thm. 2.1, Thm. 2.4] for a detailed version.

For any measure $Q \in \mathcal{P}(\Omega)$ we define,

$$\mathcal{H}(Q) := \left\{ g : [0, 1] \times \Omega \rightarrow \mathbb{R}^n; \text{ meas. and s.t. } \sigma^2 E_Q \int_0^1 |g_t|^2 dt < \infty \right\}$$

equipped with the semi-norm $\|g\|_{\mathcal{H}(Q)} := \left(\sigma^2 E_Q \int_0^1 |g_t|^2 dt \right)^{1/2}$. The starting point is the variational representation of the relative entropy (1.2) which implies,

$$H(P|R) \geq \int u dP - \log \int e^u dR. \quad (1.27)$$

We can choose $u = \int_0^1 \sigma h_s dW_s^R - \frac{\sigma^2}{2} \int_0^1 |h_s|^2 ds$. For such a function u , [Léo12b, Lemma 4.1] assures that the logarithmic term in (1.27) is non-positive, therefore we can neglect it and we obtain,

$$H(P|R) \geq E_P \sigma \int h_s dW_s^R - \frac{\sigma^2}{2} E_P \int_0^1 |h_s|^2 ds \quad (1.28)$$

We can do the same for $-h$ and replace h with h/λ for some $\lambda > 0$, to get

$$\left| E_P \frac{\sigma}{\lambda} \int_0^1 h_s \cdot dW_s^R \right| \leq H(P|R) + \frac{\sigma^2}{2\lambda^2} E_P \int_0^1 |h_s|^2 ds.$$

If we choose $\lambda = \|h\|_{\mathcal{H}(P)}$, it reduces to

$$\left| \sigma E_P \int_0^1 h_s \cdot dW_s^R \right| \leq \left(H(P|R) + \frac{1}{2} \right) \|h\|_{\mathcal{H}(P)}$$

This holds even when $\|h\|_{\mathcal{H}^P} = 0$. Indeed, when $\|h\|_{\mathcal{H}^P} = 0$ replacing h by ah in (1.28) gives

$$E_P \sigma \int h_s \cdot dW_s^R \leq H(P|R)/a$$

and at the limit $a \rightarrow \infty$ it gives $E_P \sqrt{\sigma} \int h_s \cdot dW_s^R = 0$. Identifying the equivalence classes turns $(\mathcal{H}^P, \|\cdot\|_P)$ into an Hilbert space. Hence, the Riesz representation Theorem assures, as soon as $H(P|R)$ is finite, the existence of a unique $\beta \in \mathcal{H}^P$ such that,

$$\sigma E_P \int h \cdot dW_s^R = \sigma E_P \int_0^1 \beta_t \cdot h_t dt.$$

This is equivalent to,

$$E_P \int_0^1 h_t \cdot (dM_t^R - \sigma^2 \beta_t dt) = 0, \quad \forall h \in \mathcal{H}_P.$$

And this implies that $dM_t^P := dM_t^R - \sigma^2 \beta_t dt = dX_t - b_t(X_t)dt - \sigma^2 \beta_t dt$ is a P -martingale increment. In other words, P is associated to the SDE,

$$dX_t = [b_t(X_t) + \sigma^2 \beta_t(X_t)] dt + \sigma dW_t^P$$

that is the statement in 2..

In order to prove 3. we take any $P \in \mathcal{P}(\Omega)$ that satisfies

$$dX_t = b_t^P dt + dM_t^P,$$

for some b^P such that

$$E_P \int_0^1 |b_t^P|^2 dt < \infty. \quad (1.29)$$

We define for some γ satisfying (1.29),

$$Q := \exp \left(\int_0^1 \gamma_s \cdot dW_s^P - \frac{\sigma^2}{2} \int_0^1 |\gamma_s|^2 ds \right) P.$$

It can be proved that Q satisfies

$$dX_t = [b_t^P + \sigma \gamma_t] dt + dM_t^Q, \quad Q\text{-a.e.}$$

with $dM_t^Q = \sigma dW_t^Q$. Now, if we recall that R is associated with (1.12) and take $P \in \mathcal{P}(\Omega)$ such that $H(P|R) < \infty$, then if we set $\gamma = -\beta$ that is the vector field that appears in 1. and 2., we obtain

$$\begin{aligned} dX_t &= [b_t^P - \sigma \beta_t] dt + dM_t^Q \\ &= [b_t + \sigma \beta_t - \sigma \beta_t] dt + dM_t^Q \\ &= b_t dt + dM_t^Q. \end{aligned}$$

Finally by the assumption of uniqueness on R , we obtained $Q^x = R^x$, that implies

$$\begin{aligned} P^x &= \exp \left(\int_0^1 \beta_t \cdot dM_t^P + \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right) R^x \\ &= \exp \left(\int_0^1 \beta_t \cdot (dX_t - b_t dt - \sigma^2 \beta_t dt) + \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right) R^x \\ &= \exp \left(\int_0^1 \beta_t \cdot (dX_t - b_t dt) - \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right) R^x \\ &= \exp \left(\int_0^1 \beta_t \cdot dM_t^R - \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right) R^x \end{aligned}$$

that is 3..

Point 4. is a direct consequence of 3. Indeed by definition of relative entropy and (1.26)

$$\begin{aligned} H(P|R) &= E_P \log(dP/dR) \\ &= E_P \log(dP_0/dR_0)(X_0) + E_P \left(\int_0^1 \beta_t \cdot dM_t^P + \frac{\sigma^2}{2} \int_0^1 |\beta_t|^2 dt \right) \\ &= H(P_0|R_0) + \frac{\sigma^2}{2} E_P \int_0^1 |\beta_t|^2 dt \end{aligned}$$

where the last equality is given by the martingale identity $E_P \int_0^1 \beta_t \cdot dM_t^P = 0$. \triangleright

1.1.4 Markov semigroups and Γ_2 -calculus

We recall here some well known and basic results related to the Γ_2 -calculus and the Bakry-Emery theory. Beyond a wide literature on the topic, the main references are the two monographs [BGL14] and [ABC⁺00] (in french).

Let us start by the definition of Markov semigroup and its relation to Markov processes.

Let $m \in \mathcal{M}_+(\mathbb{R}^n)$ be any σ -finite measure.

Definition 1.1.16 (Markov semigroup) *Let $(T_t)_{t \geq 0}$ be a family of linear bounded operators on $L^2(m)$ it is a Markov semigroup if:*

- (i) $T_0 = Id$,
- (ii) for all $f \in L^2(m)$, $t \rightarrow T_t f$ is continuous on $[0, +\infty)$,
- (iii) for all $s, t \geq 0$, $T_{t+s} = T_t \circ T_s$,
- (iv) $T_t \mathbb{1} = \mathbb{1}$ and $T_t f \geq 0$ for $f \geq 0$,
- (v) T_t is a contraction: $\forall f \in L^2(m)$, $\|T_t f\|_{L^2(m)} \leq \|f\|_{L^2(m)}$.

Given a Markov process $(X_t)_{t > 0}$ and $X_0 = x$, solution of the SDE (1.7) then the associated Markov semigroup is defined for any suitable real measurable function f on \mathbb{R}^n by,

$$T_t f(x) := E(f(X_t) | X_0 = x).$$

On \mathbb{R}^n Markov semigroups admit the kernel representation,

$$T_t f(x) = E(f(X_t) | X_0 = x) = \int f(y) p_t(x, dy).$$

where $p_t(x, dy)$ is the transition probability kernel of the associated Markov process. This is actually true on any space that admits a disintegration formula $\mu(dx dy) = k(x, dy) \mu_1(dx)$, such as a Polish space. Moreover the kernel $p_t(x, dy)$ may have a density with respect to the measure m , in this case we say that a Markov semigroup admits a density kernel with respect to m , if for any $t > 0$ and any $f \in L^2(m)$,

$$T_t f(x) = \int f(y) p_t(x, y) dm(y).$$

The existence of such a kernel is not always guaranteed, see for instance [BGL14, Prop.1.2.5]. For all function f such that $(T_t f - f)/t$ admits the limit for $t \rightarrow 0^+$ in $L^2(m)$ with the $L^2(m)$ norm, we give the following

Definition 1.1.17 (Infinitesimal generator) *The infinitesimal generator is the operator L associated to the semigroup $(T_t)_{t \geq 0}$ defined as*

$$Lf := \lim_{t \rightarrow 0^+} \frac{T_t f - f}{t}.$$

The set of functions f for which the limit exists in $L^2(m)$ with respect to the L^2 -norm, is called domain of L and is denoted $\mathcal{D}(L)$.

The semigroup property and the linearity of the operators $(T_t)_{t \geq 0}$ yield for any function $f \in \mathcal{D}(L)$ and for all $t \geq 0$ to,

$$\partial_t T_t f = T_t Lf = LT_t f$$

It is a delicate question to determine the domain of the generator L and the class of functions on which defining a Markov semigroup. [BGL14, Ch. 3.3] is entirely devoted to rigorously answering this question, and the name of the chapter "Heart of the darkness" is emblematic. We only mention that in order the operator L to be well defined, it is not enough to consider a class of functions smooth enough, but it is necessary a good behavior at infinity in order to use integration by parts.

The infinitesimal generator L coincides with the differential diffusion operator (1.8) associated to the martingale problem (1.9).

Let us now give the important notions of invariant and symmetric measures defined at Definition 1.1.7 in terms of Markov semigroups.

Definition 1.1.18 (Invariant measure) *A σ -finite measure m on \mathbb{R}^n is invariant for a Markov semigroup $(T_t)_{t \geq 0}$, if for all $t \geq 0$ and all bounded measurable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,*

$$\int_{\mathbb{R}^n} T_t f \, dm = \int_{\mathbb{R}^n} f \, dm$$

Not necessary a Markov semigroup admits an invariant measure; moreover it is not necessary a finite measure, and it is defined up to a multiplicative factor. When m is a finite measure we assume that it is a probability and we can see that

$$E(f(X_t)) = E(E(f(X_t)|X_0)) = \int_{\mathbb{R}^n} T_t f \, dm = \int_{\mathbb{R}^n} f \, dm = E(f(X_0)).$$

We can deduce that if the associated Markov process has initial distribution m , namely $\mathcal{L}(X_0) = m$, then $\mathcal{L}(X_t) = m$ for any time $t > 0$. It follows directly from the definition of infinitesimal generator, that Definition 1.1.18 is equivalent to say that for all $f \in L^2(m)$, such that $Lf \in L^1(m)$,

$$\int_{\mathbb{R}^n} Lf \, dm = 0.$$

Another important definition that we will often use in the next chapters, is the one of reversibility.

Definition 1.1.19 (Reversing measure) *A σ -finite measure m on \mathbb{R}^n is reversing for a Markov semigroup $(T_t)_{t \geq 0}$, if for all $f, g \in L^2(m)$ and all $t \geq 0$,*

$$\int_{\mathbb{R}^n} f T_t g \, dm = \int_{\mathbb{R}^n} g T_t f \, dm.$$

Remark that reversibility is a stronger condition and implies invariance. It can be seen by taking in the definition $g = \text{Id}$. It is an easy exercise to prove that these two definitions are equivalent to the one given at Definition 1.1.7. Indeed, we say that a measure is invariant if and only if $R_t = (X_t)_\# R = m$ for all $0 \leq t \leq 1$, hence

$$\int T_t f \, dm = E_R f(X_t) = E_R f(X_0) = \int f \, dm.$$

On the other hand, R is reversible if and only if,

$$\begin{aligned} \int f T_t g \, dm &= E_R(f(X_0)g(X_t)) = E_{R^*}(f(X_0)g(X_t)) = \\ &= E_R(f(X_1)g(X_{1-t})) = E_R(f(X_t)g(X_0)) = \int g T_t f \, dm. \end{aligned}$$

Examples 1.1.20 *Let us look at the main example that we will explore all along this manuscript. The Kolmogorov semigroup $(T_t)_{t \geq 0}$ on \mathbb{R}^n , associated to the random process introduced in Example 1.1.8 (3.), has infinitesimal generator*

$$L = (\Delta - \nabla V \cdot \nabla)/2.$$

and reversing measure $d\mathbf{m} = e^{-V} dx$.

- (a) *The first main example is $V = 0$. In this case the generator is $L = \Delta/2$, the reversing measure is the Lebesgue measure and $(T_t)_{t \geq 0}$ is the heat semigroup*

$$T_t f(x) := \int_{\mathbb{R}^n} f(y) e^{-\frac{|x-y|^2}{2t}} / (2\pi t)^{n/2} \, dy.$$

- (b) *Another explicit example is $V = |x|^2/2$. In this case, $L = (\Delta - x \cdot \nabla)/2$, the reversing measure is the Gaussian measure $d\mathbf{m} = e^{-|x|^2/2}/(2\pi)^{n/2} dx$ and $(T_t)_{t \geq 0}$ is the Ornstein-Uhlenbeck semigroup*

$$T_t f(x) := \int f(e^{-t}x + \sqrt{1-e^{-2t}}y) \frac{e^{-|y|^2/2}}{(2\pi)^{n/2}} \, dy.$$

Note that while the reversing measure of the Ornstein-Uhlenbeck semigroup is a probability measure, the reversing measure of the heat semigroup is not finite on \mathbb{R}^n .

We recall some general definition and property of the operators Γ and Γ_2 associated to a generator L with invariant measure m being fixed.

We assume the existence of an algebra $\mathfrak{A}_0 \subset L^2(m)$ dense in $L^2(m)$, stable under the action of smooth functions $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$, such that $\Phi(0) = 0$ and satisfying $L\mathfrak{A}_0 \subset \mathfrak{A}_0$. See [BGL14, Sec. 3.4.1] for further details on the existence and the properties of the algebra \mathfrak{A}_0 .

Definition 1.1.21 (Carré du champs operators) *For all couples (f, g) of functions in \mathfrak{A}_0 , the carré du champ operator and the iterated carré du champ operator associated to the generator L , are respectively defined as*

$$\begin{aligned} \Gamma(f, g) &:= \frac{1}{2}[L(fg) - fLg - gLf], \\ \Gamma_2(f, g) &:= \frac{1}{2}[L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf)]. \end{aligned}$$

We denote $\Gamma f = \Gamma(f) = \Gamma(f, f)$, and the same for Γ_2 .

We collect in the next proposition some basic properties of Γ and Γ_2 , that will be used over and over in this manuscript.

Proposition 1.1.22 *Let Γ and Γ_2 be associated to an infinitesimal generator L with reversible measure m . For any couple of functions $f, g \in \mathfrak{A}_0$ the following statements are satisfied,*

i. $\Gamma(\cdot, \cdot)$ and $\Gamma_2(\cdot, \cdot)$ are bi-linear and symmetric;

ii. $\Gamma f \geq 0$,

iii. Γ satisfies the integration by parts formula,

$$\int f Lg \, dm = - \int \Gamma(f, g) \, dm;$$

iv. For all real function $\Phi \in C^\infty(\mathbb{R}^n)$ with $\Phi(0) = 0$ and for all functions $f_1, \dots, f_n \in \mathfrak{A}_0$, the diffusion property is satisfied,

$$L\Phi(f_1, \dots, f_n) = \sum_{i=1}^n \frac{\partial \Phi}{\partial x_i}(f_1, \dots, f_n) Lf_i + \sum_{i,j=1}^n \frac{\partial^2 \Phi}{\partial x_i \partial x_j}(f_1, \dots, f_n) \Gamma(f_i, f_j); \quad (1.30)$$

v. For all real function $\Phi \in C^\infty(\mathbb{R}^n)$ with $\Phi(0) = 0$ the chain formula is satisfied,

$$\Gamma(\Phi(f), g) = \Phi'(f) \Gamma(f, g);$$

vi. For any function $h \in \mathfrak{A}_0$,

$$\int \Gamma(f, g) h \, dm = - \int [f \Gamma(g, h) + fh Lg] \, dm;$$

vii. If $f \in \mathfrak{A}_0$ depends also on a time variable $t \geq 0$,

$$\frac{d}{dt} \Gamma(f_t) = 2\Gamma(f_t, f'_t);$$

viii. Γ satisfies for any $\alpha, \beta \in \mathbb{R}$ the identity,

$$\Gamma(\alpha f + \beta g, \alpha f - \beta g) = \alpha^2 \Gamma(f) - \beta^2 \Gamma(g);$$

ix. For any $\alpha, \beta \in \mathbb{R}$, Γ_2 satisfies the parallelogram identity

$$\Gamma_2(\alpha f + \beta g) + \Gamma_2(\alpha f - \beta g) = 2\alpha^2 \Gamma_2(f) + 2\beta^2 \Gamma_2(g);$$

x. Γ_2 satisfies the integration by parts formula,

$$\int_{\mathbb{R}^n} \Gamma_2(f, g) \, dm = \int_{\mathbb{R}^n} (Lf)(Lg) \, dm;$$

xi. Γ satisfies the symmetry property ,

$$\int \Gamma(Lf, g) \, dm = \int \Gamma(f, Lg) \, dm.$$

Note that property (iv) is equivalent to the diffusion formula (1.8). An easy computation shows that L defined at (1.8) satisfies (1.30). On the other hand by choosing $f_i = x_i$ in (1.30) shows that L is actually of the form (1.8). We skip the proof of these properties, that is merely computations. We refer to [BGL14] for a deep and complete presentation on the subject.

Definition 1.1.23 (Bakry-Emery condition) *We say that an operator L satisfies a curvature-dimension inequality $CD(\kappa, N)$ with curvature $\kappa \in \mathbb{R}$ and dimension $N \in [1, \infty]$, if for all function $f \in \mathfrak{A}_0$,*

$$\Gamma_2(f) \geq \kappa \Gamma(f) + \frac{1}{N} (Lf)^2, \quad (1.31)$$

or with infinite dimension $N = \infty$,

$$\Gamma_2(f) \geq \kappa \Gamma(f). \quad (1.32)$$

Remark 1.1.24 *Define for some $\alpha \in \mathbb{R}$ the operator $L^\alpha = \alpha L$. It is an easy exercise to see that the associated carré du champ operators are,*

$$\Gamma^\alpha = \alpha \Gamma, \quad \Gamma_2^\alpha = \alpha^2 \Gamma_2,$$

where Γ and Γ_2 are associated to the generator L . Then if L satisfies a $CD(\kappa, N)$ condition, we have

$$\begin{aligned} \Gamma_2^\alpha(f) = \alpha^2 \Gamma_2(f) &\geq \alpha^2 \kappa \Gamma(f) + \frac{\alpha^2}{N} (Lf)^2 \\ &= \alpha \kappa \Gamma^\alpha(f) + \frac{1}{N} (L^\alpha f)^2 \end{aligned}$$

that is, L^α satisfies a $CD(\alpha\kappa, N)$ condition. Note that unlike the typical $L = \Delta - \nabla V \cdot \nabla$ traditionally used in the Bakry-Emery-Ledoux theory [BGL14], in the subsequent chapters we will deal with the generator $L/2$ that is a more natural choice in our setting.

Examples 1.1.25 *Easy computations show that for the Kolmogorov semigroup, with generator $L = (\Delta - \nabla V \cdot \nabla)/2$ and reversing measure $d\mathbf{m} = e^{-V} dx$, the Γ and Γ_2 operators are given by*

$$\begin{aligned} \Gamma(f) &= \frac{1}{2} |\nabla f|^2 \\ \Gamma_2(f) &= \frac{1}{4} \|\nabla^2 f\|_2^2 + \frac{1}{4} (\nabla f)^T (\nabla^2 V) (\nabla f) \end{aligned}$$

where ∇^2 denotes the hessian matrix. Then condition $CD(\kappa/2, \infty)$ is satisfied if and only if the potential V is κ -convex, that is $\nabla^2 V \geq \kappa \text{Id}$. Note that in this case, the algebra is the set of all the smooth and compactly supported functions on \mathbb{R}^n , that is, $\mathfrak{A}_0 = C_0^\infty(\mathbb{R}^n)$

- The Ornstein-Uhlenbeck semigroup has $V = |x|^2/2$, $\nabla^2 V = \text{Id}$, (V is convex) and,

$$\Gamma_2(f) = \frac{1}{4} [\|\nabla^2 f\|_2^2 + |\nabla f|^2] \geq \frac{|\nabla f|^2}{4} = \frac{1}{2} \Gamma(f)$$

hence it satisfies the $CD(1/2, \infty)$ condition.

- When $V = 0$, by the Cauchy-Schwarz inequality,

$$\Gamma_2(f) = \frac{\|\nabla^2 f\|_2^2}{4} \geq \frac{1}{4n}(\Delta f)^2 = \frac{1}{n}(Lf)^2$$

hence the heat semigroup satisfies the $CD(0, n)$ condition, with n the dimension of the Euclidean space.

In practice often the $CD(\kappa, N)$ condition is used in the following equivalent forms.

Theorem 1.1.26 *The following are equivalent,*

(i) $CD(\kappa, \infty)$ is verified;

(ii.) For any $f \in \mathfrak{A}_0$ and any $t \geq 0$,

$$\Gamma(T_t f) \leq e^{-2\kappa t} T_t(\Gamma f); \quad (1.33)$$

(iii.) For any positive $f \in \mathfrak{A}_0$ and any $t \geq 0$,

$$\frac{|\nabla T_t f|^2}{T_t f} \leq e^{-\kappa t} T_t \left(\frac{|\nabla f|^2}{f} \right);$$

(iv.) For any $f \in \mathfrak{A}_0$ and any $t \geq 0$,

$$\sqrt{\Gamma(T_t f)} \leq e^{-\kappa t} T_t(\sqrt{\Gamma f}).$$

Proof

◁ The method used to prove the equivalence (i) \iff (ii) is a classical tool in the Bakry-Emery theory. We omit the computations and give an idea of the method.

Define a function $\Lambda : [0, t] \rightarrow \mathbb{R}$ for any $t > 0$, by

$$\Lambda(s) := T_s(\Gamma(T_{t-s} f)).$$

We differentiate in $s \in [0, t]$ and apply some of the properties stated in Proposition 1.1.22 to obtain, $\Lambda'(s) \geq 2\kappa\Lambda(s)$. At this point, by the Gronwall lemma, it follows that

$$\Lambda(t) \geq \Lambda(0) e^{2\kappa t}$$

which, by definition of Λ , gives (1.33). Analogous computations show the equivalence (i) \iff (iii). In order to prove (i) \iff (iv), we use a similar strategy with the function $\tilde{\Lambda}$, defined for any $t > 0$ as

$$\tilde{\Lambda}(s) := e^{-\kappa s} \sqrt{\Gamma(T_{t-s} f)}.$$

But in order to show that $\tilde{\Lambda}$ is increasing and conclude, we need the special trick to consider the reinforced curvature-dimension condition,

$$4\Gamma f(\Gamma_2 f - \kappa\Gamma f) \geq \Gamma(\Gamma f).$$

See [BGL14, Thm3.2.4, page 515]. ▷

Taking into account the dimension, there is a corrective term that occurs and yields to the next stronger result.

Theorem 1.1.27 *The following are equivalent,*

(i) *CD(0, N) is verified;*

(ii) *For any $f \in \mathfrak{A}_0$ and any $t \geq 0$,*

$$\Gamma(T_t f) \leq T_t(\Gamma f) - \frac{2t}{N}(LT_t f)^2.$$

The proof, via the dimensional log-Sobolev inequality, can be found in [BL06].

In order to understand the geometrical meaning of the Bakry-Emery condition, let us consider on a smooth complete connected Riemannian manifold (M, g) , the generator $L = \Delta_g$ where Δ_g is the Laplace-Beltrami operator. One might consider the more general case $L = \Delta_g + X$ for some smooth vector field X on M , but let us stick to the trivial but significant case $V = 0$. In this setting the Bochner identity says that for any smooth function f on M ,

$$-\nabla f \cdot \nabla \Delta_g f + \Delta_g \|\nabla f\|^2/2 = |\nabla^2 f|^2 + \text{Ric}(\nabla f, \nabla f)$$

where Ric is the Ricci curvature on M . Comparing Bochner's formula with Definition 1.1.21, one easily sees that the left hand side of (1.1.4) is equal to the formula of Γ_2 , that is

$$\Gamma_2 f = \|\nabla^2 f\|^2 + \text{Ric}(\nabla f, \nabla f). \quad (1.34)$$

Hence the Bakry-Emery condition (1.31) holds true if and only if

$$\text{Ric}(\nabla f, \nabla f) \geq \kappa \Gamma(f),$$

that is equivalent of saying that the Ricci curvature is bounded from below by some real κ . Note that the bound on the term $\|\nabla^2 f\|^2$ in (1.34) is just a consequence of the Cauchy-Schwarz inequality,

$$\|\nabla^2 f\|^2 \geq (\Delta_g f)^2/n.$$

Let us remark that the Laplacians are the only operators for which the dimension of the manifold and the dimension in the $\text{CD}(\kappa, N)$ condition are equal, while in general it can be proved that $N \geq n$.

The curvature-dimension condition (1.31) was introduced by Bakry and Emery in the seminal paper [BE85] with the aim of providing a criterion for hypercontractivity, hence by the Gross Theorem, for logarithmic Sobolev inequalities, see for instance the lecture notes [Bak94]. Thereafter the Bakry-Emery condition has known a large number of generalizations ([EKS15]), it has proven to be equivalent to several semigroup inequalities ([Wan05]), functional, geometric and transportation inequalities. We refer to the textbook [BGL14] and the related rich bibliography for the countless applications.

1.2 Optimal transport

The optimal transport theory is nowadays a huge increasing field, at the crossroads of analysis, geometry, probability, functional inequalities, numerical applications. The main references are the two monographs of Villani, [Vil03, Vil09] the one of Ambrosio, Gigli and Savaré, [AGS08], and the one of Santambrogio [San15]. Here we recall some of the basic notions and the fundamental results, and focus mainly on the particular case of quadratic cost on \mathbb{R}^n .

1.2.1 The Monge-Kantorovich problem

Introduced by Monge in 1871 [Mon81], the basic idea of the optimal transport theory, is to move in the *cheapest* possible way some kind of material (sand, particles, goods, ...) from one configuration to another one.

We consider as a state space the Euclidean space \mathbb{R}^n , and fix a measurable non-negative function $c(x, y) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$, that represents the cost to transport a unity of particle from position x to position y . We denote by $\mu_0, \mu_1 : \mathbb{R}^n \rightarrow \mathbb{R}^+$ respectively the initial and final distribution; under the assumption that no mass is added or lost in the transportation, we assume $\int_{\mathbb{R}^n} d\mu_0 = \int_{\mathbb{R}^n} d\mu_1 = \alpha$ for some positive $\alpha < \infty$, and without loss of generality, we can assume $\alpha = 1$.

Definition 1.2.1 (Static Monge-Kantorovich problem) *Given two probability measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ and a cost function $c(x, y) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$,*

$$\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) \pi(dxdy) ; \pi \in \Pi(\mu_0, \mu_1) \right\} \quad (1.35)$$

where $\Pi(\mu_0, \mu_1)$ is the set of probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ_0, μ_1 , that is

$$\Pi(\mu_0, \mu_1) := \left\{ \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) \text{ s.t. } \begin{array}{l} \pi(A \times \cdot) = \mu_0(A) \\ \pi(\cdot \times B) = \mu_1(B), \end{array} A, B \subset \mathbb{R}^n \text{ meas.} \right\},$$

or equivalently, all the couplings π of μ_0 and μ_1 , namely, all the probability measures $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ such that for any bounded measurable functions φ and ψ on \mathbb{R}^n ,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \psi(y)] \pi(dxdy) = \int_{\mathbb{R}^n} \varphi d\mu_0 + \int_{\mathbb{R}^n} \psi d\mu_1. \quad (1.36)$$

This one is actually a relaxed version, proposed by Kantorovich in the 1940s. The original Monge's problem had one more constraint : *no mass to be split*. It means that any particle, say at position x , can be sent to a unique destination $y = T(x)$. In this case, problem (1.35) writes as,

$$\inf \left\{ \int_{\mathbb{R}^n} c(x, T(x)) d\mu_0(x); T : \mathbb{R}^n \rightarrow \mathbb{R}^n \text{ meas. s.t. } T_{\#}\mu_0 = \mu_1 \right\}.$$

The Monge Kantorovich problem (1.41) admits also a dynamical formulation, that takes into account the path of each particle for any $0 < t < 1$, and not only the initial and final positions for $t = 0, t = 1$. To this aim we consider the path space $\Omega := C([0, 1], \mathbb{R}^n)$ already introduced in (1.10).

Definition 1.2.2 (Dynamic Monge-Kantorovich problem) *Let $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ and the cost function $C(\cdot) : \Omega \rightarrow \mathbb{R}^+$ be fixed,*

$$\inf \left\{ \int_{\Omega} C((\omega_t)_{0 \leq t \leq 1}) P(d\omega); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1 \right\}. \quad (1.37)$$

In order problem (1.35) and (1.37) to be equivalent, the cost functions $c(\cdot, \cdot)$ and $C(\cdot)$ have to be related by,

$$c(x, y) = \inf \{ C((\omega_t)_{0 \leq t \leq 1}) : \omega_0 = x, \omega_1 = y \}.$$

From now on we will consider the particular case when the cost function is the quadratic distance, namely $c(x, y) = |x - y|^2$,

$$\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \pi(dx dy) ; \pi \in \Pi(\mu_0, \mu_1) \right\}, \quad (1.38)$$

or its dynamical version,

$$\inf \left\{ E_P \int_0^1 |\dot{X}_t|^2 dt ; P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1 \right\}. \quad (1.39)$$

It can be easily verified that (1.38) and (1.39) are equivalent. Indeed,

$$E_P \int_0^1 |\dot{X}_t|^2 dt = \int_{\mathbb{R}^n \times \mathbb{R}^n} E_{P^{xy}} \int_0^1 |\dot{X}_t|^2 dt P_{01}(dx dy),$$

and with Jensen's inequality, we see that any solution \hat{P} of (1.39) satisfies,

$$\hat{P}^{xy} = \delta_{\gamma^{xy}},$$

where $\gamma_t^{xy} = (1 - t)x + ty$ is the constant speed geodesic between x and y . Hence, since $E_{\delta_{\gamma^{xy}}} \int_0^1 |\dot{X}_t|^2 dt = \int_0^1 |\dot{\gamma}_t^{xy}|^2 dt = |x - y|^2$, we obtain

$$E_{\hat{P}} \int_0^1 |\dot{X}_t|^2 dt = \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \hat{P}_{01}(dx dy).$$

Thus the two problems are equivalent, in the sense that they have the same minimal value, and the respective solutions $\hat{\pi}$ and \hat{P} are related by, $\hat{P}(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \delta_{\gamma^{xy}}(\cdot) \hat{\pi}(dx dy)$ and $\hat{P}_{01} = \hat{\pi}$.

Note that (1.39) can be written in the equivalent form,

$$\inf \left\{ \int_{\mathbb{R}^n} \int_0^1 |\dot{T}_t|^2 dt d\mu_0 ; T_0 = Id, T_{1\#}\mu_0 = \mu_1 \right\}. \quad (1.40)$$

In this case, (1.35) defines the square of the *quadratic Wasserstein pseudo-distance*, for any $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ as

$$W_2^2(\mu_0, \mu_1) := \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \pi(dx dy) ; \pi \in \Pi(\mu_0, \mu_1) \right\}. \quad (1.41)$$

In order W_2 to be a distance (also known as Kantorovich-Monge-Rubinstein distance), we need to reduce the set of definition of the marginals μ_0, μ_1 . Let us denote the set of probability measures with second order moment finite, by

$$\mathcal{P}_2(\mathbb{R}^n) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^n) ; \int |x|^2 \mu(dx) < \infty \right\}, \quad (1.42)$$

then the following proposition is satisfied.

Proposition 1.2.3 $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ is a metric space.

Proof

◁ It is immediate to see that $W_2(\mu_0, \mu_1) \geq 0$ and that $W_2(\mu_0, \mu_1) = W_2(\mu_1, \mu_0)$. Moreover, if $W_2(\mu_0, \mu_1) = 0$, it means that the infimum in (1.41) is attained, therefore it exists an optimal $\gamma \in \Pi(\mu_0, \mu_1)$ such that $\int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \gamma(dx dy) = 0$. It means that $\gamma(x, y) = \delta_{\{x=y\}}$, and this implies that for any test function ϕ

$$\int \phi d\mu_0 = \int \phi(x) d\gamma = \int \phi(y) d\gamma = \int \phi d\mu_1, \quad (1.43)$$

where the first and last equality are given by the fact that $\gamma \in \Pi(\mu_0, \mu_1)$ while the central equality is given to the property that γ is concentrated on the set $\{x = y\}$ as we just saw. Hence, (1.43) implies that $\mu_0 = \mu_1$. It remains to prove the triangular inequality. Let $\mu_0, \mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^n)$ and let $\gamma_1 \in \Pi(\mu_0, \mu_1)$ and $\gamma_2 \in \Pi(\mu_1, \mu_2)$ be optimal for the Monge Kantorovich problems respectively between (μ_0, μ_1) and (μ_1, μ_2) . The gluing lemma [Vil03, Ch. 1] makes sure the existence of a probability measure $\sigma \in P(\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n)$ such that $(X_{01})_{\#}\sigma = \gamma_1$ and $(X_{12})_{\#}\sigma = \gamma_2$, where (X_{01}) and (X_{12}) are the projection respectively on the two first and two last variables. Define $\gamma := (X_{02})_{\#}\sigma$. It can be easily verified that $\gamma \in \Pi(\mu_0, \mu_2)$, hence,

$$\begin{aligned} W_2(\mu_0, \mu_2) &\leq \left(\int |x - z|^2 d\gamma \right)^{1/2} = \left(\int |x - z|^2 d\sigma \right)^{1/2} = |x - z|_{L^2(\sigma)} \\ &\leq |x - y|_{L^2(\sigma)} + |y - z|_{L^2(\sigma)} = \left(\int |x - y|^2 d\sigma \right)^{1/2} + \left(\int |y - z|^2 d\sigma \right)^{1/2} \\ &= \left(\int |x - y|^2 d\gamma_1 \right)^{1/2} + \left(\int |y - z|^2 d\gamma_2 \right)^{1/2} = W_2(\mu_0, \mu_1) + W_2(\mu_1, \mu_2). \end{aligned}$$

▷

Remarks 1.2.4 (i). Note that Proposition 1.2.3 (and its proof) can be extended to $(\mathcal{P}_p(\mathbb{R}^n), W_p)$ for any $p \geq 1$.

(ii). Note that on $\mathcal{P}(\mathbb{R}^n)$, W_2 already satisfies the axioms of distance but it is not necessary finite. Whereas for any $x, y \in \mathbb{R}^n$ the inequality $|x - y|^2 \leq 2(|x|^2 + |y|^2)$ implies

$$\int \frac{|x - y|^2}{2} \pi(dx dy) \leq \int |x|^2 \mu_0(dx) + \int |y|^2 \mu_1(dy) < \infty,$$

whenever $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$.

1.2.2 McCann interpolations

From now on we will focus on the quadratic case, (1.41). We see here some existence and uniqueness results of the static and dynamic Monge-Kantorovich problem, and how these solutions are related.

Theorem 1.2.5 (Brenier's Theorem) Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, suppose that $\mu_0 \ll \mathcal{L}$. Then there exists a convex function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ differentiable a.e. such that

$$\mu_1 = \nabla \varphi_{\#} \mu_0$$

and the optimal map $\mathsf{T} = \nabla \varphi$ is unique, in other words,

$$W_2^2(\mu_0, \mu_1) = \int_{\mathbb{R}^n} |x - \nabla \varphi(x)|^2 d\mu_0(x).$$

Note that this is equivalent to say that the optimal coupling $\hat{\pi} \in \Pi(\mu_0, \mu_1)$ is given by

$$\hat{\pi} = (Id \times \nabla\varphi)_{\#}\mu_0.$$

This Theorem was later generalized by McCann to Riemannian manifolds in [McC01]. Let's see now the solution of the dynamic problem.

Theorem 1.2.6 (McCann's theorem) *Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ with $\mu_0 \ll \mathcal{L}$, then the unique solution $(\mathbb{T}_t)_{t \in [0,1]}$ of (1.38) is,*

$$\mathbb{T}_t = (1-t)Id + t\nabla\varphi, \quad (1.44)$$

where $\nabla\varphi$ is the solution of the Monge-Kantorovich static problem given by the Brenier Theorem 1.2.5.

Definition 1.2.7 (McCann interpolation) *For $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ with $\mu_0 \ll \mathcal{L}$ we define the McCann or displacement interpolation, the family of probability measures,*

$$\mu_t^{Mc} := (\mathbb{T}_t)_{\#}\mu_0 = ((1-t)Id + t\nabla\varphi)_{\#}\mu_0 \in \mathcal{P}(\mathbb{R}^n) \quad \forall 0 \leq t \leq 1. \quad (1.45)$$

where $(\mathbb{T}_t)_{0 \leq t \leq 1}$ is the solution of (1.40) given at Theorem 1.2.6.

In the rest of this section we shall drop the index Mc since there is no ambiguity with other interpolations.

The McCann interpolations are constant speed geodesics in the space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$. Indeed, for any $0 \leq t \leq 1$,

$$\begin{aligned} W_2^2(\mu_0, \mu_t) &= \int_{\mathbb{R}^n} |x - [(1-t)x - t\nabla\varphi(x)]|^2 d\mu_0 \\ &= t^2 \int_{\mathbb{R}^n} |x - \nabla\varphi|^2 d\mu_0 \\ &= t^2 W_2^2(\mu_0, \mu_1). \end{aligned}$$

Moreover, for any $0 \leq s \leq t \leq 1$,

$$W_2^2(\mu_s, \mu_t) = \int |x - \tilde{\mathbb{T}}x|^2 d\mu_s,$$

where $\tilde{\mathbb{T}}$ is the optimal map such that $\tilde{\mathbb{T}}_{\#}\mu_s = \mu_t$. Since $\mu_s = \mathbb{T}_s_{\#}\mu_0$,

$$W_2^2(\mu_s, \mu_t) = \int |\mathbb{T}_s x - \tilde{\mathbb{T}}\mathbb{T}_s x|^2 d\mu_0.$$

On the other hand, it can be easily verified that $\tilde{\mathbb{T}}\mathbb{T}_s_{\#}\mu_0 = \mu_t$ but $\mu_t = \mathbb{T}_t_{\#}\mu_0$, hence

$$\begin{aligned} W_2^2(\mu_s, \mu_t) &= \int |\mathbb{T}_s x - \mathbb{T}_t x|^2 d\mu_0 \\ &= \int |(1-s)x + s\nabla\varphi(x) - (1-t)x - t\nabla\varphi(x)|^2 d\mu_0 \\ &= (t-s)^2 \int |x - \nabla\varphi(x)|^2 d\mu_0. \end{aligned}$$

In other words, for any $0 \leq s, t \leq 1$,

$$W_2(\mu_s, \mu_t) = |t - s|W_2(\mu_0, \mu_1).$$

Formally, the optimal transportation is driven by the system,

$$\begin{cases} \partial_t \rho_t + \nabla \cdot (\rho_t \nabla \theta_t) = 0, & \rho_{t=0} = \mu_0 \\ \partial_t \theta_t + \frac{|\nabla \theta_t|^2}{2} = 0, & \theta_{t=0} = \varphi(x) - |x|^2/2, \end{cases} \quad (1.46)$$

where $\mu_0 \in \mathcal{P}_2(\mathbb{R}^n)$ is the prescribed initial marginal and φ is the convex function given by the Brenier Theorem 1.2.5.

Let us show how to obtain formally the equations of motion of $(\mu_t)_{0 \leq t \leq 1}$. We start by showing that the McCann interpolation is a weak (in the sense of distribution) solution of the transport equation with vector field $\nabla \theta_t$. We recall that $\mu_t = \mathbf{T}_{t\#} \mu_0$ where $\mathbf{T}_t(x) = (1-t)x + t\nabla \varphi(x)$. Define the associated velocity field by $v_t = d\mathbf{T}_t/dt \circ \mathbf{T}_t^{-1}$, or equivalently

$$v(t, \mathbf{T}_t(x)) = \frac{d}{dt} \mathbf{T}_t(x). \quad (1.47)$$

By definition of push forward we have for any test function f ,

$$\int f d\mu_t = \int f(\mathbf{T}_t) d\mu_0.$$

It can be verified that the dominated convergence theorem allows us to switch integral and derivative, hence we have

$$\frac{d}{dt} \int f d\mu_t = \int \partial_t f(\mathbf{T}_t) d\mu_0. \quad (1.48)$$

The integrand on the right hand side is given by

$$\partial_t f(\mathbf{T}_t) = \nabla f(\mathbf{T}_t) \frac{d\mathbf{T}_t}{dt} = \nabla f(\mathbf{T}_t) \cdot v_t(\mathbf{T}_t)$$

where the last equality follows directly by the definition of v_t . Hence from (1.48),

$$\frac{d}{dt} \int f \mu_t dx = \int \nabla f(\mathbf{T}_t) \cdot v_t(\mathbf{T}_t) d\mu_0 \stackrel{*}{=} \int \nabla f \cdot v_t d\mu_t = \int f \nabla \cdot (v_t \mu_t) dx,$$

where the marked equality is the definition of push forward and in the last equality we integrated by parts.

We study now the equation satisfied by the vector field v_t (1.47). We see by definition of the map \mathbf{T}_t (1.44) that the second order derivative is zero. Hence,

$$0 = \frac{d^2}{dt^2} \mathbf{T}_t(x) = \frac{d}{dt} v(t, \mathbf{T}_t) = -\frac{\partial}{\partial t} v(t, \mathbf{T}_t x) + v(t, \mathbf{T}_t x) \cdot \nabla v(t, \mathbf{T}_t x)$$

that is,

$$\partial_t v + v \cdot \nabla v = 0. \quad (1.49)$$

Moreover by definition of \mathbf{T}_t the initial condition is given by

$$v(0, \mathbf{T}_0(x)) = v(0, x) = \frac{d}{dt} \mathbf{T}_t(x) \Big|_{t=0} = \nabla \varphi(x) - x. \quad (1.50)$$

It can be actually proved (see [Vil03, Ch. 5]) that v_t is a gradient vector field, i.e. $v_t = \nabla\theta_t$. We can deduce the Hamilton Jacobi equation satisfied by θ_t by replacing $\nabla\theta_t$ to v_t in (1.49),

$$\partial_t \nabla\theta_t + \nabla\theta_t \cdot \nabla^2 \nabla\theta_t = 0$$

Note that,

$$\nabla \left(\frac{|\nabla\theta_t|^2}{2} \right) = \nabla\theta_t \cdot \nabla^2 \theta_t,$$

and replacing it in the previous equation, we can see that

$$\partial_t \theta_t + \frac{|\nabla\theta_t|^2}{2} = 0.$$

Moreover, according to (1.50), the initial condition is

$$\theta_0 = \varphi(x) - |x|^2/2. \quad (1.51)$$

Corollary 1.2.8 (Conserved quantity) *The Kinetic Energy is conserved along McCann geodesics, in other words, for all $0 \leq t \leq 1$,*

$$\int |\nabla\theta_t|^2 d\mu_t = \int |\nabla\varphi(x) - x|^2 d\mu_0 = W_2^2(\mu_0, \mu_1). \quad (1.52)$$

Proof

◁ We just show through a formal computation, that the derivative of the energy is zero.

$$\begin{aligned} \frac{d}{dt} \int |\nabla\theta_t|^2 d\mu_t &= \int 2\nabla\theta_t \cdot \nabla \partial_t \theta_t d\mu_t + |\nabla\theta_t|^2 \partial_t \mu_t dx \\ &= \int -2\nabla\theta_t \cdot \nabla \frac{|\nabla\theta_t|^2}{2} \mu_t - |\nabla\theta_t|^2 \nabla \cdot (\nabla\theta_t \mu_t) dx \\ &= \int -\nabla\theta_t \cdot \nabla |\nabla\theta_t|^2 + \nabla |\nabla\theta_t|^2 \cdot \nabla\theta_t \mu_t dx = 0, \end{aligned}$$

where in the last equality we integrated by parts. ▷

Remark 1.2.9 *One bad news about optimal transport maps, is that they are in general not smooth. Alexandrov's Theorem makes sure that φ being a convex function on \mathbb{R}^n , is twice differentiable almost everywhere, but for what concerns continuity and regularity everywhere it is a delicate issue. For instance as proved by Caffarelli [Vil09, Ch. 12], even if we consider two marginal measures with smooth, compactly supported probability densities on \mathbb{R}^n , with smooth and connected supports, yet the optimal transport associated with quadratic cost, is discontinuous.*

Assume that μ_0 and μ_1 have respectively density f and g with respect to the Lebesgue measure. Then the Brenier map, $\nabla\varphi$ verifying $\nabla\varphi\#\mu_0 = \mu_1$ satisfies for all regular enough test function h ,

$$\int h(y) g(y) dy = \int h(\nabla\varphi(x)) f(x) dx.$$

On the other hand we can do at least formally, the change of variable $y = \nabla\varphi(x)$,

$$\int h(y) g(y) dy = \int h(\nabla\varphi(x)) g(\nabla\varphi(x)) \det \nabla^2\varphi(x) dx.$$

Comparing these last two equations we derive the Monge-Ampère equation

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

Despite the regularity of f and g we cannot deduce any additional regularity property for φ (See [Vil09, Ch. 12] and [Vil03, Ch. 4]). Some regularity results on the solutions of the Monge-Ampère equation have been studied first in [Caf92] and then [McC97] under lighter assumptions, provided that the Hessian is interpreted in the sense of Alexandrov.

It is a big drawback in the applications of optimal transport to require non-smooth arguments, and it is an important part of the motivation of our work as it will be made clear in the next chapter.

1.2.3 Kantorovich and Benamou-Brenier formulations

In this section we present two alternative formulations of the Monge-Kantorovich problem, that are useful when proving functional inequalities. One is the Kantorovich dual formulation by mean of the Hopf-Lax semigroup, solution of the Hamilton Jacobi equation, very practical in computation. The other is the Benamou-Brenier formulation by mean of minimal action.

Kantorovich dual formulation

We give first an intuitive interpretation of the Monge Kantorovich problem, and the dual problem. The primal problem, as we already mentioned, consists in minimizing the cost $c(x, y)$ of the transportation of some goods from position x to position y . As an alternative we can think to the transportation in terms of price. Imagine to operate the transportation for someone else. You fix a price $\psi(x)$ to charge the element in position x that you sell in position y at the price $\varphi(y)$. Your goal is to maximize your profit $\varphi(y) - \psi(x)$ with the condition not to exceed the cost $c(x, y)$ (otherwise is not convenient). This can be written as,

$$\sup \left\{ \int \varphi(y) d\mu_1(y) - \int \psi(x) d\mu_0(x); \varphi(y) - \psi(x) \leq c(x, y) \right\}.$$

The dual formulation holds in extremely generality on the state spaces, as long as they are Polish spaces, and on the cost function, as long as it is lower semi-continuous. It was first proved by Kantorovich in [Kan42b] in a particular case; a complete and general proof can be found in [Vil09]. In what follows, we will state the Kantorovich dual formulation in the specific case of the quadratic Wasserstein distance, (1.41).

Theorem 1.2.10 (Kantorovich dual formulation for W_2^2) For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, then

$$W_2^2(\mu_0, \mu_1) = \sup_{\varphi} \left\{ \int_{\mathbb{R}^n} \varphi d\mu_1 - \int_{\mathbb{R}^n} Q\varphi d\mu_0 \right\}, \quad (1.53)$$

where the supremum runs over all bounded continuous function φ and

$$Q\varphi(x) = \sup_{y \in \mathbb{R}^n} \{ \varphi(y) - |x - y|^2 \}, \quad x \in \mathbb{R}^n. \quad (1.54)$$

Intuitively it comes by the constrain on φ and ψ ,

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

Indeed it can be written as,

$$\psi(x) \geq \varphi(x) - |x - y|^2$$

and taking the sup over $y \in \mathbb{R}^n$ it gives (1.54).

Remark 1.2.11 *It is often expressed in the equivalent form,*

$$W_2^2(\mu_0, \mu_1) = \sup_{\psi} \left\{ \int_{\mathbb{R}^n} \tilde{Q}\psi d\mu_1 - \int_{\mathbb{R}^n} \psi d\mu_0 \right\}, \quad (1.55)$$

where the supremum runs over all bounded continuous function ψ and

$$\tilde{Q}\psi(y) = \inf_{x \in \mathbb{R}^n} \{ \psi(x) + |y - x|^2 \}, \quad y \in \mathbb{R}^n. \quad (1.56)$$

The map $Q\psi$ is called the sup-convolution of ψ and its defining identity is often referred to as the Hopf-Lax formula.

Let us briefly recall that the Hopf-Lax formula describes the viscosity solution of some class of Hamilton-Jacobi equations. Consider the Cauchy problem,

$$\begin{cases} \partial_t u + H(|\nabla u|) = 0 & t > 0, \\ u = f & t = 0. \end{cases} \quad (1.57)$$

Under some assumption on the Hamiltonian H and the initial datum f , it can be proved [Vil09, Ch. 22] that

$$u(t, x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + tH^* \left(\frac{|x - y|}{t} \right) \right\}, \quad t > 0$$

and $u(0, x) = f(x)$ when $t = 0$ is the solution of (1.57). Here H^* is the Legendre-Fenchel transform of H that is defined by $H^*(r) = \sup_{s \in \mathbb{R}^+} (rs - H(s))$. So, when $H(s) = s^2/2$ the also $H^* = s^2/2$. The HJ equation in (1.57) writes as,

$$\partial_t u + \frac{|\nabla u|^2}{2} = 0$$

and the the solution for all $t > 0$ is given by,

$$u(t, x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{|x - y|^2}{2t} \right\}$$

and for $t = 1$ it coincides up to a factor 1/2 with (1.56) with initial datum f .

Moreover it can be verified that the optimal ψ in (3.3) is given by $\theta_0(x) = \varphi(x) - |x|^2/2$ introduced in (1.51).

The functions φ, ψ attaining the sup are called *Kantorovich potentials*.

Benamou-Brenier formulation

The Monge Kantorovich problem admits a differential formulation, based on an Eulerian point of view on the optimal transport problem. It was first established in [BB00] by Benamou and Brenier with the purpose of new numerical methods. The idea is to associate to the cost of transportation, an *action*, as a way to measure the effort of moving the system along continuous curves. Given a particle system, moving according to the transport equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0.$$

The kinetic action is the time integral of the kinetic energy, that is, up to a factor 1/2,

$$A(\rho, v) = \int_0^1 \int_{\mathbb{R}^n} |v_t(x)|^2 \rho_t(x) dx dt.$$

Theorem 1.2.12 (Benamou-Brenier formulation for W_2^2) For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu, v)} \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 d\nu_t dt,$$

where the infimum runs over all paths $(\nu_t, v_t)_{t \in [0,1]}$ where $\nu_t \in \mathcal{P}(\mathbb{R}^n)$ and $v_t(x) \in \mathbb{R}^n$ are such that ν_t is absolutely continuous with respect to time in the sense of [AGS08, Ch. 1] for all $0 \leq t \leq 1$, $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and

$$\partial_t \nu_t + \nabla \cdot (\nu_t v_t) = 0, \quad 0 \leq t \leq 1.$$

In this equation which is understood in a weak sense, $\nabla \cdot$ stands for the standard divergence of a vector field in \mathbb{R}^n and ν_t is identified with its density with respect to Lebesgue measure. This general result is proved in [AGS08, Ch. 8]. A proof under the additional requirement that μ_0, μ_1 have compact supports is available in [Vil03, Thm. 8.1].

Roughly speaking it can be proved that

$$W_2^2(\mu_0, \mu_1) \leq \inf_{(\nu, v)} \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 d\nu_t dt.$$

Formally this is a result of the Jensen inequality. Indeed consider any ρ_t between μ_0 and μ_1 and let T_t be the associated trajectory and v_t the velocity field. Then its action is given by

$$\begin{aligned} \int_0^1 \int |v_t(x)|^2 d\rho_t(x) dt &= \int_0^1 \int |v_t(T_t(x))|^2 d\mu_0 dt \\ &= \int_0^1 \int \left| \frac{d}{dt} T_t x \right|^2 d\mu_0 dt \\ &\geq \int \left| \int_0^1 \frac{d}{dt} T_t(x) dt \right|^2 d\mu_0 \\ &= \int |T_1(x) - T_0(x)|^2 d\mu_0 = W_2^2(\mu_0, \mu_1), \end{aligned}$$

since according to the dynamical formulation (1.40) $T_0 = Id$ and $T_{1\#}\mu_0 = \mu_1$. The optimal couple, that is, the couple attaining the equality is given by $(\mu_t, \nabla\theta_t)$ where μ_t is the McCann interpolation and the vector field $\nabla\theta_t$ is defined at (1.46). It can be seen by

recalling that $\mu_t = \mathbf{T}_t\#\mu_0$ where the optimal map \mathbf{T}_t is given at (1.45) and the associated velocity field is (1.47), that can be rewritten also as,

$$v_t = \left(\frac{d}{dt} \mathbf{T}_t \right) \circ \mathbf{T}_t^{-1} = (\nabla\varphi - Id) \circ \mathbf{T}_t^{-1}.$$

It implies that,

$$\begin{aligned} \int |v_t|^2 \mu_t dx &= \int |(\nabla\varphi - Id) \circ \mathbf{T}_t^{-1}|^2 \mu_t dx \\ &\stackrel{\star}{=} \int |(\nabla\varphi - Id) \circ (\mathbf{T}_t^{-1} \circ \mathbf{T}_t)|^2 \mu_0 dx = \int |\nabla\varphi(x) - x|^2 \mu_0 dx = W_2^2(\mu_0, \mu_1), \end{aligned}$$

where the \star equality is just the push forward. Note that somehow we recovered (1.52).

Chapter 2

The Schrödinger problem

In this section we present the main protagonist of the manuscript, the Schrödinger problem. A reference on the subject is the survey of Léonard [Léo14b], and the other main references will be recalled along the chapter. Introduced by Schrödinger in 1931 [Sch31], in modern terms, it consists, roughly speaking, in minimizing the relative entropy with respect to a fixed reference path measure $R \in \mathcal{M}_+(\Omega)$, among all the path probability measures, satisfying the marginal constraints $P_0 = \mu_0$ and $P_1 = \mu_1$. In other words,

$$H(P|R) \rightarrow \min ; P \in \mathcal{P}(\Omega), \text{ s.t. } P_0 = \mu_0, P_1 = \mu_1,$$

where $P_t = (X_t)_\# P \in \mathcal{P}(\mathbb{R}^n)$ for all $0 \leq t \leq 1$. First we define the problem in its dynamic and static formulations, give some sufficient conditions for the existence of solutions and describe the minimizers, called *entropic bridges* at Section 2.1. Then starting from Section 2.2 we will focus on the case of the reference path measure associated to the Kolmogorov process, whose generator is given by $L = (\Delta - \nabla V \cdot \nabla)/2$, we define the *entropic interpolations* as continuous time projection of entropic bridges, $d\mu_t = \hat{P}_t = T_t f T_{1-t} g dm$; we explore their kinematics, by identifying the correspondent Nelson's velocities and deriving the associated time evolution equation. At Section 2.3 we discuss different possible definitions of *entropic cost* between the two probability measures μ_0, μ_1 from the minimal value attained by the minimizer. Then at Section 2.4 we will make the link between the Schrödinger problem and the Monge-Kantorovich problem introduced at Section 1.2. In particular we will see that the Schrödinger problem associated to the slowed down reference measure R^ε with generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ is a regular approximation of the square of the Wasserstein distance. At Section 2.5 we will study the behavior of the relative entropy and the Fisher information along the entropic interpolation, in analogy to their behavior along the heat flow and the McCann interpolations. And finally at Section 2.6 we give the statistical physics interpretation of the problem that motivated the question of Schrödinger and retrace the chronology and the related literature.

2.1 Solutions of the Schrödinger problem

Definition 2.1.1 (Dynamical formulation) *Let the marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ and a reference σ -finite path measure $R \in \mathcal{M}_+(\Omega)$ be fixed. The dynamic Schrödinger problem is,*

$$\inf\{H(P|R); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \in (-\infty, \infty], \quad (2.1)$$

As for the Monge-Kantorovich problem, it also admits a static formulation. Consider the projection,

$$R_{01} := (X_0, X_1)_\# R \in \mathcal{M}_+(\mathbb{R}^n \times \mathbb{R}^n)$$

of R onto the endpoint space $\mathbb{R}^n \times \mathbb{R}^n$.

Definition 2.1.2 (Static formulation) *For any $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, the static Schrödinger problem is,*

$$\inf\{H(\pi|R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi_0 = \mu_0, \pi_1 = \mu_1\} \in (-\infty, \infty], \quad (2.2)$$

where $\pi_0 := \pi(\cdot \times \mathbb{R}^n)$, and $\pi_1 := \pi(\mathbb{R}^n \times \cdot)$.

As stated in the next theorem, the dynamic and the static formulations are actually equivalent.

Theorem 2.1.3 (Equivalence between the static and dynamic formulations) *The Schrödinger problems (2.1) and (2.2) admit respectively at most one solution $\hat{P} \in \mathcal{P}(\Omega)$ and $\hat{\pi} \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ and they are related by,*

$$\hat{P}(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{xy}(\cdot) \hat{\pi}(dxdy) \in \mathcal{P}(\Omega),$$

that means $\hat{P}_{01} = \hat{\pi} \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ and that \hat{P} has the same bridges of the reference measure $\hat{P}^{xy} = R^{xy}$, for $\hat{\pi}$ -almost every (x, y) .

Proof

◁ We recover here the proof of [Föl88].

The minimization problems (2.1) and (2.2) are strictly convex, hence if a solution exists, it has to be unique. Let us assume that such a solution exists.

The additive decomposition (1.6), when $\phi = (X_0, X_1)$ writes as,

$$H(P|R) = H(P_{01}|R_{01}) + \int_{\mathbb{R}^n \times \mathbb{R}^n} H(P^{xy}|R^{xy}) P_{01}(dxdy).$$

It implies $H(P_{01}|R_{01}) \leq H(P|R)$ with equality whenever $P^{xy} = R^{xy}$ for P_{01} -almost every $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. Therefore, $\hat{P} \in \mathcal{P}(\Omega)$ is solution of (2.1) if and only if it disintegrates as,

$$\hat{P}(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} R^{xy}(\cdot) \hat{\pi}(dxdy)$$

with $\hat{\pi} = \hat{P}_{01} \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ solution of (2.2). ▷

Note that a priori the minimal value in (2.1) and (2.2) is not necessary finite and it might also take negative values. We collect some sufficient and necessary condition for the reference measure $R \in \mathcal{M}_+(\Omega)$ and the marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ in order the Schrödinger problem (2.1) (or (2.2)) to admit a solution.

Theorem 2.1.4 (Existence and uniqueness) *Let $R \in \mathcal{M}_+(\Omega)$ be Markov, with reversing measure $m \in \mathcal{P}(\mathbb{R}^n)$, and let us assume that,*

1. there exist two real non-negative measurable functions on \mathbb{R}^n , A and B such that

$$(i) \quad R_{01}(dxdy) \geq e^{-A(x)-A(y)} m(dx)m(dy),$$

$$(ii) \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{-B(x)-B(y)} R_{01}(dxdy) < \infty,$$

$$(iii) \int_{\mathbb{R}^n} A + B d\mu_i < \infty, \text{ for } \mu_0, \mu_1;$$

$$2. H(\mu_i|m) < \infty \text{ for } \mu_0, \mu_1,$$

then (2.1) (hence (2.2)) admits a unique solution.

Moreover, assumption 2. is also a necessary condition.

Proof

◁ We already mentioned in the proof of Theorem 2.1.3 that from the strictly convex nature of the problem, if a solution exists then it is unique.

Let us show that the existence of the function B such that (i)-(ii) are satisfied, makes sure that the minimal value of (2.2) is bounded from below. In the generalization of the relative entropy to unbounded measures in Section 1.1.1, we used the existence of a real valued measurable function w such that (1.4) holds true. Assumption (ii) is nothing but (1.4) when $\nu = R_{01}$ and $w(x, y) = B(x) + B(y)$. It allows us to write the relative entropy as in (1.5), for all $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, by

$$H(\pi|R_{01}) = H(\pi|R_{01}^B) - \int_{\mathbb{R}^n \times \mathbb{R}^n} [B(x) + B(y)] \pi(dxdy) - z_B. \quad (2.3)$$

Since in the minimization problem (2.2) we take into account only probability measures $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals μ_0, μ_1 , the integral on the right hand side, writes as,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [B(x) + B(y)] \pi(dxdy) = \int_{\mathbb{R}^n} B d\mu_0 + \int_{\mathbb{R}^n} B d\mu_1. \quad (2.4)$$

With decomposition (2.3) in mind, it is now clear that assumptions (ii) and (iii) make sure respectively that $z_b < \infty$, and that (2.4) is bounded from above. Moreover, the relative entropy $H(\pi|R_{01}^B) \geq 0$ since $R_{01}^B \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, thus (2.3) is bounded from below. It remains to prove that the minimal value of (2.2) is bounded from above. Consider the probability measure $\mu_0 \otimes \mu_1 \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, and compute the relative entropy of it with respect to R_{01} ,

$$\begin{aligned} H(\mu_0 \otimes \mu_1|R_{01}) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \log \frac{d\mu_0 \otimes \mu_1}{dR_{01}} \mu_0(dx) \mu_1(dy) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \log(\mu_0(x) \mu_1(y)) \mu_0(dx) \mu_1(dy) - \int_{\mathbb{R}^n \times \mathbb{R}^n} \log R_{01} \mu_0(dx) \mu_1(dy) \\ &\stackrel{(i)}{\leq} H(\mu_0|m) + H(\mu_1|m) + \int_{\mathbb{R}^n} A(x) d\mu_0(dx) + \int_{\mathbb{R}^n} A(y) d\mu_1(dy) \\ &\stackrel{(iii)+2.}{\leq} +\infty. \end{aligned}$$

Hence, since $\mu_0 \otimes \mu_1 \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ and satisfies the marginal constrains,

$$\inf\{H(\pi|R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi_0 = \mu_0, \pi_1 = \mu_1\} \leq H(\mu_0 \otimes \mu_1|R_{01}) < \infty.$$

Finally we show that 2. is actually a necessary condition. By the additive property (1.6) with $\phi = X_0$, we deduce that for all $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ satisfying the marginal condition $\pi_0 = \mu_0$, it holds $H(\pi|R_{01}) \geq H(\mu_0|m)$. Taking the inf over all the probability π satisfying the marginal constrains, we obtain

$$\inf\{H(\pi|R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi_0 = \mu_0, \pi_1 = \mu_1\} \geq H(\mu_0|m).$$

This proves that if $H(\mu_0|m) = +\infty$, the optimal value of the Schrödinger problem on the right hand side cannot be finite. \triangleright

Examples 2.1.5 We verify here that assumptions (i)-(iii) are satisfied in two main examples introduced at Examples 1.1.8.

- The simplest example is R associated to the Brownian motion, with reversing measure $m = \mathcal{L}$ on \mathbb{R}^n . In this case,

$$R_{01}(dxdy) = e^{-|x-y|^2/2}/(2\pi)^{n/2}dxdy,$$

and by the simple quadratic inequality $(a-b)^2 \leq 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$, it is immediate to see that by choosing $A(x) = B(x) = |x|^2$ assumptions (i)-(ii) are satisfied.

- Analogous computations show that assumptions (i)–(iii) are satisfied with the same functions $A(x) = B(x) = |x|^2$, up to multiplication by a real factor, when R is the law associated to the Ornstein-Uhlenbeck process, in Example 1.1.20 (b),

$$R_{01}(dxdy) = \exp\left(-\frac{|y - xe^{-1/2}|^2}{2(1 - e^{-1})}\right) / (2\pi(1 - e^{-1}))^{-n/2}dxdy.$$

In both cases the assumption on the marginals (iii) reduces to ask the measures μ_0, μ_1 to have second order moment finite, i.e.

$$\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$$

defined at (1.42) and finite entropy relative to the reversing measure, $H(\mu_i|m) < \infty$ for μ_0, μ_1 .

Provided that a solution exists, we are now interested in the minimizer.

Theorem 2.1.6 (Entropic bridge) Suppose that hypothesis 1. and 2. in Theorem 2.1.4 are satisfied. Then the unique solution of (2.1) is called entropic bridge and it is characterized by the formula,

$$\hat{P} = f(X_0)g(X_1)R \in \mathcal{P}(\Omega), \quad (2.5)$$

where f, g are measurable non-negative functions, unique solutions, up to the multiplications af, ga^{-1} for some constant $a \in \mathbb{R}$, of the Schrödinger system

$$\begin{cases} \mu_0 &= fT_1g \\ \mu_1 &= gT_1f, \end{cases} \quad (2.6)$$

where $(T_t)_{t \geq 0}$ is the Markov semigroup associated with the measure R . Moreover, the unique solution of the static problem (2.2) $\hat{\pi} \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$, is given by

$$\hat{\pi}(dxdy) = f(x)g(y)R_{01}(dxdy) \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n).$$

Sketch of the Proof

◁ Let B be the function appearing at (i)-(ii) in Theorem 2.1.4. Define the sets, $C_B(\mathbb{R}^n) := \{u : \mathbb{R}^n \rightarrow \mathbb{R} : \sup |u|/B < \infty\}$ and $\mathcal{P}_B(\mathbb{R}^n) = \{\mu \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} B d\mu < \infty\}$. We recall that to say that $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ has marginals μ_0, μ_1 is equivalent to (1.36). Hence, the variational formula (1.2) writes for any $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals μ_0, μ_1 as,

$$H(\pi|R_{01}) = \sup \left\{ \int \varphi d\mu_0 + \int \psi d\mu_1 - \log \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{\varphi \oplus \psi} dR_{01}; \varphi, \psi \in C_B(\mathbb{R}^n) \right\} \quad (2.7)$$

where $\varphi \oplus \psi(x, y) = \varphi(x) + \psi(y)$. It is proved in [Léo01a] that the minimal value in (2.2) is equal to the maximal value in (2.7). Hence, if $\hat{\pi}$ is the minimizer of (2.2), the case of equality (1.3) gives

$$\frac{d\pi}{dR_{01}} = e^{\varphi \oplus \psi}.$$

Since the variational formulation still holds for the dynamic formulation $H(P|R)$, analogous computations yield to

$$\frac{dP}{dR} = e^{\varphi(X_0) + \psi(X_1)}$$

that is (2.5), with $f = e^{\varphi}$ and $g = e^{\psi}$. ▷

The solvability of the system (2.6) was left open by Schrödinger. The existence of solutions has been proved first by Bernstein [Ber32] when the reference measure is the Wiener measure on \mathbb{R} and later in the 40s by Fortet [For40] under more general hypothesis through a method of successive approximations. Subsequently Beurling in 1960 [Beu60] proposed an alternative proof via functional analysis, and more recently, in [RT93] the existence of the solution was obtained as a consequence of Csiszár's projection theorem, whereas in [BLN94] through a variational method.

Remark 2.1.7 *Let us point out some issue of the (sketch of) the proof we presented here. Nothing assures that the supremum is attained in (2.7) by functions in $C_B(\mathbb{R}^n)$. In general we only have that $\varphi \oplus \psi$ is R_{01} -measurable [Léo01c] and this does not imply that φ and ψ are respectively R_0 and R_1 measurable. The same measurability problems are of course inherited by the functions f and g , thus we are not a priori allowed to write*

$$T_1 f = E_R(f(X_0)|X_1 = y)$$

(and the same for $T_{1-t}g$) in the marginal conditions in (2.6). To overcome these problems, it is necessary to add an assumption on the reference R ; in particular, assume that there exists some $0 < t_0 < 1$ and some measurable $F \subset \mathbb{R}^n$ such that $R_{t_0}(F) > 0$ and

$$R_{01} \ll R((X_0, X_1) \in \cdot | X_{t_0} = z), \quad \forall z \in F.$$

For more details see [LRZ14] and [Léo14b, Thm. 2.6]

For completeness we mention that there is a complementary way to think to entropic bridges, that can be found in the survey [Léo14b], through the definition of (f, g) -transforms. Given a reversible Markov reference measure R , the basic idea is to fix two functions f, g and build two measures μ_0, μ_1 by (2.6).

Definition 2.1.8 ((f, g)-transform) *Let $f, g : \mathbb{R}^n \rightarrow [0, \infty)$ be two non-negative measurable functions such that $E_R(f(X_0)g(X_1)) = 1$. The (f, g) -transform of R is defined by,*

$$P := f(X_0)g(X_1)R \in \mathcal{P}(\Omega).$$

This definition is a two-sided version of the h -transform introduced by Doob in [Doo57].

Theorem 2.1.9 *Let P be an (f, g) -transform of R , as defined at Definition 2.1.8, with the functions f, g satisfying,*

$$\int_{\mathbb{R}^n} |fg \log f| dm < \infty, \quad \int_{\mathbb{R}^n} |fg \log g| dm < \infty.$$

Then P is the entropic bridge between $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ given by,

$$\begin{aligned} \mu_0 &= fT_1g dm \\ \mu_1 &= T_1fg dm, \end{aligned}$$

where $(T_t)_{t \geq 1}$ is the Markov semigroup associated to R .

For the proof see [Léo14b, Thm.3.2].

2.2 Kinematics of entropic bridges

In this section we look closer at the minimizer of the Schrödinger problem. It is assumed once for all that hypothesis of Theorem 2.1.4 are satisfied.

The results that we discuss here can be formulated for any σ -finite measure $R \in \mathcal{M}_+(\Omega)$ satisfying the assumptions of Theorem 2.1.4. However we decided to focus on a specific $R \in \mathcal{M}_+(\Omega)$, first because it is what will be actually used in the applications in the following chapters, and then to provide some direct proofs and explicit properties. For the literature about the Schrödinger problem in general we refer to [Léo14b], and therein references.

The reference measure. Let $\Omega = C([0, 1], \mathbb{R}^n)$ be the space of continuous path on the state space \mathbb{R}^n . We consider the measure $R^\varepsilon \in \mathcal{M}_+(\Omega)$ associated to the stochastic diffusion process (1.12), with diffusion coefficient $\sigma^2 = \varepsilon$, for some $\varepsilon > 0$ and a drift vector field of the form $b = -\varepsilon \nabla V/2$, for some real valued potential V on \mathbb{R}^n , namely,

$$\begin{aligned} dX_t^\varepsilon &= -\varepsilon \nabla V(X_t^\varepsilon)/2 dt + \sqrt{\varepsilon} dW_t, \quad 0 \leq t \leq 1, \\ \mathcal{L}(X_0) &= \mathbf{m} \end{aligned} \tag{2.8}$$

where $d\mathbf{m} = e^{-V} dx$ is the reversing measure. We assume the potential V to be \mathcal{C}^2 on \mathbb{R}^n and, whenever $\int_{\mathbb{R}^n} e^{-V} dx < \infty$, such that $\int_{\mathbb{R}^n} e^{-V} dx = 1$. Furthermore, we assume that there exists some constant $c > 0$ such that one of the following assumptions holds true:

- (i) $\lim_{|x| \rightarrow \infty} V(x) = +\infty$ and $\inf\{|\nabla V|^2 - \Delta V/2\} > -\infty$, or
- (ii) $-x \cdot \nabla V(x) \leq c(1 + |x|^2)$, for all $x \in \mathbb{R}^n$.

Under this assumption, (2.8) admits a unique solution, see [Roy99, Thm.2.2.19].

Note that R^ε is not necessary a probability measure on Ω . Indeed in the case $V = 0$, R^ε is the measure associated to the standard reversible Brownian motion, that on \mathbb{R}^n has infinite mass. This is the reason why we introduced at Section 1.1.1 the generalization of the relative entropy to unbounded measures.

In the rest of the manuscript, we will mainly consider the case $V = 0$, already discussed, or the case of κ -convex V , that is, for some $\kappa \in \mathbb{R}$,

$$[\nabla V(x) - \nabla V(y)] \cdot (y - x) \leq -\kappa |x - y|^2/2, \quad \forall x, y \in \mathbb{R}^n.$$

The reason why we considered the slowed down process (2.8), instead of

$$\begin{aligned} dX_t &= -\nabla V(X_t)/2 dt + dW_t, \quad 0 \leq t \leq 1, \\ \mathcal{L}(X_0) &= \mathbf{m} \end{aligned} \quad (2.9)$$

will be clear at Section 2.4 when we will take the limit for $\varepsilon \rightarrow 0$, to connect the Schrödinger problem to the optimal transport theory, and at Section 2.6 when we will give some physical motivation/interpretation.

As we already mentioned in the Examples 1.1.20, the operators associated to the Kolmogorov SDE (2.9) are

$$Lf = (\Delta f - \nabla V \cdot \nabla f)/2, \quad \Gamma f = \frac{|\nabla f|^2}{2}, \quad \Gamma_2 f = \frac{1}{2}L(|\nabla f|^2) - \nabla f \cdot \nabla Lf. \quad (2.10)$$

For the slowed down process (2.8), we denote,

$$L^\varepsilon = \varepsilon L, \quad \Gamma^\varepsilon f = \varepsilon \Gamma f, \quad \Gamma_2^\varepsilon f = \varepsilon^2 \Gamma_2. \quad (2.11)$$

Definition 2.2.1 (Entropic interpolation) *The R^ε -entropic interpolation between μ_0 and μ_1 is defined as the marginal flow of the entropic bridge (2.5), that is*

$$\mu_t^\varepsilon := P_t^\varepsilon = (X_t)_\# P^\varepsilon \in \mathcal{P}(\mathbb{R}^n), \quad \forall 0 \leq t \leq 1.$$

In particular, it is characterized by the formula,

$$d\mu_t^\varepsilon = T_t^\varepsilon f_\varepsilon T_{1-t}^\varepsilon g_\varepsilon d\mathbf{m}, \quad \forall 0 \leq t \leq 1. \quad (2.12)$$

It is sometimes more convenient to use the equivalent exponential form,

$$d\mu_t^\varepsilon = e^{\frac{\varphi_t^\varepsilon + \psi_t^\varepsilon}{\varepsilon}} d\mathbf{m}, \quad \forall 0 \leq t \leq 1, \quad (2.13)$$

where for any $t \in [0, 1]$,

$$\begin{aligned} \varphi_t^\varepsilon &:= \varepsilon \log T_t^\varepsilon f_\varepsilon \\ \psi_t^\varepsilon &:= \varepsilon \log T_{1-t}^\varepsilon g_\varepsilon \end{aligned} \quad (2.14)$$

with $f_\varepsilon, g_\varepsilon$ solutions of (2.6) and $(T_t^\varepsilon)_{t \geq 0}$ the Markov semigroup associated with the generator L^ε . Note that in analogy to the Kantorovich potentials (attaining the sup in (1.53) and (1.55)), φ_0^ε and ψ_1^ε are referred to as *Schrödinger potentials*.

Following the traditional notation in [Vil09], we denote the density of the measure $d\mu_t^\varepsilon$ with respect to the invariant measure \mathbf{m} and with respect to the Lebesgue measure \mathcal{L} respectively as,

$$\rho_t^\varepsilon = \frac{d\mu_t^\varepsilon}{d\mathbf{m}}, \quad \mu_t^\varepsilon = \frac{d\mu_t^\varepsilon}{dx}.$$

Since we consider the case $\mathbf{m} = e^{-V} dx$, the relation between the densities is

$$\mu_t^\varepsilon = e^{-V} \rho_t^\varepsilon. \quad (2.15)$$

Proposition 2.2.2 *The potentials $\varphi_t^\varepsilon, \psi_t^\varepsilon$ are respectively solutions of the following Hamilton-Jacobi-Bellman equations,*

$$\begin{cases} \partial_t \varphi_t^\varepsilon - \Gamma \varphi_t^\varepsilon - \varepsilon L \varphi_t^\varepsilon = 0, & 0 < t \leq 1, \\ \varphi_0^\varepsilon = \varepsilon \log f_\varepsilon & t = 0 \end{cases} \quad \begin{cases} \partial_t \psi_t^\varepsilon + \Gamma \psi_t^\varepsilon + \varepsilon L \psi_t^\varepsilon = 0, & 0 \leq t < 1, \\ \psi_1^\varepsilon = \varepsilon \log g_\varepsilon, & t = 1. \end{cases} \quad (2.16)$$

where L and Γ are given at (2.10).

Proof

◁ The proof is an immediate consequence of Definition (2.14) and Γ_2 -calculus in Section 1.1.4. Without any ambiguity, we can drop the index ε in the potentials φ_t and ψ_t for the rest of the proof.

$$\begin{aligned} \partial_t \varphi_t &= \varepsilon \frac{L^\varepsilon T_t f}{T_t f} = \varepsilon e^{-\varphi_t/\varepsilon} L^\varepsilon (e^{\varphi_t/\varepsilon}) \\ &= \varepsilon e^{-\varphi_t/\varepsilon} \left[\frac{e^{\varphi_t/\varepsilon}}{\varepsilon} L^\varepsilon \varphi_t + \frac{e^{\varphi_t/\varepsilon}}{\varepsilon^2} \Gamma^\varepsilon \varphi_t \right] \\ &= L^\varepsilon \varphi_t + \frac{1}{\varepsilon} \Gamma^\varepsilon \varphi_t = \varepsilon L \varphi_t + \Gamma \varphi_t, \end{aligned}$$

where the last equality follows by the notation convention (2.11). The case of ψ_t is exactly the same but with opposite sign. ▷

Kinematics of μ_t^ε . As a consequence of Proposition 2.2.2 we can derive the time evolution equation for the entropic interpolation. Indeed, taking the sum of the two equations in (2.16), we obtain

$$\partial_t(\varphi_t^\varepsilon + \psi_t^\varepsilon) + \varepsilon L(\psi_t^\varepsilon - \varphi_t^\varepsilon) + (\Gamma \psi_t^\varepsilon - \Gamma \varphi_t^\varepsilon) = 0 \quad (2.17)$$

Thanks to property (viii) in Proposition 1.1.22, and (2.13), it becomes

$$\varepsilon \partial_t(\log \rho_t^\varepsilon) + \varepsilon L(\psi_t^\varepsilon - \varphi_t^\varepsilon) + \Gamma(\varepsilon \log \rho_t^\varepsilon, \psi_t^\varepsilon - \varphi_t^\varepsilon) = 0$$

Namely,

$$\partial_t \rho_t^\varepsilon + 2\rho_t^\varepsilon [L\theta_t^\varepsilon + \Gamma(\log \rho_t^\varepsilon, \theta_t^\varepsilon)] = 0 \quad (2.18)$$

where $\nabla \theta_t^\varepsilon := \nabla(\psi_t^\varepsilon - \varphi_t^\varepsilon)/2$ is the current velocity that will be introduced at Proposition 2.2.4.

Examples 2.2.3 *We write explicitly these equations in two particular cases.*

- When the reference measure is associated to the generator $L^\varepsilon = \varepsilon \Delta/2$ and $d\mathbf{m} = \mathcal{L}$, the equations in (2.16) write as,

$$\begin{cases} \partial_t \varphi_t^\varepsilon - \varepsilon \Delta \varphi_t^\varepsilon / 2 - |\nabla \varphi_t^\varepsilon|^2 / 2 = 0, & 0 < t \leq 1, \\ \varphi_0^\varepsilon = \varepsilon \log f_\varepsilon, & t = 0 \end{cases}$$

and,

$$\begin{cases} \partial_t \psi_t^\varepsilon + \varepsilon \Delta \psi_t^\varepsilon / 2 + |\nabla \psi_t^\varepsilon|^2 / 2 = 0, & 0 \leq t < 1, \\ \psi_1^\varepsilon = \varepsilon \log g_\varepsilon, & t = 1. \end{cases}$$

and equation (2.18) is the transport equation,

$$\partial_t \rho_t^\varepsilon + \nabla \cdot (\rho_t^\varepsilon \nabla \theta_t^\varepsilon) = 0.$$

Note that in this case $(\mu_t)_{0 \leq t \leq 1} = (\rho_t)_{0 \leq t \leq 1}$ since the reversing measure m coincides with the Lebesgue measure \mathcal{L} .

- Whereas in the general case with generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ and $d\mathbf{m} = e^{-V} dx$, equation (2.18) writes as,

$$\partial_t \rho_t^\varepsilon + \nabla \cdot (\rho_t^\varepsilon \nabla \theta_t^\varepsilon) - \rho_t^\varepsilon \nabla V \cdot \nabla \theta_t^\varepsilon = 0$$

or equivalently,

$$\partial_t \rho_t^\varepsilon + e^V \nabla \cdot (e^{-V} \rho_t^\varepsilon \nabla \theta_t^\varepsilon) = 0.$$

We can deduce from (2.15) that the density with respect to the Lebesgue measure,

$$\mu_t^\varepsilon = e^{-V} \rho_t^\varepsilon$$

satisfies the transport equation,

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (\mu_t^\varepsilon \nabla \theta_t^\varepsilon) = 0. \quad (2.19)$$

Beyond these algebraic computations, we can derive the kinematics of the entropic bridge P^ε from a probabilistic point of view. In the next proposition we identify the Nelson velocities introduced at Section 2.2 for the entropic bridge $P^\varepsilon \in \mathcal{P}(\Omega)$.

Proposition 2.2.4 *Let $P^\varepsilon \in \mathcal{P}(\Omega)$ be the entropic bridge between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$. The forward and backward velocities associated with P^ε are respectively,*

$$\vec{v}_t^{P^\varepsilon} = -\frac{\varepsilon}{2} \nabla V + \nabla \psi_t^\varepsilon, \quad \overleftarrow{v}_t^{P^\varepsilon} = -\frac{\varepsilon}{2} \nabla V + \nabla \varphi_t^\varepsilon, \quad (2.20)$$

where ψ_t^ε and φ_t^ε are defined at (2.14). And the current and osmotic velocities are respectively,

$$v_t^{cu, P^\varepsilon} = \nabla \theta_t^\varepsilon = \frac{\nabla \psi_t^\varepsilon - \nabla \varphi_t^\varepsilon}{2}, \quad v_t^{os, P^\varepsilon} = \varepsilon \nabla \log \sqrt{\mu_t^\varepsilon}. \quad (2.21)$$

Proof

◁ We recall that as reference measure we consider the Markov and reversible $R^\varepsilon \in \mathcal{P}(\Omega)$ associated with generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$.

In the rest of the proof we use the notations,

$$DX := X_{t+h} - X_t \quad \text{and} \quad g_t := E_R(g(X_1)|X_t) = T_{1-t}g$$

where g appears at the Schrödinger system (2.6), and we drop the index ε . By definition of entropic bridge we have,

$$\begin{aligned} E_P(DX|X_t) &= E_R\left(\frac{dP}{dR}DX\Big|X_t\right) \frac{1}{E_R\left(\frac{dP}{dR}\Big|X_t\right)} \\ &= \frac{E_R(f(X_0)g(X_1)DX|X_t)}{E_R(f(X_0)g(X_1)|X_t)}, \end{aligned}$$

where in the last equality we applied Theorem 2.1.6. By the Markov property of the reference measure R , the past and the future are independent, then

$$\begin{aligned} E_P(DX|X_t) &= \frac{E_R(f(X_0)|X_t)E_R(g(X_1)DX|X_t)}{E_R(f(X_0)|X_t)E_R(g(X_1)|X_t)} \\ &= \frac{E_R(g(X_1)DX|X_t)}{g_t}. \end{aligned}$$

Again, by the Markov property,

$$E_R(g(X_1)DX|X_t) = E_R(g(X_1)DX|X_{[0,t]}) = E_R(E_R(g(X_1)DX|X_{[0,t+hx]})|X_{[0,t]}).$$

Moreover, by writing $Dg = g_{t+h} - g_t$,

$$\begin{aligned} E_P(DX|X_t) &= \frac{E_R((g_t + Dg)DX|X_t)}{g_t} \\ &= E_R(DX|X_t) + \frac{E_R(DGD X|X_t)}{g_t}. \end{aligned}$$

By definition of forward velocity (Definition 1.1.9 (i)),

$$\begin{aligned} \vec{v}_t^P &= \lim_{h \rightarrow 0^+} E_P\left(\frac{DX}{h} | X_t\right) \\ &= \lim_{h \rightarrow 0^+} E_R\left(\frac{DX}{h} | X_t\right) + \lim_{h \rightarrow 0^+} \frac{E_R(DGD X|X_t)}{hg_t} \\ &\stackrel{(2.27)}{=} \vec{v}_t^R + 2\frac{\Gamma(g_t, x_t)}{g_t} \\ &= -\frac{\nabla V}{2} + \varepsilon \nabla \log g_t \\ &= -\frac{\nabla V}{2} + \nabla \psi_t. \end{aligned}$$

Similar computations show that $\overleftarrow{v}_t^{P^\varepsilon} = -\varepsilon \frac{\nabla V}{2} + \nabla \varphi_t^\varepsilon$.

It remains to identify the current and osmotic velocities. By Definition 1.1.9 we have, for any $0 \leq t \leq 1$,

$$v_t^{cu,P} = \frac{\vec{v}_t^P - \overleftarrow{v}_t^P}{2} = \frac{\nabla \psi_t^\varepsilon - \nabla \varphi_t^\varepsilon}{2}.$$

Thus we recognize that the current velocity is the vector field that appears in the transport equation $v_t^{cu,P} = \nabla \theta_t$. For what concerns the osmotic velocity, we have for any $0 \leq t \leq 1$,

$$\begin{aligned} v_t^{os,P} = \frac{\vec{v}_t^P + \overleftarrow{v}_t^P}{2} &= -\frac{\varepsilon}{2} \nabla V + \frac{1}{2} \nabla (\psi_t^\varepsilon + \varphi_t^\varepsilon) \\ &= -\frac{\varepsilon}{2} (\nabla V - \nabla \log \rho_t^\varepsilon) = \frac{\varepsilon}{2} \nabla \log (\rho_t^\varepsilon e^{-V}) \\ &\stackrel{(2.15)}{=} \varepsilon \nabla \log \sqrt{\mu_t^\varepsilon} \end{aligned} \tag{2.22}$$

recovering identity (1.25). \triangleright

Remark 2.2.5 *Note that in the proof we used the probabilistic definition of the Γ -operator, that we recall here briefly. For any suitable function u , Ito's formula gives*

$$du(X_t) = Lu(X_t) + \nabla u(X_t) \cdot dW_t^P.$$

Multiply this equation for a suitable function v ,

$$vdu(X_t) = vLu(X_t) + v\nabla u(X_t) \cdot dW_t^P. \quad (2.23)$$

We can switch the role of u and v ,

$$udv(X_t) = uLv(X_t) + u\nabla v(X_t) \cdot dW_t^P. \quad (2.24)$$

And finally consider Ito's formula for the product function uv ,

$$d(uv)(X_t) = L(uv)(X_t) + \nabla(uv)(X_t) \cdot dW_t^P. \quad (2.25)$$

We subtract (2.23) and (2.24) from (2.25), and, according to Definition 1.1.21, get

$$d(uv) - vdu - udv = L(uv) - vLu - uLv = 2\Gamma(u, v)dt.$$

Hence,

$$\Gamma(u, v)dt = d(uv) - vdu - udv. \quad (2.26)$$

Applying Ito's formula to $u(X_t) = u_t$ and $v(X_t) = v_t$,

$$d(u_t v_t) = u_t dv_t + v_t du_t + dv_t du_t.$$

This together with (2.26) gives,

$$\Gamma(u, v)dt = du(X_t)dv(X_t). \quad (2.27)$$

Note also that the limit in the proof is formal, since one has to make sure that the relative entropy is finite.

Remark 2.2.6 *We can derive as well the relation between current and osmotic velocities of the entropic bridge P^ε with those of the reference path measure R^ε . According to Examples 1.1.13 we have*

$$v^{cu, R^\varepsilon} = 0, \quad v^{os, R^\varepsilon} = -\frac{\varepsilon}{2} \nabla V.$$

Hence,

$$\begin{aligned} v_t^{cu, P^\varepsilon} - v_t^{cu, R^\varepsilon} &= \nabla \theta_t^\varepsilon; \\ v_t^{os, P^\varepsilon} - v_t^{os, R^\varepsilon} &= \frac{\varepsilon}{2} [\nabla \log \mu_t^\varepsilon + \nabla V] \\ &= \frac{\varepsilon}{2} \nabla \log(\mu_t^\varepsilon e^V) \\ &\stackrel{(2.15)}{=} \frac{\varepsilon}{2} \nabla \log \rho_t^\varepsilon. \end{aligned}$$

Let us see now which are the relations between the Nelson velocities of the the entropic Bridge P and its time reversed $P^* \in \mathcal{P}(\Omega)$, that we recall is $P^* = (X^*)_\# P$, where $X_t^* = X_{1-t}$ for all $t \in [0, 1]$, i.e.

$$E_{P^*} f(X_t^*) = E_P f(X_{1-t}).$$

Lemma 2.2.7 (Time reversal) *Given an entropic bridge P and its time reversed P^* , the forward velocity of the former and the backward velocity of the latter are related by,*

$$\vec{v}_{1-t}^{P^*}(z) = \overleftarrow{v}_t^P(z), \quad \overleftarrow{v}_{1-t}^{P^*}(z) = \vec{v}_t^P(z), \quad (2.28)$$

for any $0 \leq t \leq 1$.

Proof

◁ It follows directly by the definitions that,

$$\begin{aligned} \vec{v}_{1-t}^{P^*}(z) &= \lim_{h \rightarrow 0^+} E_{P^*} \left(\frac{X_{1-t+h} - X_{1-t}}{h} \middle| X_t = z \right) \\ &= \lim_{h \rightarrow 0^+} E_{\hat{P}} \left(\frac{X_{t-h} - X_t}{h} \middle| X_t = z \right) \\ &= \overleftarrow{v}_t^P(z). \end{aligned}$$

A similar computation shows that $\overleftarrow{v}_{1-t}^{P^*}(z) = \vec{v}_t^P(z)$. ▷

The stochastic equation associated to the entropic bridge P^ε is

$$\begin{aligned} dX_t &= \left(-\frac{\varepsilon}{2} \nabla V + \nabla \psi_t^\varepsilon \right) dt + \sqrt{\varepsilon} dW_t, \quad t > 0 \\ \mathcal{L}(X_0) &= \mu_0 \end{aligned} \quad (2.29)$$

This implies that for any regular and integrable enough test function u and any $0 \leq t \leq 1$,

$$\begin{aligned} \int_{\mathbb{R}^n} u(z) \mu_t^\varepsilon(z) dz &= E_{P^\varepsilon} u(X_0) + E_{P^\varepsilon} \int_0^t (\vec{v}_s^{P^\varepsilon} \cdot \nabla u + \varepsilon \Delta u / 2)(X_s) ds \\ &= \int_{\mathbb{R}^n} u(z) \mu_0(z) dz + \int_0^t ds \int_{\mathbb{R}^n} (\vec{v}_s^{P^\varepsilon} \cdot \nabla u + \varepsilon \Delta u / 2)(z) \mu_s(z) dz. \end{aligned}$$

After integration by parts, it yields the Fokker Planck equation

$$\partial_t \mu_t^\varepsilon + \nabla \cdot \left(\mu_t^\varepsilon \left[\vec{v}_t^{\hat{P}} - \frac{\varepsilon}{2} \nabla \log \mu_t^\varepsilon \right] \right) = 0,$$

and replacing the forward velocity by its value (2.20),

$$\partial_t \mu_t^\varepsilon + \nabla \cdot \left(\mu_t^\varepsilon \left[-\frac{\varepsilon}{2} \nabla V + \nabla \psi_t^\varepsilon - \frac{\varepsilon}{2} \nabla \log \mu_t^\varepsilon \right] \right) = 0. \quad (2.30)$$

Writing (2.30) for the time reversed $P^{\varepsilon,*} \in \mathcal{P}(\Omega)$, together with (2.20) and (2.28), gives

$$\partial_t \mu_t^{\varepsilon,*} + \nabla \cdot \left(\mu_t^{\varepsilon,*} \left[-\frac{\varepsilon}{2} \nabla V + \nabla \varphi_t^\varepsilon - \frac{\varepsilon}{2} \nabla \log \mu_t^{\varepsilon,*} \right] \right) = 0, \quad (2.31)$$

where we denoted $\mu_t^*(z) = dP_t^*/dz = dP_{1-t}^\varepsilon/dz = \mu_{1-t}(z)$. Since, $\partial_t \mu_{1-t}^* = -\partial_t \mu_t(z)$, (2.31) writes as,

$$\partial_t \mu_t^\varepsilon + \nabla \cdot \left(\mu_t^\varepsilon \left[\frac{\varepsilon}{2} \nabla V - \nabla \varphi_t^\varepsilon + \frac{\varepsilon}{2} \nabla \log \mu_t^\varepsilon \right] \right) = 0. \quad (2.32)$$

Taking the half sum of (2.30) and (2.32), we obtain,

$$\partial_t \mu_t^\varepsilon + \nabla \cdot (\mu_t^\varepsilon \nabla \theta_t^\varepsilon) = 0,$$

recovering the transport equation (2.19).

In Corollary 1.2.8 we showed that the kinetic energy is conserved along displacement interpolations and actually coincides with W_2^2 . We show here an analogous result along entropic interpolations.

Lemma 2.2.8 (Conserved quantity) *For any entropic interpolation $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$,*

$$C^\varepsilon = \int |\nabla \theta_t^\varepsilon|^2 - \frac{\varepsilon^2}{4} |\nabla \log \rho_t^\varepsilon|^2 d\mu_t = 2 \int \nabla \psi_t^\varepsilon \cdot \nabla \varphi_t^\varepsilon d\mu_t \quad (2.33)$$

is constant for each $t \geq 0$.

Proof

◁ We drop the index ε to simplify the notation. We start to prove that

$$\int \Gamma(\psi_t, \varphi_t) d\mu_t = \int \Gamma(f_t, g_t) dm, \quad (2.34)$$

where

$$\begin{cases} f_t := T_t f \\ g_t := T_{1-t} g \end{cases} \quad (2.35)$$

with f, g solutions of the Schrödinger system (2.6). Then,

$$\begin{aligned} \int \Gamma(\varphi_t, \psi_t) d\mu_t &= \int \Gamma(\log f_t, \log g_t) d\mu_t \\ &\stackrel{*}{=} \int \frac{1}{f_t g_t} \Gamma(f_t, g_t) d\mu_t = \int \Gamma(f_t, g_t) dm, \end{aligned}$$

where the marked equality is due to property *v.* in Proposition 1.1.22, and the last equality to $d\mu_t = f_t g_t dm$ as in (2.12).

We show that (2.34) is constant. From property *vii.* in Proposition 1.1.22 and (2.35) it follows,

$$\begin{aligned} \frac{d}{dt} \int \Gamma(f_t, g_t) dm &= \int \Gamma(\partial_t f_t, g_t) + \Gamma(f_t, \partial_t g_t) dm \\ &= \int \Gamma(Lf_t, g_t) - \Gamma(f_t, Lg_t) dm = 0, \end{aligned} \quad (2.36)$$

and the last equality is given by the symmetry property *xi.* in Proposition 1.1.22.

Finally in order to show that the constant quantity C^ε is (2.33) we recall the relations between Nelson's velocities in Proposition 2.2.4, and property *viii.* in Proposition 1.1.22, to obtain

$$\begin{aligned} 2 \int \Gamma(\psi_t, \varphi_t) d\mu_t &= 2 \int \Gamma(\theta_t + \varepsilon \log \sqrt{\rho_t}, \theta_t - \varepsilon \log \sqrt{\rho_t}) d\mu_t \\ &= 2 \int \Gamma(\theta_t) - \varepsilon^2 \Gamma(\log \sqrt{\rho_t}) d\mu_t \\ &= \int |\nabla \theta_t|^2 - \frac{\varepsilon^2}{4} |\nabla \log \rho_t|^2 d\mu_t, \end{aligned}$$

that is the claimed result. ▷

Remark 2.2.9 *Note that Lemma 2.2.8 is a particular case of a more general result. We introduce the Γ_n operator defined via the iterative formula for $n = 0, 1, 2, \dots$*

$$\Gamma_{n+1} = \frac{1}{2} [L\Gamma_n(f, g) - \Gamma_n(Lf, g) - \Gamma_n(f, Lg)]$$

with initial condition $\Gamma_0 = fg$, recovering Definition 1.1.21 for $n = 1, 2$. It can be proved with the same arguments as in (2.36) that for any $n \geq 0$, the function

$$[0, 1] \ni t \mapsto \int \Gamma_n(f_t^\varepsilon, g_t^\varepsilon) dm$$

is constant.

2.3 Entropic cost

Provided that an entropic bridge between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ exists, it attains the minimal value, that is

$$H(\hat{P}|R) = \inf\{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\}.$$

We could think about the value $H(\hat{P}|R)$ as a *cost* of transporting μ_0 into μ_1 according to the minimization problem (2.1). However, despite it is the most natural and intuitive way to define such a cost, it might not be the most meaningful, since for instance, it might be negative. We discuss here the different possible definitions of entropic cost with their properties.

Definition 2.3.1 (Forward and backward entropic costs) *Given two probability measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, such that there exists an entropic bridge $\hat{P} \in \mathcal{P}(\Omega)$, we define, the forward entropic cost as,*

$$\vec{\mathcal{A}}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \varepsilon H(\mu_0|m) \quad (2.37)$$

and the backward entropic cost as,

$$\overleftarrow{\mathcal{A}}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \varepsilon H(\mu_1|m).$$

The interest of considering $\vec{\mathcal{A}}$ and $\overleftarrow{\mathcal{A}}$ is that they are non negative quantities, and they take into account only the contribution of the transportation, of the dynamics. However they are not symmetric, since reversing μ_0 with μ_1 , the value of the constant term $H(\mu_i|m)$, for μ_0, μ_1 changes, hence also the value of costs.

To enforce symmetry, but still taking into account only the demand of transportation, we introduce the next definition.

Definition 2.3.2 (Entropic cost) *Given two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, such that there exists an entropic bridge $P^\varepsilon \in \mathcal{P}(\Omega)$, we define the entropic cost as,*

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \frac{\varepsilon}{2}[H(\mu_0|m) + H(\mu_1|m)].$$

The entropic cost \mathcal{A} is symmetric $\mathcal{A}(\mu, \nu) = \mathcal{A}(\nu, \mu)$ since R is reversible. To see this, let us denote $X_t^* = X_{1-t}$, $0 \leq t \leq 1$, and $Q^* := (X^*)_{\#}Q$ the time reversal of any $Q \in \mathcal{M}_+(\Omega)$. As X^* is one-to-one, we have $H(P|R) = H(P^*|R^*)$ and since we assume that $R^* = R$, we see that

$$H(P|R) = H(P^*|R), \quad \forall P \in \mathcal{P}(\Omega).$$

Hence, if P solves (2.1) with $(\mu_0, \mu_1) = (\mu, \nu)$, then $X^*_{\#}P$ solves (2.1) with $(\mu_0, \mu_1) = (\nu, \mu)$ and these Schrödinger problems share the same value. For what concerns the

constant term $[H(\mu|m) + H(\nu|m)]$ it is evident that if we exchange the role of μ and ν it remains unchanged.

Thus \mathcal{A} seems to be the best candidate to define a cost between two measures, having in mind the idea that we have no hope to define the square of an *entropic distance* between μ_0 and μ_1 , in analogy to W_2^2 . Indeed there is no reason why $\mathcal{A}(\mu, \mu) = 0$ for any $\mu \in \mathcal{P}(\mathbb{R}^n)$ (and this is true also for $\overrightarrow{\mathcal{A}}$ and $\overleftarrow{\mathcal{A}}$). This last definition is useful when we want to look at the dynamic Eulerian point of view. We will prove at Section 3.5 that the entropic cost \mathcal{A} admits a Benamou-Brenier formulation, in analogy to the Wasserstein distance (Theorem 1.2.12). In order to motivate Definition 2.3.2 we anticipate here some arguments of Section 3.5.

Let $P \in \mathcal{P}(\Omega)$ be such that $H(P|R^\varepsilon) < \infty$, by the Girsanov Theorem 1.1.14 there exists a vector field $\overrightarrow{\beta}_t$ such that

$$E_P \int_0^1 |\overrightarrow{\beta}_t|^2 / 2 dt < \infty,$$

the forward velocity of P is given by $\overrightarrow{v}_t^P = \varepsilon \overrightarrow{\beta}_t - \varepsilon \nabla V / 2$ and the relative entropy can be written as,

$$H(P|R^\varepsilon) = H(P_0|m) + \varepsilon E_P \int_0^1 |\overrightarrow{\beta}_t|^2 / 2 dt.$$

Since R^ε is reversible, we obtain by reversing time, analogously

$$H(P|R^\varepsilon) = H(P_1|m) + \varepsilon E_P \int_0^1 |\overleftarrow{\beta}_t|^2 / 2 dt,$$

where $\overleftarrow{\beta}_t$ appears in the backward velocity of P , that is $\overleftarrow{v}_t^P = \varepsilon \overleftarrow{\beta}_t - \varepsilon \nabla V / 2$. Furthermore by definition of current and osmotic velocities (Definition 1.1.9),

$$v^{cu,P} = \frac{\varepsilon \overrightarrow{\beta}_t - \varepsilon \overleftarrow{\beta}_t}{2} =: \varepsilon \beta^{cu}, \quad v^{os,P} = \frac{\varepsilon \overrightarrow{\beta}_t + \varepsilon \overleftarrow{\beta}_t}{2} - \frac{\varepsilon}{2} \nabla V =: \varepsilon \beta^{os} - \varepsilon \nabla V / 2.$$

We define here the *forward* and *backward entropic actions* of P with respect to R^ε by

$$\overrightarrow{\mathcal{A}}(P|R^\varepsilon) := H(P|R^{P_0,\varepsilon}) = \varepsilon E_P \int_0^1 |\overrightarrow{\beta}_t|^2 dt$$

$$\overleftarrow{\mathcal{A}}(P|R^\varepsilon) := H(P|R^{P_1,\varepsilon}) = \varepsilon E_P \int_0^1 |\overleftarrow{\beta}_t|^2 dt,$$

where

$$R^{P_0,\varepsilon}(\cdot) := \int_{\mathbb{R}^n} R^\varepsilon(\cdot|X_0 = x) P_0(dx), \quad R^{P_1,\varepsilon}(\cdot) := \int_{\mathbb{R}^n} R^\varepsilon(\cdot|X_1 = y) P_1(dy),$$

for which it holds,

$$H(P|R^{P_0,\varepsilon}) = H(P|R^\varepsilon) - H(P_0|R_0^\varepsilon), \quad H(P|R^{P_1,\varepsilon}) = H(P|R^\varepsilon) - H(P_1|R_1^\varepsilon).$$

Similarly, we define the *current* and *osmotic actions* as,

$$A^{cu}(P|R^\varepsilon) = \varepsilon E_P \int_0^1 |\beta^{cu}|^2 / 2 dt,$$

$$A^{os}(P|R^\varepsilon) = \varepsilon E_P \int_0^1 |\beta^{os}|^2 / 2 dt.$$

It follows that

$$A^{cu}(P|R^\varepsilon) = \varepsilon^{-1} \int_0^1 \int_{\mathbb{R}^n} |v_t^{cu}|^2 / (2) dt d\mu_t,$$

$$A^{os}(P|R^\varepsilon) = \frac{\varepsilon}{2} \int_0^1 \int_{\mathbb{R}^n} |\nabla \log \mu_t + \nabla V / 2|^2 dt d\mu_t.$$

Taking the half sum of the two expressions in (2.3) together with the parallelogram identity (1.19), leads us to,

$$H(P|R^\varepsilon) = \frac{1}{2} [H(P_0|R_0^\varepsilon) + H(P_1|R_1^\varepsilon)] + \varepsilon E_P \int_0^1 |\beta_t^{cu}|^2 + |\beta_t^{os}|^2 dt. \quad (2.38)$$

Let us introduce the last definition of the *entropic action*,

$$A(P|R^\varepsilon) := A^{cu}(P|R^\varepsilon) + A^{os}(P|R^\varepsilon) = [\vec{A}(P|R^\varepsilon) + \overleftarrow{A}(P|R^\varepsilon)] / 2, \quad (2.39)$$

and write (2.38) as,

$$H(P|R^\varepsilon) = \frac{1}{2} [H(P_0|R_0^\varepsilon) + H(P_1|R_1^\varepsilon)] + A(P|R^\varepsilon). \quad (2.40)$$

It is straightforward that the Schrödinger problem (2.1) is equivalent, up to an additive constant, to

$$\inf \{ \varepsilon A(P|R^\varepsilon); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \},$$

and it follows in particular that its minimal value is,

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf \{ H(P|R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1 \} - \frac{\varepsilon}{2} [H(\mu_0|m) + H(\mu_1|m)].$$

In the subsequent chapters we will deal both with the forward and entropic cost. The advantage of the first is that it has a duality formulation that will be proved at Section 3.4, while the advantage of the second is the Benamou-Brenier formulation proved at Section 3.5. We will choose time to time which is the most convenient definition to adopt. When we will use the two different entropic costs $\vec{\mathcal{A}}$ and \mathcal{A} in the same expression, it might seem confusing but keep in mind that they differ just for a constant term, that is

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \vec{\mathcal{A}}^\varepsilon(\mu_0, \mu_1) + \frac{\varepsilon}{2} [H(\mu_0|\mathcal{L}) - H(\mu_1|\mathcal{L})].$$

Remark 2.3.3 As a consequence of (2.40), we deduce that the conserved quantity at Lemma 2.2.8 is finite, indeed,

$$|C^\varepsilon| < 2H(P^\varepsilon|R^\varepsilon) < \infty,$$

where \hat{P}^ε is an entropic bridge.

2.4 Limit for $\varepsilon \rightarrow 0$

In this section we explore the relation between the Schrödinger problem and the Monge-Kantorovich problem. We will see that the normalized entropic cost, no matter which of the previous definitions we choose, converges to the minimal value of the Monge-Kantorovich problem, whenever the reference path measure $R^\varepsilon \in \mathcal{P}(\Omega)$ of the former, is properly related to the cost function of the latter $C : \Omega \rightarrow \mathbb{R}^+$ in (1.37). In particular, we need that $(R^\varepsilon)_{\varepsilon>0}$ satisfies a *large deviation principle* with scale $1/\varepsilon$, as ε tends to zero, and rate function C , that is,

$$R^\varepsilon(A) \underset{\varepsilon \rightarrow 0^+}{\asymp} e^{-\inf_{\omega \in A} C(\omega)/\varepsilon}$$

for any measurable subset $A \subset \Omega$. The paper [Léo12a] of Léonard gives the proof of this convergence result in all generality. Let us first give an idea of how the family of Schrödinger problems with reference measure R^ε associated to the Brownian motion is related to the quadratic Wasserstein distance.

Indeed when $R_{01}^\varepsilon \in \mathcal{M}_+(\mathbb{R}^n \times \mathbb{R}^n)$ and $c : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ are given respectively by

$$R_{01}^\varepsilon(dxdy) = e^{-\frac{|x-y|^2}{2\varepsilon}} / (2\pi\varepsilon)^{n/2} dxdy, \quad c(x, y) = \frac{|x-y|^2}{2},$$

we can compute explicitly the relative entropy for any $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ as,

$$\begin{aligned} \varepsilon H(\pi | R_{01}^\varepsilon) &= \varepsilon \int \log \frac{d\pi}{dR_{01}^\varepsilon} d\pi \\ &= \varepsilon \int \log \frac{d\pi}{dxdy} d\pi - \varepsilon \int \log \left(e^{-\frac{|x-y|^2}{2\varepsilon}} / (2\pi\varepsilon)^{n/2} \right) d\pi \\ &= \varepsilon \int \log \frac{d\pi}{dxdy} d\pi + \int \frac{|x-y|^2}{2} d\pi + \frac{n\varepsilon}{2} \log(2\pi\varepsilon). \end{aligned}$$

Taking limit for $\varepsilon \rightarrow 0$, when $H(\pi | dxdy)$ is finite, shows that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon H(\pi | R_{01}^\varepsilon) = \int \frac{|x-y|^2}{2} d\pi.$$

And taking the infimum over all $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ satisfying the marginal constraints, we obtain

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{W_2^2(\mu_0, \mu_1)}{2}. \quad (2.41)$$

Remark 2.4.1 Note that either $\overrightarrow{\mathcal{A}}^\varepsilon$ or $\overleftarrow{\mathcal{A}}^\varepsilon$ or \mathcal{A}^ε defined at Definitions 2.3.1 and 2.3.2 can play the role of the entropic cost \mathcal{A}^ε in this theorem, since they differ by some constant quantity times ε , namely $\varepsilon H(\mu_{0,1} | m)$ that vanishes at the limit.

Despite formal, these computations give an enlightening intuition on how the entropy minimization problem (2.1) can be actually an approximation of a linear deterministic minimization problem as (1.37), without entering in the heavy machinery of large deviation theory and Γ -convergence.

The relation between the Schrödinger problem with reference measure $(R^\varepsilon)_{\varepsilon>0}$ with generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ to the quadratic Wasserstein distance, is due to the fact that they satisfy a large deviation principle. We recall here that a sequence $(p_\varepsilon)_{\varepsilon>0}$ of probability measures satisfies the large deviation principle with scale $1/\varepsilon$ and rate function I , if for each Borel measurable set A we have,

$$-\inf_{x \in \text{int}A} I(x) \leq \liminf_{\varepsilon \rightarrow 0} \varepsilon \log p_\varepsilon(A) \leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \log p_\varepsilon(A) \leq -\inf_{x \in \text{cl}A} I(x)$$

where $\text{int}A$ and $\text{cl}A$ are respectively the interior and the closure of the set A .

We recall that when $V = 0$ and $R^\varepsilon \in \mathcal{M}_+(\Omega)$ is associated to the SDE,

$$Y_t^{\varepsilon,x} = x + \sqrt{\varepsilon}W_t, \quad 0 \leq t \leq 1.$$

where the initial condition $Y_0^{x,\varepsilon} = x$ is deterministic, Schilder's theorem makes sure that such a $(R^\varepsilon)_{\varepsilon>0}$ satisfies a large deviation principle with scale $1/\varepsilon$ and rate function,

$$C^x(\omega) = \begin{cases} \int_0^1 \frac{|\dot{\omega}_t|^2}{2} dt & \in [0, \infty], \quad \text{if } \omega \in \Omega \text{ s. t. } \omega_0 = x, \text{ and } \omega \text{ is abs.cont.} \\ +\infty & \text{otherwise.} \end{cases} \quad (2.42)$$

Theorem 2.4.2 (Schilder's theorem) *The sequence of stochastic processes (2.4) satisfies a large deviation principle in $\Omega = C([0, 1], \mathbb{R}^n)$ equipped with the topology of uniform convergence with scale $1/\varepsilon$ and rate function (2.42).*

For the proof see [DZ98, Thm. 5.2.3]. This result was extended in the Freidlin-Wentzell Theorem, which states that the solution of the SDE (1.12),

$$dX_t = b(X_t)dt + \sqrt{\varepsilon}dW_t, \quad \forall 0 \leq t \leq 1, \quad X_0 = 0$$

satisfies a large deviation principle as ε tends to zero with rate function

$$I(\omega) = \frac{1}{2} \int |\dot{\omega}_t - b(\omega_t)|^2 dt,$$

see [DZ98, Thm. 5.6.3]. Since we consider the drift $b = -\varepsilon \nabla V/2$, it can be proved through exponential equivalence, that the rate function $I(\omega)$ coincides with the quadratic cost (2.42) as ε tends to zero.

Theorem 2.4.3 (Convergence of \mathcal{A}^ε to $W_2^2/2$) *Let R^ε have generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ and reversing measure $\mathbf{m} = e^{-V}$ and let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ have finite relative entropy with respect to \mathbf{m} . Then,*

$$\Gamma\text{-}\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \cdot) = \frac{W_2^2(\mu_0, \cdot)}{2}$$

in $\mathcal{P}(\mathbb{R}^n)$. Moreover, for any $\mu_1 \in \mathcal{P}(\mathbb{R}^n)$, there exists a sequence $\mu_1^\varepsilon \in \mathcal{P}(\mathbb{R}^n)$ such that $\lim_{\varepsilon \rightarrow 0} \mu_1^\varepsilon = \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, and

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \mu_1^\varepsilon) = \frac{W_2^2(\mu_0, \mu_1)}{2}.$$

If $\mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, then for any small enough $\varepsilon > 0$ the dynamic Schrödinger problem (2.1) admits a unique solution $\hat{P}^\varepsilon \in \mathcal{P}(\Omega)$ such that

$$\lim_{\varepsilon \rightarrow 0} \hat{P}^\varepsilon = \hat{P},$$

where $\hat{P} \in \mathcal{P}(\Omega)$ is the unique solution of the dynamic Monge-Kantorovich problem with quadratic cost, (1.40).

Let us give a flavor of the general arguments of the proof in [Léo12a]. We first recall that the Γ -limit of a sequence $(f_k)_{k \geq 1}$ of $(-\infty, \infty]$ -valued functions on \mathbb{R}^n , is given for all $x \in \mathbb{R}^n$ by,

$$\Gamma\text{-}\lim_{k \rightarrow \infty} f_k(x) = \sup_{V \in N(x)} \liminf_{k \rightarrow \infty} \inf_{y \in V} f_k(y)$$

where $N(x)$ is the set of all neighborhoods of x . This is equivalent to say that, for any sequence $(x_k)_{k \geq 1}$ such that $\lim_{k \rightarrow \infty} x_k = x$,

$$\liminf_{k \rightarrow \infty} f_k(x_k) \geq f(x),$$

and there exists a sequence $(\tilde{x}_k)_{k \geq 1}$ such that $\lim_{k \rightarrow \infty} \tilde{x}_k = x$, and

$$\limsup_{k \rightarrow \infty} f_k(\tilde{x}_k) \leq f(x).$$

The proof of Theorem 2.4.3 relies on the next preliminary result. First we introduce some useful notation. We denote,

$$R^{\varepsilon, \mu_0}(\cdot) := \int_{\mathbb{R}^n} R^{\varepsilon, x}(\cdot) \mu_0(dx), \quad \in \mathcal{P}(\Omega),$$

with $R^{\varepsilon, x}$ the law of the stochastic process (2.4), and for every $x \in \mathbb{R}^n$ and $A \subset \mathbb{R}^n$,

$$\iota_{x \in A} = \begin{cases} 0, & \text{if } x \in A \\ \infty & \text{otherwise.} \end{cases}$$

We also denote for any $P \in \mathcal{P}(\Omega)$,

$$\xi^{\varepsilon, \mu_0}(P) := \varepsilon H(P|R^{\varepsilon, \mu_0}) + \iota_{P_0 = \mu_0} \quad (2.43)$$

$$\xi^{\mu_0}(P) := \int_{\Omega} C dP + \iota_{P_0 = \mu_0}. \quad (2.44)$$

Observe that,

$$\begin{aligned} \mathcal{A}^\varepsilon(\mu_0, \nu) &= \inf \{ \xi^{\varepsilon, \mu_0}(P) : P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \nu \} \\ \frac{W_2^2(\mu_0, \nu)}{2} &= \inf \{ \xi^{\mu_0}(P) : P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \nu \}. \end{aligned}$$

Proposition 2.4.4 For any $x \in \mathbb{R}^n$ and $\mu_0 \in \mathcal{P}(\mathbb{R}^n)$, it holds for any $P \in \mathcal{P}(\Omega)$,

$$\Gamma\text{-}\lim_{\varepsilon \rightarrow 0^+} \varepsilon H(P|R^{\varepsilon, \mu_0}) + \iota_{P_0 = \mu_0} = \int_{\Omega} C dP + \iota_{P_0 = \mu_0}. \quad (2.45)$$

where R^{ε, μ_0} and C are given respectively at (2.4) and (2.42). Equivalently according to Definitions (2.43) and (2.44),

$$\Gamma\text{-}\lim_{\varepsilon \rightarrow 0^+} \xi^{\varepsilon, \mu_0} = \xi^{\mu_0} \in \mathcal{P}(\Omega).$$

Sketch of the Proof

◁ We give the recipe of the proof step by step.

Step1. Furnish the space $C_b(\Omega)$ with the norm $\|f\| = \sup_{\Omega} |f|$, $f \in C_b(\Omega)$ and denote $C_b(\Omega)'$ its topological dual space. Prove that for each $P \in C_b(\Omega)'$,

$$H(P|R^{\mu_0}) + \iota_{\{P_0 = \mu_0\}} = \sup_{f \in C_b(\Omega)} \left\{ \int_{\Omega} f dQ - \int_{\mathbb{R}^n} \log \left(\int_{\Omega} e^{f/\varepsilon} R^{\varepsilon, x}(dw) \right) \mu_0(dx) \right\}.$$

Compare this formula to the variational formula (1.2).

Step2. For any $f \in C_b(\Omega)$ denote

$$\Lambda(f) := \int_{\mathbb{R}^n} \sup_{\Omega} \{f - C^x\} \mu_0(dx) = \int_{\mathbb{R}^n} \sup_{\Omega^x} \{f - C\} \mu_0(dx),$$

where $\Omega^x := \{X_0 = x\} \subset \Omega$, and its convex conjugate by,

$$\Lambda^*(Q) := \sup_{f \in C_b(\Omega)} \left\{ \int_{\Omega} f dQ - \Lambda(f) \right\}.$$

Prove that for any $Q \in \mathcal{P}(\Omega)$,

$$\Lambda^*(Q) = \int_{\Omega} C dQ + \iota_{\{Q_0 = \mu_0\}},$$

or shortly, $\Lambda^*(Q) = \xi^{\mu_0}(Q)$, $\forall Q \in \mathcal{M}_b(\Omega)$.

Step3. Define, for any $f \in C_b(\Omega)$

$$\Lambda^\varepsilon(f) := \varepsilon \int_{\mathbb{R}^n} \log \left(\int_{\Omega} e^{f/\varepsilon} R^{\varepsilon, x}(d\omega) \right) \mu_0(dx).$$

Prove that

- $\lim_{\varepsilon \rightarrow 0} \Lambda^\varepsilon(f) = \Lambda(f)$;
- $\sup_{\varepsilon > 0} |\Lambda^\varepsilon(f)| \leq \|f\|$, $|\Lambda(f)| \leq \|f\| := \sup_{\Omega} |f|$;
- Λ, Λ^* are convex.

From the Laplace-Varadhan principle [DZ98, Thm. 4.3.1], it follows that for all $x \in \mathbb{R}^n$,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \log \int_{\Omega} e^{f/\varepsilon} R^{\varepsilon, x}(d\omega) = \sup_{\Omega} \{f - C^x\}.$$

Moreover, it is easy to see that $|\varepsilon \log \int_{\Omega} e^{f/\varepsilon} R^{\varepsilon, x}(d\omega)| \leq \|f\|_{\infty}$. Hence by the dominated convergence theorem, it follows that $\lim_{\varepsilon \rightarrow 0} \Lambda^\varepsilon(f) = \Lambda(f)$, for all $f \in C_b(\Omega)$.

Prove that

$$\Lambda^{\varepsilon, *}(Q) = \varepsilon H(P|R^{\varepsilon, \mu_0}) + \iota_{P_0 = \mu_0}.$$

Step4. From Step3. and [Léo12a, Cor.5.4]¹, deduce that

$$\lim_{\varepsilon \rightarrow 0} \Lambda^{\varepsilon, *}(Q) = \Lambda^*(Q)$$

From Step1. it follows that $\Lambda^{*, \varepsilon} = \xi^{\varepsilon, \mu_0}$, and from Step2. $\Lambda^* = \xi^{\mu_0}$. Hence, by Definitions (2.43) and (2.44), conclude the proof. \triangleright

Proof of Theorem 2.4.3

\triangleleft The proof of Theorem 2.4.3 is an immediate consequence of Proposition 2.4.4 and

¹It states that given a family of real valued convex functions $(g_k)_{k \geq 1}$ on a normed space $(Y, \|\cdot\|)$, such that g_k converges pointwise to g as $k \rightarrow \infty$ and there exists $c > 0$ such that $|g_k(y)| \leq c(1 + \|y\|)$ for all $y \in Y$ and $k \geq 1$, then $(g_k^*)_{k \geq 1}$ Γ -converges to g^* with respect to $\sigma(X, Y)$ where X is the topological dual space of Y .

[Léo12a, Thm. 6.1]. The last insures that, given the Γ -convergence (2.45), we have that the minimization problem

$$\inf \left\{ \varepsilon H(\pi | R_{01}^{\varepsilon, \mu_0}) + \iota_{\{P_0 = \mu_0\}} : \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : P_1 = \mu_1 \right\} \quad (2.46)$$

Γ -converges as ε tends to zero to

$$\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x - y|^2}{2} \pi(dx dy) + \iota_{\{P_0 = \mu_0\}} : \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : P_1 = \mu_1 \right\} \quad (2.47)$$

in $\mathcal{P}(\mathbb{R}^n)$. Moreover it states that for any $\mu_1 \in \mathcal{P}(\mathbb{R}^n)$, it exists a sequence $(\mu_1^\varepsilon)_{\varepsilon > 0}$ such that $\lim_{\varepsilon \rightarrow 0} \mu_1^\varepsilon = \mu_1$ and such that

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \inf \left\{ \varepsilon H(\pi | R_{01}^{\varepsilon, \mu_0}) + \iota_{\{P_0 = \mu_0\}} : \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : P_1 = \mu_1^\varepsilon \right\} = \\ \inf \left\{ \int \frac{|x - y|^2}{2} d\pi + \iota_{\{P_0 = \mu_0\}} : \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : P_1 = \mu_1 \right\}. \end{aligned} \quad (2.48)$$

Note that the Γ -convergence of (2.46) to (2.47) can be written equivalently as,

$$\Gamma\text{-}\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \cdot) = \frac{W_2^2(\mu_0, \cdot)}{2}$$

in $\mathcal{P}(\mathbb{R}^n)$. Hence (2.48) writes as,

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \mu_1^\varepsilon) = \frac{W_2^2(\mu_0, \mu_1)}{2}.$$

Moreover [Léo12a, Thm. 6.1] states that if $\mu_1 \in \mathcal{P}(\mathbb{R}^n)$ is such that $W_2^2(\mu_0, \mu_1) < \infty$, then $\mathcal{A}^\varepsilon(\mu_0, \mu_1^\varepsilon) < \infty$ for a small enough $\varepsilon > 0$, and it admits a minimizer $\hat{\pi}^\varepsilon$. The sequence $(\hat{\pi}^\varepsilon)_{\varepsilon > 0}$ admits at least a cluster point $\hat{\pi}$, minimizer of $W_2^2(\mu_0, \mu_1)$. \triangleright

As a consequence of Theorem 2.4.3, we can derive some convergence result for the current and osmotic actions.

Lemma 2.4.5 *Let $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$ be the entropic interpolation between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, then*

$$\lim_{\varepsilon \rightarrow 0^+} \int_0^1 \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon dt = W_2^2(\mu_0, \mu_1) \quad (2.49)$$

$$\lim_{\varepsilon \rightarrow 0^+} \varepsilon^2 \int_0^1 I(\mu_t^\varepsilon | \mathbf{m}) dt = \lim_{\varepsilon \rightarrow 0^+} \varepsilon^2 \int_0^1 \int |\nabla \log \rho_t^\varepsilon|^2 d\mu_t^\varepsilon dt = 0, \quad (2.50)$$

where $\nabla \theta_t^\varepsilon$ is the current velocity of the entropic bridge P^ε in (2.21) and $I(\mu_t | \mathbf{m})$ denotes the Fisher information $I(\rho | \nu) = \int |\nabla \log(d\rho/d\nu)|^2 d\rho$.

Proof

\triangleleft Let $P^\varepsilon \in \mathcal{P}(\Omega)$ be the entropic bridge between $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, then according to (2.39) it minimizes

$$\inf \{ \varepsilon A(P | R^\varepsilon); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1 \}$$

thus,

$$\varepsilon A(P^\varepsilon | R^\varepsilon) = \frac{1}{2} \int_0^1 \int_{\mathbb{R}^n} |v^{cu, P^\varepsilon}|^2 d\mu_t^\varepsilon dt + \frac{\varepsilon^2}{8} \int_0^1 I(\mu_t^\varepsilon | \mathbf{m}) dt$$

where $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$ is the entropic interpolation between μ_0 and μ_1 . Theorem 2.4.3 implies that

$$\lim_{\varepsilon \rightarrow 0^+} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \lim_{\varepsilon \rightarrow 0^+} \int_0^1 \int_{\mathbb{R}^n} |v^{cu, P^\varepsilon}|^2 d\mu_t^\varepsilon dt + \frac{\varepsilon^2}{8} \int_0^1 I(\mu_t^\varepsilon | \mathbf{m}) dt = W_2^2(\mu_0, \mu_1). \quad (2.51)$$

Since the continuity equation and the other assumptions in Theorem 1.2.12 are satisfied by $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$, the Benamou-Brenier formulation makes sure that for any $\varepsilon > 0$,

$$W_2^2(\mu_0, \mu_1) \leq \int_0^1 \int_{\mathbb{R}^n} |v_t^{cu, P^\varepsilon}|^2 d\mu_t^\varepsilon dt.$$

Together with limit (2.51), it implies the claimed limits (2.49) and (2.50). \triangleright

Lemma 2.4.6 *Let $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$ be the entropic interpolation between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$, then*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \int_0^1 t \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon dt &= \frac{W_2^2(\mu_0, \mu_1)}{2} \\ \lim_{\varepsilon \rightarrow 0^+} \varepsilon^2 \int_0^1 t I(\mu_t^\varepsilon | \mathbf{m}) dt &= \lim_{\varepsilon \rightarrow 0^+} \varepsilon^2 \int_0^1 t \int |\nabla \log \rho_t^\varepsilon|^2 d\mu_t^\varepsilon dt = 0, \end{aligned} \quad (2.52)$$

where $\nabla \theta_t^\varepsilon$ is the current velocity of the entropic bridge P^ε in (2.21) and $I(\mu_t | \mathbf{m})$ denotes the Fisher information.

Proof

\triangleleft The proof is a consequence of Lemma 2.4.5 and the conserved quantity at Lemma 2.2.8. Indeed since

$$C^\varepsilon = \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon - \frac{\varepsilon^2}{2} \int |\nabla \log \rho_t^\varepsilon|^2 d\mu_t^\varepsilon = \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon - \frac{\varepsilon^2}{2} I(\mu_t^\varepsilon | \mathbf{m}),$$

taking the limit

$$\lim_{\varepsilon \rightarrow 0^+} C^\varepsilon = \lim_{\varepsilon \rightarrow 0^+} \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon - \lim_{\varepsilon \rightarrow 0^+} \frac{\varepsilon^2}{2} I(\mu_t^\varepsilon | \mathbf{m}) = W_2^2(\mu_0, \mu_1),$$

where the two limits are given at (2.49) and (2.50). At the same time from (2.50) it follows,

$$0 \leq \lim_{\varepsilon \rightarrow 0^+} \int_0^1 t I(\mu_t^\varepsilon | \mathbf{m}) dt \leq \lim_{\varepsilon \rightarrow 0^+} \int_0^1 I(\mu_t^\varepsilon | \mathbf{m}) dt = 0 \quad (2.53)$$

that is (2.52). Finally,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \int_0^1 t \int |\nabla \theta_t^\varepsilon|^2 d\mu_t^\varepsilon &= \lim_{\varepsilon \rightarrow 0^+} \int_0^1 t C^\varepsilon + \lim_{\varepsilon \rightarrow 0^+} \int_0^1 t I(\mu_t^\varepsilon | \mathbf{m}) dt \\ &= \lim_{\varepsilon \rightarrow 0^+} C^\varepsilon \int_0^1 t dt = \frac{W_2^2(\mu_0, \mu_1)}{2}, \end{aligned}$$

where the first equality is a consequence of (2.53). \triangleright

Another convergence result is about the kinematics of the entropic bridge with velocity field $v_t^{\hat{P}^\varepsilon} = -\frac{\varepsilon}{2} \nabla V + \nabla \psi_t^\varepsilon$ introduced at Proposition 2.2.4 to the one associated with the McCann interpolation introduced at (1.46). Since in this convergence the potential V doesn't play any role, we can assume $V = 0$, that is $L^\varepsilon = \varepsilon \Delta / 2$.

Proposition 2.4.7 *The following limit holds true,*

$$\lim_{\varepsilon \rightarrow 0^+} \psi_t^\varepsilon = \theta_t$$

where θ_t is introduced at (1.46) and ψ_t^ε in (2.14).

Proof

◁ We recall that ψ_t^ε is defined as,

$$\psi_t^\varepsilon = \varepsilon \log T_{1-t} g =: \mathcal{Q} \psi_1^\varepsilon$$

where the semigroup \mathcal{Q} is defined for any suitable function f as $\mathcal{Q}_t f = \varepsilon \log T_t e^{f/\varepsilon}$ and $\psi_1^\varepsilon = \varepsilon \log g^\varepsilon$. Moreover as we proved in Proposition 2.2.2 it satisfies the Hamilton-Jacobi-Bellman equation for any $0 \leq t < 1$

$$\partial_t \psi_t^\varepsilon + \varepsilon \Delta \psi_t^\varepsilon / 2 + |\nabla \psi_t^\varepsilon|^2 / 2 = 0$$

with initial condition $\psi_1^\varepsilon = \varepsilon \log g_\varepsilon$. By the vanishing viscosity theory [Eva98] it follows that for $\varepsilon \rightarrow 0$, ψ_t^ε converges to the solution of the associated Hamilton-Jacobi equation $\partial_t \psi_t + |\nabla \psi_t|^2 / 2 = 0$ with condition

$$\psi_1 := \lim_{\varepsilon \rightarrow 0} \psi_1^\varepsilon. \quad (2.54)$$

It will be clear at Section 3.4 that the limit (2.54) is a consequence of the dual Kantorovich formulation for the entropic cost given at Theorem 3.4.1 and that in particular

$$\psi_1(x) = \inf_{y \in \mathbb{R}^n} \left\{ \varphi(y) - |y|^2 + \frac{|x - y|^2}{2} \right\} \quad (2.55)$$

where $\psi_0(y) := \varphi(y) - |y|^2$ is given at (1.46) where φ is the convex function of the Brenier Theorem 1.2.5. ▷

Remark 2.4.8 *The convergence of ψ_t^ε might be seen as well as the result of the Schilder Theorem 2.4.2. Indeed the Laplace-Varadhan principle [DZ98, Thm. 4.3.1] states that,*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \psi_t^\varepsilon(z) &= \sup \left\{ \psi_1(\omega_1) - \int_t^1 \frac{|\dot{\omega}_s|^2}{2} ds; \omega \in \Omega : \omega_t = z \right\} \\ &= \sup_{y \in \mathbb{R}^n} \left\{ -\frac{|y - z|^2}{2(1-t)} + \psi_1(y) \right\} \end{aligned}$$

where ψ_1 is given at (2.55). Again we recovered the solution of the Hamilton-Jacobi equation in (1.46).

2.5 Entropy along entropic interpolation

In this section we look at the evolution along entropic interpolations of two information functionals: the relative entropy and the Fisher Information introduced at Section 1.1.1. Let $(\mu_t^\varepsilon)_{t \in [0,1]}$ be the entropic interpolation between two measures $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ chosen

in order assumptions in Theorem 2.1.4 to be satisfied. In the rest of the Section, the reference measure is R^ε with generator

$$L^\varepsilon = \frac{\varepsilon}{2}(\Delta - \nabla V \cdot \nabla), \quad \text{and} \quad \Gamma^\varepsilon(f, g) = \frac{\varepsilon}{2} \nabla f \cdot \nabla g$$

In the following we will adopt the convention (2.11) to have a clearer and explicit notation. We recall that by definition, the relative entropy and the Fisher Information of the entropic interpolation with respect to the invariant measure \mathbf{m} are, respectively,

$$H(\mu_t^\varepsilon | \mathbf{m}) = \int \rho_t^\varepsilon \log \rho_t^\varepsilon d\mathbf{m}, \quad I(\mu_t^\varepsilon | \mathbf{m}) = \frac{\varepsilon}{2} \int |\nabla \log \rho_t^\varepsilon|^2 \rho_t^\varepsilon d\mathbf{m}$$

We recall also that the densities with respect to the Lebesgue and the reversing measure \mathbf{m} are respectively

$$\mu_t^\varepsilon = e^{\frac{\varphi_t^\varepsilon + \psi_t^\varepsilon}{\varepsilon}} e^{-V}, \quad \rho_t^\varepsilon = e^{\frac{\varphi_t^\varepsilon + \psi_t^\varepsilon}{\varepsilon}}$$

related by (2.15). The question of the behavior of the relative entropy along entropic interpolations was first addressed in [Léo17], in which the first and second derivative, that we compute in the next two Propositions, are studied. Here we won't make any assumption on the marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ in addition to those of Theorem 2.1.4 which assure the existence of entropic interpolations. Further assumptions are necessary to make sure that Γ_2 -computations are allowed. We will make them clear later at Chapter 4 when these results will be used.

Proposition 2.5.1 (First derivative of the entropy) *Let $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ and the reference measure $R^\varepsilon \in \mathcal{P}(\Omega)$ be such that hypothesis in Theorem 2.1.4 are satisfied. Then the first order derivative of the relative entropy along the entropic interpolation $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$, is*

$$\frac{d}{dt} H(\mu_t^\varepsilon | \mathbf{m}) = \frac{1}{\varepsilon} \int \Gamma \psi_t^\varepsilon - \Gamma \varphi_t^\varepsilon d\mu_t^\varepsilon = 2 \int \Gamma(\log \rho_t^\varepsilon, \theta_t^\varepsilon) d\mu_t^\varepsilon$$

where Γ is the carré du champs operator associated with the generator L of the reference measure R .

Proposition 2.5.2 (Second derivative of the entropy) *In the same setting as in Proposition 2.5.1, the second order derivative of the relative entropy along the entropic interpolation $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$, is*

$$\frac{d^2}{dt^2} H(\mu_t^\varepsilon | \mathbf{m}) = 2 \int \Gamma_2 \psi_t^\varepsilon + \Gamma_2 \varphi_t^\varepsilon d\mu_t^\varepsilon = \int 4\Gamma_2 \theta_t^\varepsilon + \varepsilon^2 \Gamma_2 \log \rho_t^\varepsilon d\mu_t^\varepsilon$$

where Γ_2 is the iterated carré du champ operator associated with the generator L of the reference measure R .

Proposition 2.5.3 (First derivative of the Fisher Information) *In the same setting as in Proposition 2.5.1, the first order derivative of the Fisher information along the entropic interpolation $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$, is*

$$\frac{d}{dt} I(\mu_t^\varepsilon | \mathbf{m}) = 2 \int \Gamma_2 \psi_t^\varepsilon - \Gamma_2 \varphi_t^\varepsilon d\mu_t^\varepsilon = 4\varepsilon \int \Gamma_2(\log \rho_t^\varepsilon, \theta_t^\varepsilon) d\mu_t^\varepsilon$$

where Γ_2 is the iterated carré du champ operator associated with the generator L of the reference measure R .

Remarks 2.5.4 *Let us do some formal considerations.*

1. *It is interesting to remark the formal analogy between (both the expressions of) the first derivative of the Fisher information in Proposition 2.5.3 and the first derivative of the entropy in Proposition 2.5.1.*

2. *The entropic interpolation can be seen somehow as a link between the heat flow and the McCann interpolation. Indeed consider equation (2.18),*

$$\dot{\rho}_t^\varepsilon = -2\rho_t^\varepsilon[L\theta_t^\varepsilon + \Gamma(\log \rho_t^\varepsilon, \theta_t^\varepsilon)]$$

for some θ_t^ε . If, formally, we assume $v^{cu} + v^{os} = 0$, that is equivalent to

$$\theta_t^\varepsilon = -\varepsilon \log \rho_t^\varepsilon / 2,$$

we get the heat equation $\dot{\rho}_t^\varepsilon = L^\varepsilon \rho_t^\varepsilon$. It is well known that in this case the Fisher information represents the opposite of the first derivative of the entropy along the heat flow. Indeed we recover this result, if we replace $\theta_t^\varepsilon = -\varepsilon \log \rho_t^\varepsilon / 2$ in the second expression in Proposition 2.5.1, and as a consequence, we get also, from Propositions 2.5.2 and 2.5.3 that, in this specific case, $d^2 H(\mu_t^\varepsilon | \mathbf{m}) / dt^2 = -dI(\mu_t^\varepsilon | \mathbf{m}) / dt$.

On the other hand by considering the case $\varepsilon = 0$, we get,

$$\dot{\rho}_t = -\rho_t[\Delta\theta_t + \Gamma(\log \rho_t, \theta_t)]$$

that is the transport equation satisfied by the McCann interpolation as stated in (1.46). Hence, with $\varepsilon = 0$, and replacing $\nabla\theta_t^{\varepsilon=0} = \nabla\theta_t$ also given at (1.46), we obtain in Propositions 2.5.1 and 2.5.2 the formal expression of the first and second order derivatives of the entropy along the McCann interpolation.

3. *It is interesting to look at the equation satisfied by the potential θ_t^ε , that can be easily verified,*

$$\dot{\theta}_t^\varepsilon = \Gamma\theta_t^\varepsilon - \frac{\varepsilon^2}{2}L \log \rho_t^\varepsilon - \frac{\varepsilon^2}{4}\Gamma \log \rho_t^\varepsilon.$$

If we consider the example $L = \Delta/2$, we recognize the Hamilton-Jacobi equation with the potential $\log \rho_t^\varepsilon$, [Eva98].

2.5.1 Proofs of the preceding results.

Proof of Proposition 2.5.1

◁ We drop the index ε to simplify the notation.

$$\begin{aligned} \frac{d}{dt}H(\mu_t | \mathbf{m}) &= \frac{d}{dt} \int \rho_t \log \rho_t d\mathbf{m} \\ &= \int \dot{\rho}_t (\log \rho_t + 1) d\mathbf{m} \\ &\stackrel{*}{=} \frac{1}{\varepsilon} \int (-\varepsilon L\psi_t - \Gamma\psi_t + \varepsilon L\varphi_t + \Gamma\varphi_t) \frac{1}{\varepsilon} (\psi_t + \varphi_t) d\mu_t \\ &= \frac{1}{\varepsilon} \underbrace{\int (\varphi_t + \psi_t)L(\varphi_t - \psi_t) d\mu_t}_{(*)} + \frac{1}{\varepsilon^2} \int (\Gamma\varphi_t - \Gamma\psi_t)(\varphi_t + \psi_t) d\mu_t \end{aligned} \quad (2.56)$$

where the star equality is given to (2.17). We integrate (*) by parts,

$$\begin{aligned}
(*) &= - \int \Gamma(\varphi_t - \psi_t, (\varphi_t + \psi_t) e^{\frac{\psi_t + \varphi_t}{\varepsilon}}) d\mathbf{m} \\
&\stackrel{*}{=} - \int \Gamma(\varphi_t - \psi_t, \varphi_t + \psi_t) \left(1 + \frac{1}{\varepsilon}(\varphi_t + \psi_t)\right) d\mu_t \\
&\stackrel{**}{=} - \int (\Gamma\varphi_t - \Gamma\psi_t) \left(1 + \frac{1}{\varepsilon}(\varphi_t + \psi_t)\right) d\mu_t \\
&= \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t - \frac{1}{\varepsilon} \int (\Gamma\varphi_t - \Gamma\psi_t)(\varphi_t + \psi_t) d\mu_t.
\end{aligned}$$

Where in the the star and double star equality we used respectively properties (v.) and (viii.) of Proposition 1.1.22. If we replace it in (2.56) we have,

$$\begin{aligned}
\frac{d}{dt} H(\mu_t | \mathbf{m}) &= \frac{1}{\varepsilon} \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t - \frac{1}{\varepsilon^2} \int (\Gamma\varphi_t - \Gamma\psi_t)(\varphi_t + \psi_t) d\mu_t + \frac{1}{\varepsilon^2} \int (\Gamma\varphi_t - \Gamma\psi_t)(\varphi_t + \psi_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t.
\end{aligned}$$

In order to get the second equality in Proposition 2.5.1 by property (viii.) in Proposition 1.1.22, together with the current velocity (2.2) and the characterization formula (2.13),

$$\begin{aligned}
\frac{1}{\varepsilon} \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t &= \frac{1}{\varepsilon} \int \Gamma(\psi_t + \varphi_t, \psi_t - \varphi_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int \Gamma(2\theta_t, \varepsilon \log \rho_t) d\mu_t \\
&= 2 \int \Gamma(\theta_t, \log \rho_t) d\mu_t = \int \nabla \theta_t \cdot \nabla \log \rho_t d\mu_t
\end{aligned}$$

and this concludes the proof.

Proof of Proposition 2.5.2

◁ Again for simplicity we drop the index ε .

$$\frac{d^2}{dt^2} H(\mu_t | \mathbf{m}) = \frac{1}{\varepsilon} \frac{d}{dt} \int \Gamma\psi_t - \Gamma\varphi_t d\mu_t.$$

By property (vii.) in Proposition 1.1.22

$$\begin{aligned}
\frac{d^2}{dt^2} H(\mu_t | \mathbf{m}) &= \frac{1}{\varepsilon} \int 2\Gamma(\psi_t, \dot{\psi}_t) - 2\Gamma(\varphi_t, \dot{\varphi}_t) + \frac{1}{\varepsilon} (\Gamma\psi_t - \Gamma\varphi_t)(\dot{\psi}_t + \dot{\varphi}_t) d\mu_t \\
&\stackrel{*}{=} \frac{1}{\varepsilon} \int 2\Gamma(\psi_t, -\varepsilon L\psi_t - \Gamma\psi_t) - 2\Gamma(\varphi_t, \varepsilon L\varphi_t + \Gamma\varphi_t) + \\
&\quad + \frac{1}{\varepsilon} (\Gamma\psi_t - \Gamma\varphi_t) (\varepsilon L\varphi_t - \varepsilon L\psi_t + \Gamma\varphi_t - \Gamma\psi_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int -2\varepsilon\Gamma(\psi_t, L\psi_t) - 2\Gamma(\psi_t, \Gamma\psi_t) - 2\varepsilon\Gamma(\varphi_t, L\varphi_t) - 2\Gamma(\varphi_t, \Gamma\varphi_t) \\
&\quad + \Gamma\psi_t L\varphi_t - \Gamma\psi_t L\psi_t - \Gamma\varphi_t L\varphi_t + \Gamma\varphi_t L\psi_t \\
&\quad + \frac{1}{\varepsilon} \Gamma\psi_t \Gamma\varphi_t - \frac{1}{\varepsilon} (\Gamma\psi_t)^2 - \frac{1}{\varepsilon} (\Gamma\varphi_t)^2 + \frac{1}{\varepsilon} \Gamma\varphi_t \Gamma\psi_t d\mu_t \\
&= \frac{1}{\varepsilon} \int 2\varepsilon\Gamma_2\psi_t + 2\varepsilon\Gamma_2\varphi_t - \underbrace{\varepsilon L\Gamma\psi_t - \varepsilon L\Gamma\varphi_t}_{(**)} - 2\Gamma(\psi_t, \Gamma\psi_t) - 2\Gamma(\varphi_t, \Gamma\varphi_t) \\
&\quad + \Gamma\psi_t L\varphi_t - \Gamma\psi_t L\psi_t - \Gamma\varphi_t L\varphi_t + \Gamma\varphi_t L\psi_t \\
&\quad + \frac{1}{\varepsilon} \Gamma\psi_t \Gamma\varphi_t - \frac{1}{\varepsilon} (\Gamma\psi_t)^2 - \frac{1}{\varepsilon} (\Gamma\varphi_t)^2 + \frac{1}{\varepsilon} \Gamma\varphi_t \Gamma\psi_t d\mu_t. \tag{2.57}
\end{aligned}$$

Where the star equality is given by (2.16), and the last equality is given by definition of Γ_2 (Definition 1.1.21). Integrating the terms $(**)$ twice by parts, we obtain

$$\begin{aligned}
(**) &= -\varepsilon \int L(\Gamma\psi_t + \Gamma\varphi_t) e^{\frac{\psi_t + \varphi_t}{\varepsilon}} d\mathbf{m} \\
&= \varepsilon \int \Gamma(\Gamma\psi_t + \Gamma\varphi_t, e^{\frac{\psi_t + \varphi_t}{\varepsilon}}) d\mathbf{m} \\
&= \int \Gamma(\Gamma\psi_t + \Gamma\varphi_t, \varphi_t + \psi_t) d\mu_t \\
&\stackrel{*}{=} - \int (\Gamma\psi_t + \Gamma\varphi_t) \Gamma(\varphi_t + \psi_t, e^{\frac{\psi_t + \varphi_t}{\varepsilon}}) + (\Gamma\psi_t + \Gamma\varphi_t) L(\varphi_t + \psi_t) e^{\frac{\psi_t + \varphi_t}{\varepsilon}} d\mathbf{m} \\
&= - \int \frac{1}{\varepsilon} (\Gamma\psi_t + \Gamma\varphi_t) \Gamma(\varphi_t + \psi_t, \varphi_t + \psi_t) + (\Gamma\psi_t + \Gamma\varphi_t) L(\varphi_t + \psi_t) d\mu_t \\
&= -\frac{1}{\varepsilon} \int (\Gamma\psi_t + \Gamma\varphi_t) (\Gamma\varphi_t + 2\Gamma(\varphi_t, \psi_t) + \Gamma\psi_t) d\mu_t + \\
&\quad - \int \Gamma\psi_t L\varphi_t + \Gamma\psi_t L\psi_t + \Gamma\varphi_t L\varphi_t + \Gamma\varphi_t L\psi_t d\mu_t
\end{aligned}$$

where the marked equality is again due to property (vi.) in Proposition 1.1.22. Therefore, in (2.57) we get,

$$\begin{aligned}
\frac{d^2}{dt^2}H(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int 2\varepsilon\Gamma_2\psi_t + 2\varepsilon\Gamma_2\varphi_t - 2\Gamma(\psi_t, \Gamma\psi_t) - 2\Gamma(\varphi_t, \Gamma\varphi_t) \\
&\quad + \Gamma\psi_t L\varphi_t - \Gamma\psi_t L\psi_t - \Gamma\varphi_t L\varphi_t + \Gamma\varphi_t L\psi_t - \frac{1}{\varepsilon}(\Gamma\psi_t)^2 \\
&\quad - \frac{1}{\varepsilon}\Gamma\psi_t\Gamma\varphi_t - \frac{2}{\varepsilon}\Gamma\psi_t\Gamma(\varphi_t, \psi_t) - \frac{1}{\varepsilon}(\Gamma\varphi_t)^2 - \frac{2}{\varepsilon}\Gamma\varphi_t\Gamma(\psi_t, \varphi_t) \\
&\quad - \frac{1}{\varepsilon}\Gamma\psi_t\Gamma\varphi_t - \Gamma\psi_t L\varphi_t - \Gamma\psi_t L\psi_t - \Gamma\varphi_t L\varphi_t - \Gamma\varphi_t L\psi_t \\
&\quad + \frac{1}{\varepsilon}\Gamma\psi_t\Gamma\varphi_t - \frac{1}{\varepsilon}(\Gamma\psi_t)^2 - \frac{1}{\varepsilon}(\Gamma\varphi_t)^2 + \frac{1}{\varepsilon}\Gamma\varphi_t\Gamma\psi_t d\mu_t.
\end{aligned}$$

After simplification it reduces to

$$\begin{aligned}
\frac{d^2}{dt^2}H(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int 2\varepsilon\Gamma_2\psi_t + 2\varepsilon\Gamma_2\varphi_t \underbrace{-2\Gamma(\psi_t, \Gamma\psi_t)}_{(***)} \underbrace{-2\Gamma(\varphi_t, \Gamma\varphi_t)}_{(****)} - 2\Gamma\psi_t L\psi_t \\
&\quad - 2\Gamma\varphi_t L\varphi_t - \frac{2}{\varepsilon}(\Gamma\psi_t)^2 - \frac{2}{\varepsilon}(\Gamma\varphi_t)^2 - \frac{2}{\varepsilon}\Gamma(\varphi_t, \psi_t)(\Gamma\varphi_t + \Gamma\psi_t) d\mu_t.
\end{aligned}$$

We integrate by parts (**),

$$\begin{aligned}
(***) &= -2 \int \Gamma(\psi_t, \Gamma\psi_t) e^{\frac{\psi_t + \varphi_t}{\varepsilon}} d\mathbf{m} \\
&= 2 \int \Gamma\psi_t \Gamma(\psi_t, e^{\frac{\psi_t + \varphi_t}{\varepsilon}}) + \Gamma\psi_t L\psi_t e^{\frac{\psi_t + \varphi_t}{\varepsilon}} d\mathbf{m} \\
&= 2 \int \frac{1}{\varepsilon} \Gamma\psi_t \Gamma(\psi_t, \varphi_t + \psi_t) + \Gamma\psi_t L\psi_t d\mu_t \\
&= 2 \int \frac{1}{\varepsilon} \Gamma\psi_t \Gamma(\psi_t, \varphi_t) + \frac{1}{\varepsilon} (\Gamma\psi_t)^2 + \Gamma\psi_t L\psi_t d\mu_t.
\end{aligned}$$

Analogously, we can deduce that

$$(****) = 2 \int \frac{1}{\varepsilon} \Gamma\varphi_t \Gamma(\psi_t, \varphi_t) + \frac{1}{\varepsilon} (\Gamma\varphi_t)^2 + \Gamma\varphi_t L\varphi_t d\mu_t.$$

So, replacing these two terms in (2.57),

$$\begin{aligned}
\frac{d^2}{dt^2}H(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int 2\varepsilon\Gamma_2\psi_t + 2\varepsilon\Gamma_2\varphi_t + \frac{2}{\varepsilon}\Gamma\psi_t\Gamma(\psi_t, \varphi_t) + \frac{2}{\varepsilon}(\Gamma\psi_t)^2 + 2\Gamma\psi_t L\psi_t \\
&\quad + \frac{2}{\varepsilon}\Gamma\varphi_t\Gamma(\psi_t, \varphi_t) + \frac{2}{\varepsilon}(\Gamma\varphi_t)^2 + 2\Gamma\varphi_t L\varphi_t - 2\Gamma\psi_t L\psi_t \\
&\quad - 2\Gamma\varphi_t L\varphi_t - \frac{2}{\varepsilon}(\Gamma\psi_t)^2 - \frac{2}{\varepsilon}(\Gamma\varphi_t)^2 - \frac{2}{\varepsilon}\Gamma(\varphi_t, \psi_t)(\Gamma\varphi_t + \Gamma\psi_t) d\mu_t \\
&= \int 2\Gamma_2\psi_t + 2\Gamma_2\varphi_t d\mu_t.
\end{aligned}$$

And after simplification of equal opposite terms, we get the announced result. Finally to get the second equality in Proposition 2.5.2, we use property (ix.) in Proposition 1.1.22, together with the current velocity (2.2) and the characterization formula (2.13),

$$\begin{aligned}
2 \int \Gamma_2\psi_t + \Gamma_2\varphi_t d\mu_t &= \int \Gamma_2(\psi_t + \varphi_t) + \Gamma_2(\psi_t - \varphi_t) d\mu_t \\
&= \int 4\Gamma_2\theta_t + \varepsilon^2\Gamma_2 \log \rho_t d\mu_t.
\end{aligned}$$

Proof of Proposition 2.5.3

◁

$$\begin{aligned}
\frac{d}{dt}I(\mu_t|\mathbf{m}) &= \frac{d}{dt} \frac{\varepsilon}{2} \int |\nabla \log \rho_t|^2 d\mu_t \\
&= \frac{d}{dt} \frac{1}{\varepsilon} \int \Gamma(\varphi_t + \psi_t) e^{\frac{\varphi_t + \psi_t}{\varepsilon}} d\mathbf{m} \\
&= \frac{1}{\varepsilon} \int 2\Gamma(\varphi_t + \psi_t, \dot{\varphi}_t + \dot{\psi}_t) + \frac{1}{\varepsilon} \Gamma(\varphi_t + \psi_t)(\dot{\varphi}_t + \dot{\psi}_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int 2\Gamma(\varphi_t + \psi_t, -\varepsilon L(\psi_t - \varphi_t) - \Gamma\psi_t + \Gamma\varphi_t) \\
&\quad + \frac{1}{\varepsilon} \Gamma(\varphi_t + \psi_t)(-\varepsilon L(\psi_t - \varphi_t) - \Gamma\psi_t + \Gamma\varphi_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int -2\varepsilon\Gamma(\varphi_t, L\psi_t) + 2\varepsilon\Gamma(\varphi_t, L\varphi_t) - 2\varepsilon\Gamma(\psi_t, L\psi_t) + 2\varepsilon\Gamma(\psi_t, L\varphi_t) \\
&\quad + 2\Gamma(\varphi_t + \psi_t, \Gamma\varphi_t - \Gamma\psi_t) - \Gamma(\varphi_t + \psi_t)L(\psi_t - \varphi_t) + \frac{1}{\varepsilon} \Gamma(\varphi_t + \psi_t)(\Gamma\varphi_t - \Gamma\psi_t) d\mu_t \\
&= \frac{1}{\varepsilon} \int -2\varepsilon\Gamma_2\varphi_t + \underbrace{\varepsilon L\Gamma\varphi_t}_{(*)} + 2\varepsilon\Gamma_2\psi_t - \underbrace{\varepsilon L\Gamma\psi_t}_{(**)} - 2\varepsilon\Gamma(\varphi_t, L\psi_t) + 2\varepsilon\Gamma(\psi_t, L\varphi_t) \\
&\quad + 2\Gamma(\varphi_t + \psi_t, \Gamma\varphi_t - \Gamma\psi_t) - \Gamma(\varphi_t + \psi_t)L(\psi_t - \varphi_t) \\
&\quad + \frac{1}{\varepsilon} \Gamma(\varphi_t + \psi_t)(\Gamma\varphi_t - \Gamma\psi_t) d\mu_t. \tag{2.58}
\end{aligned}$$

Integrating by parts (*) and (**)

$$\begin{aligned}
(*) &= \varepsilon \int L\Gamma\varphi_t d\mu_t \\
&= \int \Gamma(\Gamma\varphi_t, \varphi_t + \psi_t) d\mu_t \\
(**) &= - \int \Gamma(\Gamma\psi_t, \varphi_t + \psi_t) d\mu_t.
\end{aligned}$$

It follows that,

$$(*) + (**) = - \int \Gamma(\varphi_t + \psi_t, \Gamma\varphi_t - \Gamma\psi_t) d\mu_t.$$

So, replacing these terms in (2.58),

$$\begin{aligned}
\frac{d}{dt}I(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int -2\varepsilon[\Gamma_2\varphi_t - \Gamma_2\psi_t] + \underbrace{\Gamma(\varphi_t + \psi_t, \Gamma\varphi_t - \Gamma\psi_t)}_{(***)} - 2\varepsilon[\Gamma(\varphi_t, L\psi_t) + \Gamma(\psi_t, L\varphi_t)] \\
&\quad - \Gamma(\varphi_t + \psi_t)L(\psi_t - \varphi_t) + \frac{1}{\varepsilon} \Gamma(\varphi_t + \psi_t)(\Gamma\varphi_t - \Gamma\psi_t) d\mu_t. \tag{2.59}
\end{aligned}$$

Integrating (***) by parts,

$$\begin{aligned}
(***) &= \int \Gamma(\varphi_t + \psi_t, \Gamma\varphi_t - \Gamma\psi_t) d\mu_t \\
&= - \int (\Gamma\varphi_t - \Gamma\psi_t) [\Gamma(\varphi_t + \psi_t) \frac{1}{\varepsilon} + L(\varphi_t + \psi_t)] d\mu_t.
\end{aligned}$$

Therefore, after simplification we get,

$$\begin{aligned} \frac{d}{dt}I(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int -2\varepsilon\Gamma_2\varphi_t + 2\varepsilon\Gamma_2\psi_t - \underbrace{2\varepsilon\Gamma(\varphi_t, L\psi_t)}_{****} + \underbrace{2\varepsilon\Gamma(\psi_t, L\varphi_t)}_{*****} \\ &\quad - \Gamma(\varphi_t + \psi_t)L(\psi_t - \varphi_t) + (\Gamma\varphi_t - \Gamma\psi_t)L(\varphi_t + \psi_t)d\mu_t. \end{aligned}$$

Integrate by parts (***)

$$\begin{aligned} (***) &= 2\varepsilon \int \Gamma(\varphi_t, L\psi_t) d\mu_t \\ &= -2\varepsilon \int L\psi_t\Gamma(\varphi_t, \varphi_t + \psi_t)\frac{1}{\varepsilon} + L\psi_tL\varphi_t d\mu_t. \end{aligned}$$

In the same way,

$$(***) = -2\varepsilon \int L\varphi_t\Gamma(\psi_t, \varphi_t + \psi_t)\frac{1}{\varepsilon} + L\psi_tL\varphi_t d\mu_t.$$

We replace these terms in (2.59), and after further simplification,

$$\begin{aligned} \frac{d}{dt}I(\mu_t|\mathbf{m}) &= \frac{1}{\varepsilon} \int -2\varepsilon\Gamma_2\varphi_t + 2\varepsilon\Gamma_2\psi_t - 2L\varphi_t\Gamma\psi_t - 2L\varphi_t\Gamma(\varphi_t, \psi_t) \\ &\quad - \Gamma(\varphi_t + \psi_t)L(\psi_t - \varphi_t) + (\Gamma\varphi_t - \Gamma\psi_t)L(\varphi_t + \psi_t)d\mu_t \\ &= 2 \int -\Gamma_2\varphi_t + \Gamma_2\psi_t d\mu_t. \end{aligned}$$

The second equality in Proposition 2.5.3 is obtained similarly to the others in Propositions 2.5.1 and 2.5.2.

2.6 Physical motivation and chronology

The Schrödinger problem appears for the first time in a 1931 paper [Sch31] of the Austrian physicist Nobel prize Erwin Schrödinger. Interested by quantum mechanics phenomenon, in the last section of his paper, translated in french the next year [Sch32], he proposed a probability problem in the classic theory of Brownian motion that at first sight has no link with the rest of his paper. But as he advises the reader at the end he will show an astonishing analogy.

The original question of Schrödinger didn't look like the modern formulation in terms of minimal entropy (2.1) but it was motivated by a statistical physics consideration.

Consider the classical heat equation,

$$\partial_t\rho = \Delta\rho, \tag{2.60}$$

where ρ is the probability density of a particles system. Assume that we know at time t_0 that $\rho|_{t=t_0} = \rho_0$ and at a later time t_1 that $\rho|_{t=t_1} = \rho_1$, that is,

$$\begin{cases} \partial_t\rho = \Delta\rho \\ \rho|_{t=t_0} = \rho_0, \quad \rho|_{t=t_1} = \rho_1. \end{cases} \tag{2.61}$$

With only one of the two *initial* conditions the solution of the Cauchy problems,

$$(i) \begin{cases} \partial_t\rho = \Delta\rho \\ \rho|_{t=t_0} = \rho_0 \end{cases} \quad (ii) \begin{cases} \partial_t\rho = -\Delta\rho \\ \rho|_{t=t_1} = \rho_1 \end{cases} \tag{2.62}$$

is well known. But if we know simultaneously the two limit conditions, then the solution of (2.60) will not be equal to the solution of any of the Cauchy problems in (2.62). Schrödinger did not prove existence and uniqueness results for problem (2.61) but he proved that if a solution exists then, it is given by the product of the unique respective solution of (i) and (ii) in (2.62). Indeed this is what we obtained in (2.12).

Therefore the Schrödinger problem can be addressed in the following way: according to (i) in (2.62) we are supposed to find the particles system at time $t_1 > t_0$ at the configuration $\hat{\rho}|_{t=t_1}$ where $\hat{\rho}$ is the solution of the classical Cauchy problem with initial condition ρ_0 . If we find the system at time t_1 in the configuration $\rho_1 \neq \hat{\rho}|_{t=t_1}$ it means that something unexpected perturbed the evolution of the system. Hence the problem is to determine a drift $\nabla\psi$ (a priori it is not unique), able to drive the system in the observed ρ_1 , i.e. a drift such that the problem

$$\begin{cases} \partial_t \rho = \Delta \rho + \nabla \psi \cdot \nabla \rho \\ \rho|_{t=t_0} = \rho_0, \quad \rho|_{t=t_1} = \rho_1 \end{cases} \quad (2.63)$$

admits a solution. We can still formulate this problem in more probabilistic terms. To this aim we basically follow [Léo14b, Sec. 6].

Consider a large number N of particles moving in the state space, that so far we assumed to be \mathbb{R}^n . The particles are independent, non interacting and indistinguishable. The position of each particle is associated to a random variable $Y^i, i = 1, \dots, N$ on Ω , the set of continuous path on \mathbb{R}^n (1.10), such that

$$P(Y^i \in \cdot) = R(\cdot | Y_0 = y_0^i) \in \mathcal{P}(\Omega), \quad \forall 1 \leq i \leq N,$$

where y_0^i is the initial deterministic position of the i -th particle and R is the reference path measure. For instance if

$$Y_t^i = y^i + W_t^i \quad (2.64)$$

we recover the Brownian motion, namely, (2.60). Consider the empirical measure,

$$L^N = \frac{1}{N} \sum_{i=1}^N \delta_{Y^i}$$

that is a random element in $\mathcal{P}(\Omega)$. For any $0 \leq t \leq 1$ the projection of L^N is an element of $\mathcal{P}(\mathbb{R}^n)$ given by,

$$L_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^i}.$$

Suppose that the initial positions are close to some configuration $\mu_0 \in \mathcal{P}(\mathbb{R}^n)$,

$$L_0^N \xrightarrow{N \rightarrow \infty} \mu_0 \in \mathcal{P}(\mathbb{R}^n).$$

By the law of large numbers, we know that for $N \rightarrow \infty$, L^N converges in law to the law of Y^i , that is, taking into account the conditioning with respect to the initial position,

$$L_t^N(dy) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}^n} R(X_t \in dy | X_0 = x) \mu_0(dx). \quad (2.65)$$

For instance in the example (2.64), we expect L_1^N to be close to $\mathcal{N}(0, 1)$ for N large. Schrödinger's question is then the following. Suppose that at time $t = 1$, we observe the

system in a configuration μ_1 different from the one we expect according to (2.65). This is possible since N is finite, but is a rare event. Conditionally to this rare event what is the most likely dynamics of the whole system described by L^N ? Mathematically this is equivalent to determine the limit,

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \text{Prob}(L^N \in \cdot | L_1^N \in B(\mu_1, \varepsilon))$$

in $\mathcal{P}(\mathcal{P}(\Omega))$, where B is the ball of radius ε centered in μ_1 according to some metric on $\mathcal{P}(\mathbb{R}^n)$. In these terms it is clear that it is a large deviations problem, hence it can be solved via the Sanov Theorem [DZ98]. Define $C_\varepsilon = \{P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 \in B(\mu_1, \varepsilon)\}$, a modified version of Sanov's Theorem states that,

$$\text{Prob}(L^N \in A | L^N \in C_\varepsilon) \underset{N \rightarrow \infty}{\asymp} \exp \left[-n \left\{ \inf_{P \in A \cap C_\varepsilon} H(P|R) - \inf_{P \in C_\varepsilon} H(P|R) \right\} \right]. \quad (2.66)$$

Roughly speaking, if $\hat{P} \in C_\varepsilon$ is the solution of

$$\inf_{P \in C_\varepsilon} H(P|R), \quad (2.67)$$

there are two possible situations: $\hat{P} \in A \cap C_\varepsilon$ or $\hat{P} \in C_\varepsilon \setminus (A \cap C_\varepsilon)$. In the first case it means that the two optimal values in (2.66) are equal, thus,

$$\text{Prob}(L^N \in A | L^N \in C_\varepsilon) \underset{N \rightarrow \infty}{\asymp} 1.$$

In the latter case, it means that the two optimal values are different, the exponent is strictly negative, hence at the limit for $N \rightarrow \infty$,

$$\text{Prob}(L^N \in A | L^N \in C_\varepsilon) \underset{N \rightarrow \infty}{\asymp} 0.$$

Summarizing we obtained that $\lim_{N \rightarrow \infty} \text{Prob}(L^N \in \cdot | L^N \in C_\varepsilon) = \delta_{\hat{P}_\varepsilon}$ where \hat{P}_ε is the solution of (2.67). Finally it can be proved that at the limit $\varepsilon \rightarrow 0^+$, $\hat{P}_\varepsilon \rightarrow \hat{P}$ solution of

$$\inf \{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} \quad (2.68)$$

hence we can conclude that

$$\lim_{\varepsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \text{Prob}(L^N \in \cdot | L_1^N \in B(\mu_1, \varepsilon)) = \delta_{\hat{P}} \in \mathcal{P}(\mathcal{P}(\Omega)),$$

where \hat{P} is the unique solution of (2.68) that coincides with the dynamical definition of the Schrödinger problem (2.1). This is equivalent to say that for large enough N , L^N converges in law to \hat{P} .

In the same way the probability that the system evolves spontaneously from μ_0 to the unexpected profile μ_1 is given by,

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} \frac{1}{N} \log \text{Prob}(L_1^N \in B(\mu_1, \varepsilon) | L_0^N \in B(\mu_0, \varepsilon)) = \\ - \inf \{H(P|R); P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} + H(\mu_0|m) \end{aligned} \quad (2.69)$$

where the quantity on the right hand side is equal to the forward entropic cost defined at (2.37).

To summarize, the entropic cost defined at Section 2.3 is an evaluation of the (low) probability to end up in the prescribed μ_1 , starting from μ_0 . The entropic interpolation defined at Section 2.2 is the most likely dynamics followed by the system, and $\nabla \psi_t$ in (2.29) is the unexpected drift which perturbed the reference (Brownian) motion in the formulation (2.63).

2.6.1 Brief history

Schrödinger's work has been remembered by the large public for the "cat", by the mathematical and physical community for the "equation", and by a quite tiny subset of mathematicians for the problem we are dealing with. Although it is an interesting and classical question in probability, Schrödinger's problem has never been mainstream, rather it was forgotten for a long time before being properly studied. After 1932's article, the first contributions for the solvability of system (2.6), were accomplished first by Bernstein [Ber32], successively by Fortet in 1940 [For40], 20 years later by Beurling [Beu60] and still 15 years later in the 70s by Jamison [Jam74, Jam75].

It was Föllmer in the 80s [Föl88] to give a new birth to the Schrödinger problem through its modern formulation in terms of large deviation theory. At the same time, inspired by the work of Schrödinger, Zambrini in [Zam86] started to develop a work of connections between stochastic processes and quantum physics that is still productive, [CZ91, CWZ00, CZ08]. More recently, are due to Mikami and Léonard the results on the convergence of the Schrödinger problem to the Monge-Kantorovich problem, [Mik04, Léo12a].

The last few years have known an increasing popularity of the Schrödinger problem. In connection with the optimal transport theory we recall the works of Chen, Georgiu and Pavon [CGP16, CGP17], and, with applications to functional inequalities, the works of Gentil, Léonard and the author [GLR17, Rip17]. The Schrödinger problem has recently been studied by Gigli and Tamanini in [GT17] in RDC^* spaces and applied to the study of incompressible viscid fluids in [ABLZ17].

In applied mathematics the Schrödinger problem is used as a regularizing approximation of the Monge-Kantorovich problem, being much more performing in numerical computations. We mention the works of Benamou, Carlier, Cuturi, Duval, Nenna, Peyré, Schmitzer such as [Cut13, BCC⁺15, CDPS17]. The Schrödinger problem counts also applications in economics, given by Galichon and Salanié in [GS09].

Chapter 3

Analogies between optimal transport and minimal entropy

This chapter contains a slightly modified version of [GLR17]. We tried to uniform notations to the rest of the manuscript and we decided to keep the original structure of the paper to leave the chapter self-contained. We invite the reader who already passed through Chapter 1 and Chapter 2 to skip Sections 3.2 and 3.3, which are basically repetitions and start directly from Section 3.4.

Abstract We describe some analogy between optimal transport and the Schrödinger problem where the transport cost is replaced by an entropic cost with a reference path measure. A dual Kantorovich type formulation and a Benamou-Brenier type representation formula of the entropic cost are derived, as well as contraction inequalities with respect to the entropic cost. This analogy is also illustrated with some numerical examples where the reference path measure is given by the Brownian motion or the Ornstein-Uhlenbeck process.

Our point of view is measure theoretical, rather than based on stochastic optimal control, and the relative entropy with respect to path measures plays a prominent role.

Resumé Nous décrivons des analogies entre le transport optimal et le problème de Schrödinger lorsque le coût du transport est remplacé par un coût entropique avec une mesure de référence sur les trajectoires. Une formule duale de Kantorovich, une formulation de type Benamou-Brenier du coût entropique sont démontrées, ainsi que des inégalités de contraction par rapport au coût entropique. Cette analogie est aussi illustrée par des exemples numériques où la mesure de référence sur les trajectoires est donnée par le mouvement Brownien ou bien le processus d'Ornstein-Uhlenbeck.

Notre approche s'appuie sur la théorie de la mesure, plutôt que sur le contrôle optimal stochastique, et l'entropie relative joue un rôle fondamental.

3.1 Introduction

In this article, some analogy between optimal transport and the Schrödinger problem is investigated. A Kantorovich type dual equality, a Benamou-Brenier type representation of the entropic cost and contraction inequalities with respect to the entropic cost are derived when the transport cost is replaced by an entropic one. This analogy is also illustrated with some numerical examples.

Our point of view is measure theoretical rather than based on stochastic optimal control as is done in the recent literature; the relative entropy with respect to path measures plays a prominent role.

Before explaining the Schrödinger problem which is associated with an entropy minimization, we first introduce the Wasserstein quadratic transport cost W_2^2 and its main properties. For simplicity, our results are stated in \mathbb{R}^n rather than in a general Polish space. Let us note that properties of the quadratic transport cost can be found in the monumental work by C. Villani [Vil03, Vil09]. In particular its square root W_2 is a (pseudo-)distance on the space of probability measures which is called Wasserstein distance. It has been intensively studied and has many interesting applications. For instance it is an efficient tool for proving convergence to equilibrium of evolution equations, concentration inequalities for measures or stochastic processes and it allows to define curvature in metric measure spaces, see the textbook [Vil09] for these applications and more.

The Wasserstein quadratic cost W_2^2 and the Monge-Kantorovich problem

Let $\mathcal{P}(\mathbb{R}^n)$ be the set of all probability measures on \mathbb{R}^n . We denote its subset of probability measures with a second moment by $\mathcal{P}_2(\mathbb{R}^n) = \{\mu \in \mathcal{P}(\mathbb{R}^n); \int |x|^2 \mu(dx) < \infty\}$. For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, the Wasserstein quadratic cost is

$$W_2^2(\mu_0, \mu_1) = \inf_{\pi} \int_{\mathbb{R}^n \times \mathbb{R}^n} |y - x|^2 \pi(dxdy), \quad (3.1)$$

where the infimum is running over all the couplings π of μ_0 and μ_1 , namely, all the probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ_0 and μ_1 , that is for any bounded measurable functions φ and ψ on \mathbb{R}^n ,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (\varphi(x) + \psi(y)) \pi(dxdy) = \int_{\mathbb{R}^n} \varphi d\mu_0 + \int_{\mathbb{R}^n} \psi d\mu_1.$$

In restriction to $\mathcal{P}_2(\mathbb{R}^n)$, the pseudo-distance W_2 becomes a genuine distance. The *Monge-Kantorovich problem* with a quadratic cost function, consists in finding the optimal couplings π that minimize (3.1). For more details see Section 1.2.

The entropic cost \mathcal{A}^R and the Schrödinger problem

Let fix some reference non-negative measure R on the path space $\Omega = C([0, 1], \mathbb{R}^n)$ and denote R_{01} the measure on $\mathbb{R}^n \times \mathbb{R}^n$. It describes the joint law of the initial position X_0 and the final position X_1 of a random process on \mathbb{R}^n whose law is R . This means that

$$R_{01} = (X_0, X_1)_\# R$$

is the push-forward of R by the mapping (X_0, X_1) . Recall that the push-forward of a measure α on the space \mathbf{A} by the measurable mapping $f : \mathbf{A} \rightarrow \mathbf{B}$ is defined by

$$f_\# \alpha(db) = \alpha(f^{-1}(db)), \quad db \subset \mathbf{B},$$

in other words, for any positive function H ,

$$\int H d(f_\# \alpha) = \int H(f) d\alpha.$$

For any probability measures μ_0 and μ_1 on \mathbb{R}^n , the entropic cost $\mathcal{A}^R(\mu_0, \mu_1)$ of (μ_0, μ_1) is defined up to a constant term by

$$\mathcal{A}^R(\mu_0, \mu_1) = \inf_{\pi} H(\pi|R_{01})$$

where $H(\pi|R_{01}) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \log(d\pi/dR_{01}) d\pi$ is the relative entropy of π with respect to R_{01} and π runs through all the couplings of μ_0 and μ_1 . The *Schrödinger problem* consists in finding the unique optimal entropic plan π that minimizes the above infimum.

In this article, we choose R as the reversible Kolmogorov continuous Markov process specified by the generator $\frac{1}{2}(\Delta - \nabla V \cdot \nabla)$ and the initial reversing measure $e^{-V(x)} dx$.

Aim of the paper

Below in this introductory section, we are going to focus onto four main features of the quadratic transport cost W_2^2 . Namely:

- the Kantorovich dual formulation of W_2^2 ;
- the Benamou-Brenier dynamical formulation of W_2^2 ;
- the displacement interpolations, that is the W_2 -geodesics in $\mathcal{P}_2(\mathbb{R}^n)$;
- the contraction of the heat equation with respect to W_2^2 .

The goal of this article is to recover analogous properties for \mathcal{A}^R instead of W_2^2 , by replacing the Monge-Kantorovich problem with the Schrödinger problem.

Several aspects of the quadratic Wasserstein cost

Let us provide some detail about these four properties.

Kantorovich dual formulation of W_2^2

The following duality result was proved by Kantorovich in [Kan42b]. For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0, \mu_1) = \sup_{\psi} \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} Q\psi d\mu_0 \right\}, \quad (3.2)$$

where the supremum runs over all bounded continuous function ψ and

$$Q\psi(x) = \sup_{y \in \mathbb{R}^n} \{ \psi(y) - |x - y|^2 \}, \quad x \in \mathbb{R}^n.$$

It is often expressed in the equivalent form,

$$W_2^2(\mu_0, \mu_1) = \sup_{\varphi} \left\{ \int_{\mathbb{R}^n} \tilde{Q}\varphi d\mu_1 - \int_{\mathbb{R}^n} \varphi d\mu_0 \right\}, \quad (3.3)$$

where the supremum runs over all bounded continuous function φ and

$$\tilde{Q}\varphi(y) = \inf_{x \in \mathbb{R}^n} \{ \varphi(x) + |y - x|^2 \}, \quad y \in \mathbb{R}^n.$$

The map $Q\psi$ is called the sup-convolution of ψ and its defining identity is sometimes referred to as the Hopf-Lax formula.

Benamou-Brenier formulation of W_2^2

The Wasserstein cost admits a dynamical formulation: the so-called Benamou-Brenier formulation which was proposed in [BB00]. It states that for any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu, v)} \int_{\mathbb{R}^n \times [0,1]} |v_t|^2 d\nu_t dt, \quad (3.4)$$

where the infimum runs over all paths $(\nu_t, v_t)_{t \in [0,1]}$ where $\nu_t \in \mathcal{P}(\mathbb{R}^n)$ and $v_t(x) \in \mathbb{R}^n$ are such that ν_t is absolutely continuous with respect to time in the sense of [AGS08, Ch. 1] for all $0 \leq t \leq 1$, $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and

$$\partial_t \nu_t + \nabla \cdot (\nu_t v_t) = 0, \quad 0 \leq t \leq 1.$$

In this equation which is understood in a weak sense, $\nabla \cdot$ stands for the standard divergence of a vector field in \mathbb{R}^n and ν_t is identified with its density with respect to Lebesgue measure. This general result is proved in [AGS08, Ch. 8]. A proof under the additional requirement that μ_0, μ_1 have compact supports is available in [Vil03, Thm. 8.1].

Displacement interpolations

The metric space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ is geodesic. This means that for any probability measure $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, there exists a path $(\mu_t^{Mc})_{t \in [0,1]}$ in $\mathcal{P}_2(\mathbb{R}^n)$ such that for any $s, t \in [0, 1]$,

$$W_2(\mu_s, \mu_t) = |t - s| W_2(\mu_0, \mu_1).$$

Such a path is a constant speed geodesic in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, see [AGS08, Ch. 7]. Moreover when ν is absolutely continuous with respect to the Lebesgue measure, there exists a convex function φ on \mathbb{R}^n such that for any $t \in [0, 1]$, the geodesic is given by

$$\mu_t = ((1 - t)\text{Id} + t\nabla\varphi)_\# \mu_0. \quad (3.5)$$

This interpolation is called the McCann displacement interpolation in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, see [Vil03, Ch. 5].

Contractions in Wasserstein distance

Contraction in Wasserstein distance is a way to define the curvature of the underlying space or of the reference Markov operator. In its general formulation, the von Renesse-Sturm Theorem tells that the heat equation in a smooth, complete and connected Riemannian manifold satisfies a contraction property with respect to the Wasserstein distance if and only if the Ricci curvature is bounded from below, see [SvR09]. In the context of the present article where Kolmogorov semigroups on \mathbb{R}^n are considered, two main contraction results will be explained with more details in Section 3.6.

Organization of the paper

The setting of the present work and notation are introduced in Section 3.2. The entropic cost \mathcal{A}^R is defined with more detail in Section 3.3 together with the related notion of entropic interpolation, an analogue of the displacement interpolation. A dual Kantorovich type formulation and a Benamou-Brenier type formulation of the entropic cost are derived respectively at Sections 3.4 and 3.5. Section 3.6 is dedicated to the contraction properties of the heat flow with respect to the entropic cost. In the last Section 3.7, we give some examples of entropic interpolations between Gaussian distributions when the reference path measure is given by the Brownian motion or the Ornstein-Uhlenbeck process.

Literature

The Benamou-Brenier formulation of the entropic cost which is stated at Corollary 3.5.3 was proved recently by Chen, Georgiou and Pavon in [CGP16] (in a slightly less general setting) without any mention to optimal transport (in this respect Corollary 3.5.6 relating the entropic and Wasserstein costs is new). Although our proof of Corollary 3.5.3 is close to their proof, we think it is worth including it in the present article to emphasize the analogies between displacement and entropic interpolations. In addition, we also provide a time asymmetric version of this formulation at Theorem 3.5.1.

Both [MT06] and [CGP16] share the same stochastic optimal control viewpoint. This differs from the entropic approach of the present paper.

Let us notice that Theorem 3.6.1 is a new result: it provides contraction inequalities with respect to the entropic cost. Moreover, examples and comparison proposed at the end of the paper, with respect two different kernels (Gaussian and Ornstein-Uhlenbeck) are new.

3.2 The reference path measure

We make precise the reference path measure R to which the entropic cost \mathcal{A} is associated. Although more general reversible path measures R would be all right to define a well-suited entropic cost, we prefer to consider the specific class of Kolmogorov Markov measures introduced at Section 1.1.2. This choice is motivated by the fact that, as presented in [Léo12a], the Monge Kantorovich problem is the limit of a sequence of entropy minimization problems, when a proper fluctuation parameter tends to zero. The Kolmogorov Markov measures, as reference measures in the Schrödinger problem, admit as a limit case the Monge Kantorovich problem with quadratic cost function, namely the Wasserstein distance.

Notation

For any measurable set Y , we denote respectively by $\mathcal{P}(Y)$ and $\mathcal{M}(Y)$ the set of all the probability measures and all positive σ -finite measures on Y . The *relative entropy* of a probability measure $p \in \mathcal{P}(Y)$ with respect to a positive measure $r \in \mathcal{M}(Y)$ is loosely defined by

$$H(p|r) := \begin{cases} \int_Y \log(dp/dr) dp \in (-\infty, \infty], & \text{if } p \ll r, \\ \infty, & \text{otherwise.} \end{cases}$$

For some assumptions on the reference measure r that guarantee the above integral to be meaningful and bounded from below, see Section 1.1.2. For a rigorous definition and some properties of the relative entropy with respect to an unbounded measure see Section 1.1.1 and [Léo14a]. The state space \mathbb{R}^n is equipped with its Borel σ -field and the path space Ω with the canonical σ -field $\sigma(X_t; 0 \leq t \leq 1)$ generated by the canonical process

$$X_t(\omega) := \omega_t \in \mathbb{R}^n, \quad \omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega, \quad 0 \leq t \leq 1.$$

For any path measure $Q \in \mathcal{M}_+(\Omega)$ and any $0 \leq t \leq 1$, we denote

$$Q_t(\cdot) := Q(X_t \in \cdot) = (X_t)_\# Q \in \mathcal{M}(\mathbb{R}^n),$$

the push-forward of Q by X_t . When Q is a probability measure, Q_t is the law of the random location X_t at time t when the law of the whole trajectory is Q .

The Kolmogorov Markov measure R and its semigroup

Most of our results can be stated in the general setting of a Polish state space. For the sake of simplicity, the setting of the present paper is particularized. The state space is \mathbb{R}^n and the reference path measure R is the Markov path measure associated with the generator

$$\frac{1}{2}(\Delta - \nabla V \cdot \nabla) \quad (3.6)$$

and the corresponding reversible measure

$$\mathbf{m} = e^{-V} \mathcal{L}$$

as its initial measure, where \mathcal{L} is the Lebesgue measure. It is assumed that the potential V is a \mathcal{C}^2 function on \mathbb{R}^n such that the martingale problem associated with the generator (3.6) on the domain \mathcal{C}^2 and the initial measure \mathbf{m} admits a unique solution $R \in \mathcal{M}_+(\Omega)$. This is the case for instance when the following hypothesis are satisfied.

Existence hypothesis (Exi)

There exists some constant $c > 0$ such that one of the following assumptions holds true:

- (i) $\lim_{|x| \rightarrow \infty} V(x) = +\infty$ and $\inf\{|\nabla V|^2 - \Delta V/2\} > -\infty$, or
- (ii) $-x \cdot \nabla V(x) \leq c(1 + |x|^2)$, for all $x \in \mathbb{R}^n$.

See [Roy99, Thm. 2.2.19] for the existence result under the assumptions (i) or (ii). For any initial condition $X_0 = x \in \mathbb{R}^n$, the path measure $R_x := R(\cdot \mid X_0 = x) \in \mathcal{P}(\Omega)$ is the law of the weak solution of the stochastic differential equation

$$dX_t = -\nabla V(X_t)/2 dt + dW_x(t), \quad 0 \leq t \leq 1$$

where W_x is an R_x -Brownian motion. The Kolmogorov Markov measure is

$$R(\cdot) = \int_{\mathbb{R}^n} R_x(\cdot) \mathbf{m}(dx) \in \mathcal{M}(\Omega).$$

Recall that $\mathbf{m} = e^{-V} \mathcal{L}$ is not necessary a probability measure. The Markov semigroup associated to R is defined for any bounded measurable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and any $t \geq 0$, by

$$T_t f(x) = E_{R_x} f(X_t), \quad x \in \mathbb{R}^n.$$

It is reversible with reversing measure \mathbf{m} as defined in [BGL14].

Regularity hypothesis (Reg1)

We also assume for simplicity that $(T_t)_{t \geq 0}$ admits for any $t > 0$, a *density kernel* with respect to \mathbf{m} , a probability density $p_t(x, y)$ such that

$$T_t f(x) = \int_{\mathbb{R}^n} f(y) p_t(x, y) \mathbf{m}(dy). \quad (3.7)$$

For instance, when $V(x) = |x|^2/2$, then R is the path measure associated to the Ornstein-Uhlenbeck process with the Gaussian measure as its reversing measure. When $V = 0$, we

recover the Brownian motion with Lebesgue measure as its reversing measure. We studied these examples in Sections 1.1.2 and 1.1.4. Examples of Kolmogorov semigroups admitting a density kernel can be found for instance in [BBGM12, Cor. 4.2]. This semigroup is fixed once for all.

Properties of the path measure R

The measure R is our *reference* path measure and it satisfies the following properties.

- (a) It is Markov, that is for any $t \in [0, 1]$, $R(X_{[t,1]} \in \cdot | X_{[0,t]}) = R(X_{[t,1]} \in \cdot | X_t)$. See [Léo14a] for the definition of the conditional law for unbounded measures since R is not necessary a probability measure.
- (b) It is reversible. This means that for all $0 \leq T \leq 1$, the restriction $R_{[0,T]}$ of R to the sigma-field $\sigma(X_{[0,T]})$ generated by $X_{[0,T]} = (X_t)_{0 \leq t \leq T}$, is invariant with respect to time-reversal, that is $[(X_{T-t})_{0 \leq t \leq T}]_{\#} R_{[0,T]} = R_{[0,T]}$. Any reversible measure R is stationary, i.e. $R_t = m$, for all $0 \leq t \leq 1$ for some $m \in \mathcal{M}(\mathbb{R}^n)$. This measure m is called the reversing measure of R and is often interpreted as an equilibrium of the dynamics specified by the kernel $(R_x; x \in \mathbb{R}^n)$. One says for short that R is m -reversible.

3.3 Entropic cost and entropic interpolations

We recall the definitions of the Schrödinger problem, the entropic cost and the entropic interpolation which are respectively the analogues of the Monge-Kantorovich problem, the Wasserstein cost and the displacement interpolation that were briefly described in the introduction and in Chapter 2.

Let us state the definition of the entropic cost associated with R .

Definition 3.3.1 (Entropic cost) *Consider the projection*

$$R_{01} := (X_0, X_1)_{\#} R \in \mathcal{M}(\mathbb{R}^n \times \mathbb{R}^n)$$

of R onto the endpoint space $\mathbb{R}^n \times \mathbb{R}^n$. For any $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$,

$$\mathcal{A}(\mu_0, \mu_1) = \inf\{H(\pi | R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi_0 = \mu_0, \pi_1 = \mu_1\} - \frac{1}{2}[H(\mu_0 | \mathbf{m}) + H(\mu_1 | \mathbf{m})]$$

is the R -entropic cost of (μ_0, μ_1) .

This definition is related to a static Schrödinger problem. It also admits a dynamical formulation.

Definition 3.3.2 (Dynamical formulation of the Schrödinger problem) *The Schrödinger problem associated to R , μ_0 and μ_1 consists in finding the minimizer \hat{P} of the relative entropy $H(\cdot | R)$ among all the probability path measures $P \in \mathcal{P}(\Omega)$ with prescribed initial and final marginals $P_0 = \mu_0$ and $P_1 = \mu_1$,*

$$H(\hat{P} | R) = \min\{H(P | R), P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\}. \quad (3.8)$$

It is easily seen that its minimal value is the entropic cost,

$$\mathcal{A}(\mu_0, \mu_1) = \inf\{H(P|R); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} - \frac{1}{2}[H(\mu_0|\mathbf{m}) + H(\mu_1|\mathbf{m})] \quad (3.9)$$

see for instance [Léo14b, Lemma 2.4].

Remark 3.3.3 (1) *First of all, when R is not a probability measure, the relative entropy might take some negative values and even the value $-\infty$. However, because of the decrease of information by push-forward mappings, we have*

$$H(P|R) \geq \max(H(\mu_0|m), H(\mu_1|m)),$$

see [Léo14a, Thm. 2.4] for instance. Hence $H(P|R)$ is well defined in $(-\infty, \infty]$ whenever $H(\mu_0|m) > -\infty$ or $H(\mu_1|m) > -\infty$. This will always be assumed.

(2) *Even the nonnegative quantity $\mathcal{A}(\mu_0, \mu_1) \geq 0$ cannot be the square of a distance such as the Wasserstein cost W_2^2 . There is no reason for $\mathcal{A}(\mu_0, \mu_1) = 0$ in the special situation where $\mu_0 = \mu_1 = \mu$,*

(3) *A good news about \mathcal{A} is that since R is reversible, it is symmetric: $\mathcal{A}(\mu, \nu) = \mathcal{A}(\nu, \mu)$. To see this, let us denote $X_t^* = X_{1-t}$, $0 \leq t \leq 1$, and $Q^* := (X^*)_{\#}Q$ the time reversal of any $Q \in \mathcal{M}_+(\Omega)$. As X^* is one-to-one, we have $H(P|R) = H(P^*|R^*)$ and since we assume that $R^* = R$, we see that*

$$H(P|R) = H(P^*|R), \quad \forall P \in \mathcal{P}(\Omega). \quad (3.10)$$

Hence, if P solves (3.8) with $(\mu_0, \mu_1) = (\mu, \nu)$, then $X^*_{\#}P$ solves (3.8) with $(\mu_0, \mu_1) = (\nu, \mu)$ and these Schrödinger problems share the same value.

Another definition of entropic cost is possible, depending on the time direction.

Definition 3.3.4 *The forward and backward entropic cost are defined respectively as,*

$$\vec{\mathcal{A}}(\mu_0, \mu_1) = \inf\{H(P|R); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} - H(\mu_0|\mathbf{m})$$

$$\overleftarrow{\mathcal{A}}(\mu_0, \mu_1) = \inf\{H(P|R); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} - H(\mu_1|\mathbf{m}).$$

Existence of a minimizer. Entropic interpolation

We recall some general results from [Léo14b, Thm. 2.12] about the solution of the dynamical Schrödinger problem (3.8). Let us denote by $p(x, y)$ the probability density introduced at (3.7), at time $t = 1$, so that

$$R_{01}(dxdy) = m(dx)p(x, y)m(dy).$$

In order for (3.8) to admit a unique solution, it is enough that it satisfies the following hypothesis:

Regularity hypothesis (Reg2)

- (i) $p(x, y) \geq e^{-A(x)-A(y)}$ for some nonnegative measurable function A on \mathbb{R}^n ;
- (ii) $\int_{\mathbb{R}^n \times \mathbb{R}^n} e^{-B(x)-B(y)} p(x, y) m(dx)m(dy) < \infty$ for some nonnegative measurable function B on \mathbb{R}^n ;
- (iii) $\int_{\mathbb{R}^n} (A + B) d\mu_0, \int_{\mathbb{R}^n} (A + B) d\mu_1 < \infty$ where A appears at (i) and B appears at (ii);
- (iv) $-\infty < H(\mu_0|m), H(\mu_1|m) < \infty$;

Assumptions (ii)-(iii) are useful to define rigorously $H(\mu_0|m)$ and $H(\mu_1|m)$. Under these assumptions the entropic cost $\mathcal{A}(\mu_0, \mu_1)$ is finite and the minimizer P (entropic bridge) of the Schrödinger problem (3.8) is characterized, in addition to the marginal constraints $P_0 = \mu_0, P_1 = \mu_1$, by the product formula

$$P = f(X_0)g(X_1)R \quad (3.11)$$

for some measurable functions f and g on \mathbb{R}^n . The uniqueness of the solution is a direct consequence of the fact that (3.8) is a *strictly* convex minimization problem.

Definition 3.3.5 (Entropic interpolation) *The R -entropic interpolation between μ_0 and μ_1 is defined as the marginal flow of the minimizer P of (3.8), that is $\mu_t := P_t \in \mathcal{P}(\mathbb{R}^n), 0 \leq t \leq 1$.*

Proposition 3.3.6 *Under the hypotheses (Exi), (Reg1) and (Reg2), the R -entropic interpolation between μ_0 and μ_1 is characterized by*

$$\mu_t = e^{\varphi_t + \psi_t} m, \quad 0 \leq t \leq 1, \quad (3.12)$$

where

$$\varphi_t = \log T_t f, \quad \psi_t = \log T_{1-t} g, \quad 0 \leq t \leq 1, \quad (3.13)$$

and the measurable functions f, g solve the following system

$$\frac{d\mu_0}{dm} = fT_1 g, \quad \frac{d\mu_1}{dm} = gT_1 f. \quad (3.14)$$

The system (3.14) is often called the Schrödinger system. It simply expresses the marginal constraints. Its solutions (f, g) are precisely the functions that appear in the identity (3.11). Actually it is difficult or impossible to solve explicitly the system (3.14). However, in Section 3.7, we will see some particular examples in the Gaussian setting where the system admits an explicit solution. Some numerical algorithms have been proposed recently in [BCC⁺15].

In our setting where R is the Kolmogorov path measure defined at (3.6), the entropic interpolation μ_t admits a density $\mu_t(z) := d\mu_t/dz$ with respect to the Lebesgue measure and $\rho_t := d\mu_t/dm$ with respect to the equilibrium measure. It is important to notice that, contrary to the McCann interpolation, the $(t, x) \mapsto \mu_t(x)$ is smooth on $(t, x) \in]0, 1[\times \mathbb{R}$ and solves the transport equation

$$\partial_t \mu_t + \nabla \cdot (\mu_t v^{cu}(t, \mu_t)) = 0 \quad (3.15)$$

with the initial condition μ_0 and where $v^{cu}(t, \mu_t, z) = \nabla \psi_t(z) - \nabla V(z)/2 + \nabla \log \mu_t(z)/2$ refers to the *current velocity* (see Section 2.2), introduced by Nelson in [Nel67, Chap. 11]

(this will be recalled at Section 3.5). The current velocity is a smooth function and the ordinary differential equation

$$\dot{x}_t(x) = v^{cu}(t, x_t(x)), \quad x_0 = x$$

admits a unique solution for any initial position x , the solution of the continuity equation (3.15) admits the following push-forward expression,

$$\mu_t = (x_t)_\# \mu_0, \quad 0 \leq t \leq 1, \quad (3.16)$$

in analogy with the displacement interpolation given at (3.5).

Remark 3.3.7 (From the entropic cost to the Wasserstein cost) *The Wasserstein distance is a limit case of the entropic cost. We shall use this result to compare contraction properties in Section 3.6 and also to illustrate the examples in Section 3.7.*

Let us consider the following dilatation in time with ratio $\varepsilon > 0$ of the reference path measure R : $R^\varepsilon := (X^\varepsilon)_\# R$ where $X^\varepsilon(t) := X_{\varepsilon t}$, $0 \leq t \leq 1$. It is shown in [Léo12a] and presented in Section 2.4 that some re-normalization of the entropic cost $\mathcal{A}^{R^\varepsilon}$ converges to the Wasserstein distance when ε goes to 0. Namely,

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \mathcal{A}^\varepsilon(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1)/2.$$

Even better, when μ_0 and μ_1 are absolutely continuous, the entropic interpolation $(\mu_t^\varepsilon)_{0 \leq t \leq 1}$ between μ_0 and μ_1 converges as ε tends to zero towards the McCann displacement interpolation $(\mu_t)_{0 \leq t \leq 1}$, see (3.5).

3.4 Kantorovich dual equality for the entropic cost

We derive the analogue of the Kantorovich dual equality (3.2) when the Wasserstein cost is replaced by the entropic cost.

Theorem 3.4.1 (Kantorovich dual equality for the entropic cost) *Let V, μ_0 and μ_1 be such that the hypothesis (Exi), (Reg1) and (Reg2) stated in Section 3.3 are satisfied. We have*

$$\vec{\mathcal{A}}(\mu_0, \mu_1) = \sup \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}^R \psi d\mu_0; \psi \in C_b(\mathbb{R}^n) \right\}$$

where for every $\psi \in C_b(\mathbb{R}^n)$, $\mathcal{Q}^R \psi(x) := \log E_{R^x} e^{\psi(X_1)} = \log T_1(e^\psi)(x)$, $x \in \mathbb{R}^n$.

This result was obtained by Mikami and Thieullen in [MT06] with an alternate statement and a different proof. The present proof is based on an abstract dual equality which is stated below at Lemma 3.4.2. Let us first describe the setting of this lemma.

Let \mathbb{U} be a vector space and $\Phi : \mathbb{U} \rightarrow (-\infty, \infty]$ be an extended real valued function on \mathbb{U} . Its convex conjugate Φ^* on the algebraic dual space \mathbb{U}^* of \mathbb{U} is defined by

$$\Phi^*(\ell) := \sup_{u \in \mathbb{U}} \left\{ \langle \ell, u \rangle_{\mathbb{U}^*, \mathbb{U}} - \Phi(u) \right\} \in [-\infty, \infty], \quad \ell \in \mathbb{U}^*.$$

We consider a linear map $A : \mathbb{U}^* \rightarrow \mathbb{V}^*$ defined on \mathbb{U}^* with values in the algebraic dual space \mathbb{V}^* of some vector space \mathbb{V} .

Lemma 3.4.2 (Abstract dual equality) *We assume that:*

- (a) Φ is a convex lower $\sigma(\mathbb{U}, \mathbb{U}^*)$ -semicontinuous function and there is some $\ell_o \in \mathbb{U}^*$ such that for all $u \in \mathbb{U}$, $\Phi(u) \geq \Phi(0) + \langle \ell_o, u \rangle_{\mathbb{U}^*, \mathbb{U}}$;
- (b) Φ^* has $\sigma(\mathbb{U}^*, \mathbb{U})$ -compact level sets: $\{\ell \in \mathbb{U}^* : \Phi^*(\ell) \leq a\}$, $a \in \mathbb{R}$;
- (c) The algebraic adjoint A^\dagger of A satisfies $A^\dagger \mathbb{V} \subset \mathbb{U}$.

Then, the dual equality

$$\inf \{ \Phi^*(\ell); \ell \in \mathbb{U}^*, A\ell = v^* \} = \sup_{v \in \mathbb{V}} \left\{ \langle v, v^* \rangle_{\mathbb{V}, \mathbb{V}^*} - \Phi(A^\dagger v) \right\} \in (-\infty, \infty] \quad (3.17)$$

holds true for any $v^* \in \mathbb{V}^*$.

Proof

◁ [Proof of Lemma 3.4.2] In the special case where $\Phi(0) = 0$ and $\ell_o = 0$, this result is [Léo01b, Thm. 2.3]. Considering $\Psi(u) := \Phi(u) - [\Phi(0) + \langle \ell_o, u \rangle]$, $u \in \mathbb{U}$, we see that $\inf \Psi = \Psi(0) = 0$, Ψ is a convex lower $\sigma(\mathbb{U}, \mathbb{U}^*)$ -semicontinuous function and $\Psi^*(\ell) = \Phi^*(\ell_o + \ell) + \Phi(0)$, $\ell \in \mathbb{U}^*$, has $\sigma(\mathbb{U}^*, \mathbb{U})$ -compact level sets. As Ψ^* satisfies the assumptions of [Léo01b, Thm. 2.3], we have the dual equality

$$\inf \{ \Psi^*(\ell); \ell \in \mathbb{U}^*, A\ell = v^* - A\ell_o \} = \sup_{v \in \mathbb{V}} \left\{ \langle v, v^* - A\ell_o \rangle_{\mathbb{V}, \mathbb{V}^*} - \Psi(A^\dagger v) \right\} \in [0, \infty]$$

which is (3.17). ▷

Proof

◁ [Proof of Theorem 3.4.1] Let us denote

$$R_{\mu_0}(\cdot) := \int_{\mathbb{R}^n} R_x(\cdot) \mu_0(dx) \in \mathcal{P}(\Omega).$$

R_{μ_0} is a measure on paths, its initial marginal, as a probability measure in \mathbb{R}^n , is $R_{\mu_0,0} = \mu_0$. When $\mu_0 = m$, we have $R_m = R$. If U is a functional on paths,

$$\begin{aligned} E_{R_{\mu_0}}(U) &= \int E_R(U|X_0 = x) \mu_0(dx) = \int E_R(U|X_0 = x) \frac{d\mu_0}{dR_0}(x) R_0(dx) \\ &= E_R(E_R(U \frac{d\mu_0}{dR_0} | X_0)) = E_R(E_R(U \frac{d\mu_0}{dm}(X_0) | X_0)) = E_R(U \frac{d\mu_0}{dm}(X_0)). \end{aligned}$$

So $R_{\mu_0}(\cdot) = \frac{d\mu_0}{dm}(X_0) R(\cdot)$, we see that for any $P \in \mathcal{P}(\Omega)$ such that $P_0 = \mu_0$,

$$H(P|R) = H(\mu_0|m) + H(P|R_{\mu_0}).$$

Consequently, the minimizer of (3.8) is also the minimizer of

$$H(P|R_{\mu_0}) \rightarrow \min; \quad P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \quad (3.18)$$

and

$$\vec{A}(\mu_0, \mu_1) = \inf \{ H(P|R_{\mu_0}), P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \}. \quad (3.19)$$

Therefore, all we have to prove is

$$\inf\{H(P|R_{\mu_0}), P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} = \sup\left\{\int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}^R \psi d\mu_0, \psi \in C_b(\mathbb{R}^n)\right\}.$$

This is an application of Lemma 3.4.2 with $\mathbb{U} = C_b(\Omega)$, $\mathbb{V} = C_b(\mathbb{R}^n)$ and

$$\begin{aligned} \Phi(u) &= \int_{\mathbb{R}^n} \log\left(\int_{\Omega} e^u dR^x\right) \mu_0(dx), \quad u \in C_b(\Omega), \\ A^\dagger \psi &= \psi(X_1) \in C_b(\Omega), \quad \psi \in C_b(\mathbb{R}^n). \end{aligned}$$

Let $C_b(\Omega)'$ be the topological dual space of $(C_b(\Omega), \|\cdot\|)$ equipped with the uniform norm $\|u\| := \sup_{\Omega} |u|$. It is shown at [Léo12a, Lemma 4.2] that for any $\ell \in C_b(\Omega)'$,

$$\Phi^*(\ell) = \begin{cases} H(\ell|R_{\mu_0}), & \text{if } \ell \in \mathcal{P}(\Omega) \text{ and } (X_0)_\# \ell = \mu_0 \\ +\infty, & \text{otherwise} \end{cases} \quad (3.20)$$

But according to [Léo01a, Lemma 2.1], the effective domain $\{\ell \in C_b(\Omega)^* : \Phi^*(\ell) < \infty\}$ of Φ^* is a subset of $C_b(\Omega)'$. Hence, for any ℓ in the algebraic dual $C_b(\Omega)^*$ of $C_b(\Omega)$, $\Phi^*(\ell)$ is given by (3.20). The assumption (c) of Lemma 3.4.2 on A^\dagger is obviously satisfied. Let us show that Φ and Φ^* satisfy the assumptions (a) and (b).

Let us start with (a). It is a standard result of the large deviation theory that $u \mapsto \log \int_{\Omega} e^u dR^x$ is convex (a consequence of Hölder's inequality). It follows that Φ is also convex. As Φ is upper bounded on a neighborhood of 0 in $(C_b(\Omega), \|\cdot\|)$:

$$\sup_{u \in C_b(\Omega), \|u\| \leq 1} \Phi(u) \leq 1 < \infty \quad (3.21)$$

(note that Φ is increasing and $\Phi(1) = 1$) and its effective domain is the whole space $\mathbb{U} = C_b(\Omega)$, it is $\|\cdot\|$ -continuous everywhere. Since Φ is convex, it is also lower $\sigma(C_b(\Omega), C_b(\Omega)')$ -semicontinuous and a fortiori lower $\sigma(C_b(\Omega), C_b(\Omega)^*)$ -semicontinuous. Finally, a direct calculation shows that $\ell_o = R_{\mu_0}$ is a subgradient of Φ at 0. This completes the verification of (a). The assumption (b) is also satisfied because the upper bound (3.21) implies that the level sets of Φ^* are $\sigma(C_b(\Omega)^*, C_b(\Omega))$ -compact, see [Léo01a, Cor. 2.2]. So far, we have shown that the assumptions of Lemma 3.4.2 are satisfied. It remains to show that $A\ell = v^*$ corresponds to the final marginal constraint. Since $\{\Phi^* < \infty\}$ consists of probability measures, it is enough to specify the action of A on the vector subspace $\mathcal{M}_b(\Omega) \subset C_b(\Omega)^*$ of all bounded measures on Ω . For any $Q \in \mathcal{M}_b(\Omega)$ and any $\psi \in C_b(\Omega)$, we have

$$\langle \psi, AQ \rangle_{C_b(\mathbb{R}^n), C_b(\mathbb{R}^n)^*} = \langle A^\dagger \psi, Q \rangle_{C_b(\Omega), C_b(\Omega)^*} = \int_{\Omega} \psi(X_1) dQ = \int_{\mathbb{R}^n} \psi dQ_1.$$

This means that for any $Q \in \mathcal{M}_b(\Omega)$, $AQ = Q_1 \in \mathcal{M}_b(\mathbb{R}^n)$. With these considerations, choosing $v^* = \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ in (3.17) leads us to

$$\begin{aligned} &\inf\{H(Q|R_{\mu_0}); Q \in \mathcal{P}(\Omega) : Q_0 = \mu_0, Q_1 = \mu_1\} \\ &= \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \log \langle e^{\psi(X_1)}, R^x \rangle \mu_0(dx) \right\} \end{aligned}$$

which is the desired identity. \triangleright

Remark 3.4.3 Alternatively, considering $R^y := R^{\mu_1}(\cdot | X_1 = y)$, for m -almost all $x \in \mathbb{R}^n$ and

$$R^{\mu_1}(\cdot) := \int_{\mathbb{R}^n} R^y(\cdot) \mu_1(dy) \in \mathcal{P}(\Omega)$$

we would obtain a formulation analogous to (3.3) for the backward entropic cost $\overleftarrow{\mathcal{A}}$ (Definition 2.3.1).

Remark 3.4.4 We did not use any specific property of the Kolmogorov semigroup. The dual equality can be generalized, without changing its proof, to any reference path measure $R \in \mathcal{P}(\Omega)$ on any Polish state space \mathcal{X} .

3.5 Benamou-Brenier formulation of the entropic cost

We derive some analogue of the Benamou-Brenier formulation (3.4) for the entropic cost.

Theorem 3.5.1 (Benamou-Brenier formulation of the entropic cost) Let V, μ_0 and μ_1 be such that hypothesis (Exi), (Reg1) and (Reg2) stated in Section 3.3 are satisfied. We have

$$\overrightarrow{\mathcal{A}}(\mu_0, \mu_1) = H(\mu_0 | m) + \inf_{(\nu, v)} \int_{\mathbb{R}^n \times [0,1]} \frac{|v_t(z)|^2}{2} \nu_t(dz) dt, \quad (3.22)$$

where the infimum is taken over all $(\nu_t, v_t)_{0 \leq t \leq 1}$ such that, $\nu_t(dz)$ is identified with its density with respect to Lebesgue measure $\nu(t, z) := d\nu_t/dz$, satisfying $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and the following continuity equation

$$\partial_t \nu + \nabla \cdot (\nu [v - \nabla(V + \log \nu)/2]) = 0, \quad (3.23)$$

is satisfied in a weak sense.

Moreover, these results still hold true when the infimum in (3.22) is taken among all (ν, v) satisfying (3.23) and such that v is a gradient vector field, that is

$$v_t(z) = \nabla \psi_t(z), \quad 0 \leq t \leq 1, z \in \mathbb{R}^n,$$

for some function $\psi \in C^\infty([0, 1] \times \mathbb{R}^n)$.

Remark 3.5.2

(1) The continuity equation (3.23) is the linear Fokker-Planck equation

$$\partial_t \nu + \nabla \cdot (\nu [v - \nabla V/2]) - \Delta \nu/2 = 0.$$

Its solution $(\nu_t)_{0 \leq t \leq 1}$, with v considered as a known parameter, is the time marginal flow $\nu_t = P_t$ of a weak solution $P \in \mathcal{P}(\Omega)$ (if it exists) of the stochastic differential equation

$$dX_t = [v_t(X_t) - \nabla V(X_t)/2] dt + dW_t^P, \quad P\text{-a.s.}$$

where W^P is a P -Brownian motion, $P_0 = \mu_0$ and $(X_t)_{1 \leq t \leq 1}$ is the canonical process.

(2) Clearly, one can restrict the infimum in the identity (3.22) to (ν, v) such that

$$\int_{\mathbb{R}^n \times [0,1]} |v_t(z)|^2 \nu_t(dz) dt < \infty.$$

Proof

◁ Because of (3.9) and (3.19), all we have to show is

$$\inf\{H(P|R_{\mu_0}); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} = \inf_{(\nu, v)} \int_{\mathbb{R}^n \times [0,1]} \frac{|v_t(z)|^2}{2} \nu_t(dz) dt,$$

where (ν, v) satisfies (3.23), $\nu_0 = \mu_0$ and $\nu_1 = \mu_1$. As R_{μ_0} is Markov, by [Léo14b, Prop. 2.10] we can restrict the infimum to the set of all Markov measures $P \in \mathcal{P}(\Omega)$ such that $P_0 = \mu_0, P_1 = \mu_1$ and $H(P|R_{\mu_0}) < \infty$. For each such Markov measure P , Girsanov's theorem introduced in Section 1.1.3 (see for instance [Léo12b, Thm. 2.1] for a proof related to the present setting) states that there exists a measurable vector field $\beta_t^P(z)$ such that

$$dX_t = [\beta_t^P(X_t) - \nabla V(X_t)/2] dt + dW_t^P, \quad P\text{-a.s.}, \quad (3.24)$$

where W^P is a P -Brownian motion. Moreover, β^P satisfies $E_P \int_0^1 |\beta_t^P|^2(X_t) dt < \infty$ and

$$H(P|R_{\mu_0}) = \frac{1}{2} \int_{\mathbb{R}^n \times [0,1]} |\beta_t^P|^2(z) P_t(dz) dt. \quad (3.25)$$

For any P with $P_0 = \mu_0, H(\mu_0|m) < \infty$ and $H(P|R_{\mu_0}) < \infty$, we have $P_t \ll R_t = m \ll \mathcal{L}$ for all t . Taking $\nu = (P_t)_{0 \leq t \leq 1}$ and $v = \beta^P$, the stochastic differential equation (3.24) gives (3.23) and optimizing the left hand side of (3.25) leads us to

$$\inf\{H(P|R_{\mu_0}); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \leq \inf_{(\nu, v)} \int_{\mathbb{R}^n \times [0,1]} \frac{|v_t(z)|^2}{2} \nu_t(dz) dt.$$

On the other hand, it is proved in [Zam86, Léo14b] that the solution P of the Schrödinger problem (3.18) is such that (3.24) is satisfied with $\beta_t^P(z) = \nabla \psi_t(z)$ where ψ is given in (3.13). This completes the proof of the theorem. ▷

Corollary 3.5.3 *Let V, μ_0 and μ_1 be such that the hypotheses stated in Section 3.3 are satisfied. We have*

$$\mathcal{A}(\mu_0, \mu_1) = \inf_{(\rho, v)} \int_{\mathbb{R}^n \times [0,1]} \left(\frac{1}{2} |v_t(z)|^2 + \frac{1}{8} |\nabla \log \rho_t(z)|^2 \right) \rho_t(z) m(dz) dt, \quad (3.26)$$

where the infimum is taken over all $(\rho_t, v_t)_{0 \leq t \leq 1}$ such that $\rho_0 m = \mu_0, \rho_1 m = \mu_1$ and the following continuity equation

$$\partial_t \rho + e^V \nabla \cdot (e^{-V} \rho v) = 0 \quad (3.27)$$

is satisfied in a weak sense.

Moreover, these results still hold true when the infimum in (3.26) is taken among all (ν, v) satisfying (3.27) and such that v is a gradient vector field, that is

$$v_t(z) = \nabla \theta_t(z), \quad 0 \leq t \leq 1, z \in \mathbb{R}^n,$$

for some function $\theta \in C^\infty([0, 1] \times \mathbb{R}^n)$.

Remark 3.5.4 *The density ρ in the statement of the corollary must be understood as a density $\rho = dv/dm$ with respect to the reversing measure m . Indeed, with $\nu(t, z) = dv_t/dz$, we see that $\nu = e^{-V}\rho$ and the evolution equation (3.27) writes as the current equation $\partial_t \nu + \nabla \cdot (\nu v) = 0$.*

This result was proved recently by Chen, Georgiou and Pavon in [CGP16] in the case where $V = 0$ without any mention to gradient type vector fields. The present proof is essentially the same as in [CGP16]: we take advantage of the time reversal invariance of the relative entropy $H(\cdot|R)$ with respect to the reversible path measure R .

Proof

\triangleleft The proof follows almost the same line as Theorem 3.5.1's one. The additional ingredient is the time-reversal invariance (3.10): $H(P|R) = H(P^*|R)$. Let $P \in \mathcal{P}(\Omega)$ be the solution of (3.8). We have already noted that P^* is the solution of the Schrödinger problem where the marginal constraints μ_0 and μ_1 are inverted. We obtain

$$\begin{aligned} dX_t &= v_t^P(X_t) dt + dW_t^P, & P\text{-a.s.} \\ dX_t &= v_t^{P^*}(X_t) dt + dW_t^{P^*}, & P^*\text{-a.s.} \end{aligned}$$

where W^P and W^{P^*} are respectively Brownian motions with respect to P and P^* and

$$\begin{aligned} H(P|R) &= H(\mu_0|m) + E_P \frac{1}{2} \int_0^1 |\beta_t^P(X_t)|^2 dt \\ H(P^*|R) &= H(\mu_1|m) + \frac{1}{2} E_{P^*} \int_0^1 |\beta_t^{P^*}(X_t)|^2 dt = H(\mu_1|m) + \frac{1}{2} E_P \int_0^1 |\beta_{1-t}^{P^*}(X_t)|^2 dt \end{aligned}$$

with $v^P = -\nabla V/2 + \beta^P$ and $v^{P^*} = -\nabla V/2 + \beta^{P^*}$. Taking the half sum of the above equations, the identity $H(P|R) = H(P^*|R)$ implies that

$$H(P|R) = \frac{1}{2} [H(\mu_0|m) + H(\mu_1|m)] + \frac{1}{4} E_P \int_0^1 (|\beta_t^P|^2 + |\beta_{1-t}^{P^*}|^2) dt.$$

Let us recall the definition of current velocities of P and P^* introduced in Section 1.1.2

$$\begin{aligned} v_t^{cu,P}(z) &:= v_t^P(z) - \frac{1}{2} \nabla \log \nu_t^P(z) = \beta_t^P(z) - \frac{1}{2} \nabla \log \rho_t^P(z) \\ v_t^{cu,P^*}(z) &:= v_t^{P^*}(z) - \frac{1}{2} \nabla \log \nu_t^{P^*}(z) = \beta_t^{P^*}(z) - \frac{1}{2} \nabla \log \rho_t^{P^*}(z) \end{aligned}$$

where for any $0 \leq t \leq 1, z \in \mathbb{R}^n$,

$$\nu_t^P(z) := \frac{dP_t}{dz}, \quad \rho_t^P(z) := \frac{dP_t}{dm}(z) \quad \text{and} \quad \nu_t^{P^*}(z) := \frac{dP_t^*}{dz}, \quad \rho_t^{P^*}(z) := \frac{dP_t^*}{dm}(z).$$

The naming *current velocity* is justified by the current equations

$$\begin{aligned} \partial_t \nu^P + \nabla \cdot (\nu^P v^{cu,P}) &= 0 \quad \text{and} \quad \partial_t \rho^P + e^V \nabla \cdot (e^{-V} \rho^P v^{cu,P}) = 0, \\ \partial_t \nu^{P^*} + \nabla \cdot (\nu^{P^*} v^{cu,P^*}) &= 0 \quad \text{and} \quad \partial_t \rho^{P^*} + e^V \nabla \cdot (e^{-V} \rho^{P^*} v^{cu,P^*}) = 0. \end{aligned}$$

To see that the first equation $\partial_t \nu^P + \nabla \cdot (\nu^P v^{cu,P}) = 0$ is valid, remark that ν^P satisfies the Fokker-Planck equation (3.23) with v replaced by β^P . The equation for ρ^P follows immediately and the equations for ν^{P^*} and ρ^{P^*} are derived similarly.

The very definition of P^* implies that $\rho_t^{P^*} = \rho_{1-t}^P$ and the time reversal invariance $R^* = R$ implies that

$$v_t^{cu, P^*}(z) = -v_{1-t}^{cu, P}(z), \quad 0 \leq t \leq 1, z \in \mathbb{R}^n.$$

Therefore, $\beta_{1-t}^{P^*} = -v_t^{cu, P} + \frac{1}{2}\nabla \log \rho_t^P$ and $\frac{1}{4}(|\beta_t^P|^2 + |\beta_{1-t}^{P^*}|^2) = \frac{1}{2}|v_t^{cu, P}|^2 + \frac{1}{8}|\nabla \log \rho_t^P|^2$. This completes the proof of the first statement of the corollary.

For the second statement about $v = \nabla\theta$, remark that as in Theorem 3.5.1's proof, the solution P of the Schrödinger problem is such that $\beta^P = \nabla\psi$ for some smooth function ψ . One concludes with $v^{cu, P} = \beta^P - \frac{1}{2}\nabla \log \rho^P$, by taking $\theta = \psi - \log \sqrt{\rho^P}$. \triangleright

Remark 3.5.5

- (1) The current velocity $v^{cu, P}$ of a diffusion path measure P has been introduced by Nelson in [Nel67] together with its osmotic velocity $v^{os, P} := \frac{1}{2}\nabla \log \rho^P$.
- (2) Up to a multiplicative factor, $\int_{\mathbb{R}^n} |\nabla \log \rho_t(z)|^2 \rho_t(z) m(dz)$ is the entropy production or Fischer information. The average osmotic action is $A^{os}(P) := \int_{\mathbb{R}^n \times [0,1]} \frac{1}{2}|v^{os, P}|^2 dP_t dt = \int_{\mathbb{R}^n \times [0,1]} \frac{1}{8}|\nabla \log \rho|^2 \rho dm dt$ and it is directly connected to a variation of entropy. It's worth remarking that by considering the dilatation in time of the reference path measure as introduced in Remark 3.3.7, the osmotic action vanishes in the limit for $\varepsilon \rightarrow 0$. Let us define now the osmotic cost

$$I^{os}(\mu_0, \mu_1) := \inf\{A^{os}(P); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\}$$

and the current cost $I^{cu}(\mu_0, \mu_1) := \inf_{(\rho, v)} \int_{\mathbb{R}^n \times [0,1]} \frac{1}{2}|v_t(z)|^2 \rho_t(z) m(dz) dt$ where the infimum runs through all the (ρ, v) satisfying (3.27). The standard Benamou-Brenier formula precisely states that $I^{cu}(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1)/2$. Therefore, Corollary 3.5.3 implies that

$$\mathcal{A}(\mu_0, \mu_1) \geq \frac{1}{2}W_2^2(\mu_0, \mu_1) + I^{os}(\mu_0, \mu_1).$$

In particular, by the positivity of the entropic cost I^{os} we obtain the following relation between the entropic and Wasserstein costs:

Corollary 3.5.6 *Let V, μ_0 and μ_1 be such that the hypotheses stated in Section 3.3 are satisfied. We have*

$$\mathcal{A}(\mu_0, \mu_1) \geq \frac{1}{2}W_2^2(\mu_0, \mu_1).$$

3.6 Contraction with respect to the entropic cost

The analogy between optimal transport and minimal entropy can also be observed in the context of contractions.

As explained in the introduction, contraction in Wasserstein distance depends on the curvature. Even if there are actually many contraction inequalities in Wasserstein distance, we focus here on two main results. The first one depends on the curvature and the second one includes the dimension. These results can be written for more general semigroups satisfying the curvature-dimension condition as defined in the Bakry-Émery theory.

In the context of the Kolmogorov semigroup of Section 3.2 with a generator given by (3.6) in \mathbb{R}^n , the two main contraction inequalities can be formulated as follows.

- Let assume that for some real κ , we have $\text{Hess}(V) \geq \kappa \text{Id}$ in the sense of symmetric matrices. Then for any f, g probability densities with respect to the measure m and any $t \geq 0$,

$$W_2(T_t f m, T_t g m) \leq e^{-\frac{\kappa}{2}t} W_2(f m, g m). \quad (3.28)$$

Let us recall that this result was proved in [SvR09] in the general context of Riemannian manifold. Although in the context of Kolmogorov semigroups the proof is easy, its generalization for the entropic cost to a Riemannian setting is not trivial.

- When $L = \Delta/2$ that is $V = 0$, the heat equation in \mathbb{R}^n satisfies the following dimension dependent contraction property:

$$W_2^2(T_t f \mathcal{L}, T_s g \mathcal{L}) \leq W_2^2(f \mathcal{L}, g \mathcal{L}) + n(\sqrt{t} - \sqrt{s})^2, \quad (3.29)$$

for any $s, t \geq 0$ and any f, g probability densities with respect to the Lebesgue measure \mathcal{L} . This contraction was proved in a more general context in [BGL15, Kuw15].

The two inequalities (3.28) and (3.29) can be proved in terms of entropic cost. Let us choose the reference path measure R associated with the potential V and take $\varepsilon, u > 0$ and $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$. In order to extend for each $u, \varepsilon > 0$ the dual formulation for the entropic cost of Theorem 3.4.1, consider the semigroup $(T_{\varepsilon u t})_{t \geq 0}$ and the corresponding path measure $R^{\varepsilon u}$: time is dilated by the factor $(\varepsilon u)^{-1}$. Theorem 3.4.1 implies that

$$\mathcal{A}^{R^{\varepsilon u}}(\mu_0, \mu_1) = H(\mu_0 | m) + \sup \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \log T_{\varepsilon u}(e^\psi) d\mu_0, \quad \psi \in C_b(\mathbb{R}^n) \right\}.$$

Now by changing ψ with ψ/ε we see that

$$\varepsilon \mathcal{A}^{R^{\varepsilon u}}(\mu_0, \mu_1) = \varepsilon H(\mu_0 | m) + \sup \left\{ \int_{\mathbb{R}^n} \psi d\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}_u^\varepsilon \psi d\mu_0, \quad \psi \in C_b(\mathbb{R}^n) \right\} \quad (3.30)$$

where for any $\psi \in C_b(\mathbb{R}^n)$,

$$\mathcal{Q}_u^\varepsilon \psi = \varepsilon \log T_{\varepsilon u}(e^{\psi/\varepsilon}). \quad (3.31)$$

For simplicity, we denote $\varepsilon \mathcal{A}^{R^{\varepsilon u}} = \mathcal{A}_u^\varepsilon$ and $\mathcal{A}_1^\varepsilon = \mathcal{A}^\varepsilon$.

As explained in Remark 3.3.7, we have

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}_u^\varepsilon(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1)/2u. \quad (3.32)$$

The entropic cost associated to the Kolmogorov semigroup has the following properties.

Theorem 3.6.1 (Contraction in entropic cost) *Let $\varepsilon > 0$ be fixed.*

(a) *If V satisfies $\text{Hess}(V) \geq \kappa \text{Id}$ for some $\kappa \in \mathbb{R}$, then for any $t \geq 0$,*

$$\mathcal{A}_b^\varepsilon(T_{u t(b)} f m, T_t g m) \leq \mathcal{A}_{v_t(b)}^\varepsilon(f m, g m) + \varepsilon [H(T_{u t(b)} f m | m) - H(f m | m)], \quad (3.33)$$

where f, g are probability densities with respect to m , and

$$u_t(b) = t + \frac{1}{\kappa} \log \left(\frac{e^{-\varepsilon \kappa b}}{1 + e^{\kappa t} (e^{-\varepsilon \kappa b} - 1)} \right) \quad v_t(b) = -\frac{1}{\kappa \varepsilon} \log(1 + e^{\kappa t} (e^{-\varepsilon \kappa b} - 1)) \quad (3.34)$$

where: if $\kappa \leq 0$, $b \in (0, \infty)$ and if $\kappa > 0$, $b \in (0, -\frac{1}{\kappa \varepsilon} \log(1 - e^{-\kappa t}))$.

(b) If $V = 0$ then for any $t \geq 0$,

$$\mathcal{A}^\varepsilon(T_t f m, T_s g m) \leq \mathcal{A}^\varepsilon(f m, g m) + \frac{n}{2}(\sqrt{t} - \sqrt{s})^2 + \varepsilon[H(T_t f m|m) - H(f m|m)].$$

The proof of this theorem relies on the following commutation property between the Markov semigroup T_t and the semigroup $\mathcal{Q}_t^\varepsilon$ defined at (3.31). Let us notice that the second statement of next lemma was proved in [BGL15, Section 5].

Lemma 3.6.2 (Commutation property) *Let $s, t \geq 0$, $\varepsilon > 0$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be any bounded measurable function.*

(a) If $\text{Hess}(V) \geq \kappa \text{Id}$ for some real κ , then

$$\mathcal{Q}_{v_t(b)}^\varepsilon(T_t f) \leq T_{u_t(b)}(\mathcal{Q}_b^\varepsilon f) \quad (3.35)$$

where for each $t \geq 0$, the numbers $u_t(b)$, $v_t(b)$ and b are given in (3.34).

Moreover, for ε small enough and $t > 0$ fixed, (3.35) is valid for all b positive.

(b) If $V = 0$ then

$$\mathcal{Q}_1^\varepsilon(T_t f) \leq T_s(\mathcal{Q}_1^\varepsilon f) + \frac{n}{2}(\sqrt{t} - \sqrt{s})^2.$$

Proof

◁ We only have to prove the first statement (a). Let us define for each $s \leq t$ the function

$$\Lambda(s) = T_\alpha \mathcal{Q}_\beta^\varepsilon(T_{t-s} f)$$

with $\alpha : [0, t] \rightarrow [0, \infty)$ an increasing function such that $\alpha(0) = 0$, and $\beta : [0, t] \rightarrow [0, \infty)$ and we call $\beta(t) = b$. Setting $g = \exp(T_{t-s} f / \varepsilon)$, using the chain rule for the diffusion operator L we obtain

$$\begin{aligned} \Lambda'(s) &= \varepsilon T_\alpha \left[\alpha' L \log T_{\varepsilon\beta} g + \frac{1}{T_{\varepsilon\beta} g} T_{\varepsilon\beta} (\varepsilon \beta' L g - g L \log g) \right] \\ &= \varepsilon T_\alpha \left[\alpha' \left(\frac{L T_{\varepsilon\beta} g}{T_{\varepsilon\beta} g} - \frac{|\nabla T_{\varepsilon\beta} g|^2}{2(T_{\varepsilon\beta} g)^2} \right) + \frac{1}{T_{\varepsilon\beta} g} T_{\varepsilon\beta} \left(\varepsilon \beta' L g - L g + \frac{|\nabla g|^2}{2g} \right) \right] \\ &= \varepsilon T_\alpha \left[\frac{1}{T_{\varepsilon\beta} g} \left(L T_{\varepsilon\beta} g (\alpha' + \varepsilon \beta' - 1) + T_{\varepsilon\beta} \left(\frac{|\nabla g|^2}{2g} \right) - \alpha' \frac{|\nabla T_{\varepsilon\beta} g|^2}{2T_{\varepsilon\beta} g} \right) \right] \\ &\geq \varepsilon T_\alpha \left[\frac{1}{T_{\varepsilon\beta} g} \left(L T_{\varepsilon\beta} g (\alpha' + \varepsilon \beta' - 1) + \frac{1}{2} T_{\varepsilon\beta} \left(\frac{|\nabla g|^2}{g} \right) (1 - e^{-\kappa\varepsilon\beta} \alpha') \right) \right] \end{aligned}$$

where the last inequality is given by the commutation,

$$\frac{|\nabla T_t g|^2}{T_t g} \leq e^{-\kappa t} T_t \left(\frac{|\nabla g|^2}{g} \right)$$

which is implied by the condition $\text{Hess}(V) \geq \kappa \text{Id}$ (see for instance [BGL14, Section 3.2]). If the following conditions on α and β hold

$$\begin{cases} \alpha' + \varepsilon \beta' - 1 = 0 \\ 1 - e^{-\kappa\varepsilon\beta} \alpha' = 0, \end{cases} \quad (3.36)$$

we have $\Lambda'(s) \geq 0$ for each $0 \leq s \leq t$. In particular $\Lambda(0) \leq \Lambda(t)$ for each $t \geq 0$, that is

$$\mathcal{Q}_{v_t(b)}^\varepsilon(T_t f) \leq T_{u_t(b)}(\mathcal{Q}_b^\varepsilon f)$$

where $v_t(b) = \beta(0)$ and $u_t(b) = \alpha(t)$. Finally solving system (3.36) together with the conditions $\alpha(0) = 0$, $\beta(t) = b$, we can compute the explicit formulas for v and u as in statement (a). In particular, substituting α' in the second equation of the system and integrating from 0 to t we obtain the following relation

$$e^{-\varepsilon\kappa\beta(0)} = 1 + e^{\kappa t}(e^{-\varepsilon\kappa b} - 1). \quad (3.37)$$

If we assume for a while that the term on the right hand side is positive, we obtain

$$\beta(0) = -\frac{1}{\kappa\varepsilon} \log(1 + e^{\kappa t}(e^{-\varepsilon\kappa b} - 1))$$

and

$$\alpha(t) = t + \frac{1}{\kappa} \log\left(\frac{e^{-\varepsilon\kappa b}}{1 + e^{\kappa t}(e^{-\varepsilon\kappa b} - 1)}\right).$$

Let us study now the sign of the right hand side in (3.37).

- If $\kappa \leq 0$, it is positive for each $b \in \mathbb{R}$;
- If $\kappa > 0$ in order to be positive, we need the condition for b ,

$$b < -\frac{1}{\varepsilon\kappa} \log(1 - e^{-\kappa t}) := b_0.$$

Finally let us consider the case when $\varepsilon > 0$ is small. From (3.37) we obtain the relation,

$$\beta(0) = be^{\kappa t} + o(\varepsilon)$$

for each $\kappa \in \mathbb{R}$ and b positive.

This completes the proof of the lemma. \triangleright

Proof

\triangleleft [Proof of Theorem 3.6.1] The proof is based on the dual formulation stated in Theorem 3.4.1. Let $\psi \in C_b(\mathbb{R}^n)$, by Lemma 3.6.2 under the condition $\text{Hess}(V) \geq \kappa \text{Id}$ and by time reversibility,

$$\begin{aligned} \int_{\mathbb{R}^n} \psi T_t g \, dm - \int_{\mathbb{R}^n} \mathcal{Q}_b^\varepsilon \psi T_{u_t(b)} f \, dm &= \int_{\mathbb{R}^n} T_t \psi g \, dm - \int_{\mathbb{R}^n} T_{u_t(b)} \mathcal{Q}_b^\varepsilon \psi f \, dm \\ &\leq \int_{\mathbb{R}^n} T_t \psi g \, dm - \int_{\mathbb{R}^n} \mathcal{Q}_{v_t(b)}^\varepsilon T_t \psi f \, dm \\ &\leq \mathcal{A}_{v_t(b)}^\varepsilon(fm, gm) - \varepsilon H(fm|m). \end{aligned}$$

Finally taking the supremum over $\psi \in C_b(\mathbb{R}^n)$ we obtain the desired inequality in (i). The same argument can be used to prove the contraction property in (ii), applying the second commutation inequality in Lemma 3.6.2. \triangleright

Remark 3.6.3 *Let observe that if $\kappa < 0$, the function $\beta(s)$, for $s \in [0, t]$, is decreasing, while for $\kappa > 0$ it is increasing and if $\kappa = 0$ it is the constant function $\beta(t) = b$. In particular by choosing $b = 1$ (3.35) writes as follows*

$$\mathcal{Q}_1^\varepsilon(T_t f) \leq T_t(\mathcal{Q}_1^\varepsilon f).$$

Remark 3.6.4 *Lemma 3.6.2 can be proved in the general context of a Markov diffusion operator under the Bakry-Émery curvature-dimension condition. Its application to more general problems is actually a working paper of the third author.*

Remark 3.6.5 *Let us point out two converse assertions.*

- (1) *The contraction in entropic cost in Theorem 3.6.1 implies back the contraction in Wasserstein cost. Indeed, under the assumptions of Section 3.3, it can be easily checked that when $\varepsilon \rightarrow 0$, we have $u(t) \rightarrow t$ and $v(t) \rightarrow be^{\kappa t}$. Therefore, with (3.32) and (3.33), one recovers (3.28). Analogous arguments can be applied to recover the contraction of the Wasserstein cost (3.29) when $V = 0$.*
- (2) *The commutation property in Lemma 3.6.2 implies back the convexity of the potential V . This can be seen by differentiating (3.35) with respect to b around 0. We believe also that for $\varepsilon > 0$ fixed, inequality (3.33) implies back the convexity of the potential.*

3.7 Examples

In this section we will compute explicitly the results discussed in the previous sections, between two given measures. We first compute the Wasserstein cost, its dual and Benamou-Brenier formulations and the displacement interpolation, as exposed in the introduction. Then, we will do the same for the entropic cost, taking in consideration two different reference path measures R . In particular, we will compute (3.30), for $u = 1$ and $\varepsilon > 0$ and look at the behavior in the limit $\varepsilon \rightarrow 0$ to recover the classical results of the optimal transport. For abuse of notation we will denote with μ_t both the interpolation and its density with respect to the Lebesgue measure dx . We introduce for Gaussian measures the following notation: for any $m \in \mathbb{R}^n$ and $v \in \mathbb{R}$, the density with respect to the Lebesgue measure of $\mathcal{N}(m, v^2)$ is given by

$$(2\pi v^2)^{-n/2} \exp\left(-\frac{|x - m|^2}{2v^2}\right),$$

As marginal measures we consider for $x_0, x_1 \in \mathbb{R}^n$

$$\mu_0(x) := \mathcal{N}(x_0, 1), \quad \mu_1(x) := \mathcal{N}(x_1, 1). \quad (3.38)$$

Note that the entropic interpolation between two Dirac measures δ_x and δ_y should be the Bridge R^{xy} between x and y with respect to the reference measure R . But unfortunately $H(\delta_x|m), H(\delta_y|m) = \infty$, hence we consider only marginal measures with a density with respect to m .

3.7.1 Wasserstein cost

The Wasserstein cost between μ_0, μ_1 as in (3.38), is

$$W_2^2(\mu_0, \mu_1) = d(x_0, x_1)^2.$$

In its dual formulation, the supremum is reached by the function

$$\psi(x) = (x_1 - x_0)x$$

and in the Benamou-Brenier formulation the minimizer vector field is

$$v^{MC} = x_1 - x_0$$

The displacement or McCann interpolation is given by

$$\mu_t^{MC}(x) = \mathcal{N}(x_t, 1) \tag{3.39}$$

where $x_t = (1 - t)x_0 + tx_1$. In other words using the push-forward notation (3.5),

$$\mu_t^{MC}(x) = (\hat{x}_t^{MC})_{\#}\mu_0$$

with $\hat{x}_t^{MC}(x) := (1 - t)x + t(x + x_1 - x_0)$ a trajectory whose associated velocity field is $v^{MC} = x_1 - x_0$.

3.7.2 Schrödinger cost

Heat semigroup

As a first example we consider on the state space \mathbb{R}^n the Heat (or Brownian) semigroup, that corresponds to the case $V = 0$ in our main example in Section 3.2, whose infinitesimal generator is the Laplace operator $L = \Delta/2$ and the reversing reference measure is the Lebesgue measure dx . Since we are interested in the ε -entropic interpolation, with $\varepsilon > 0$, we take in consideration the Heat semigroup with a dilatation in time, whose density kernel is given by

$$p_t^\varepsilon(x, y) = (2\pi\varepsilon t)^{-n/2} \exp\left(-\frac{|x - y|^2}{2\varepsilon t}\right)$$

i.e. $p_t^\varepsilon(x, y) = \mathcal{N}(y, \varepsilon t)$ for $t > 0, (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$.

- The entropic interpolation (3.12) is

$$\mu_t^\varepsilon(x) = \mathcal{N}(x_t, D_t^\varepsilon) \tag{3.40}$$

where x_t is like in (3.39) and $D_t^\varepsilon : [0, 1] \rightarrow \mathbb{R}^+$ is a polynomial function given by

$$D_t^\varepsilon = \alpha^\varepsilon t(1 - t) + 1$$

with $\alpha^\varepsilon = \delta^2/(1 + \delta)$ where $\delta = (\varepsilon - 2 + \sqrt{4 + \varepsilon^2})/2$. We observe that D_t^ε is such that $D_0 = D_1 = 1$ with a maximum in $t = 1/2$ for each $\varepsilon > 0$, (see Figure 3.1).

It's worth to point out how we managed to derive an explicit formula for the entropic interpolation. The key point is the resolution of the Schrödinger system (3.14) that in our example writes as

$$\begin{cases} (2\pi)^{-n/2} \exp\left(-\frac{|x-x_0|^2}{2}\right) &= f(x) \int g(y)(2\pi\epsilon t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2\epsilon t}\right) dy \\ (2\pi)^{-n/2} \exp\left(-\frac{|x-x_1|^2}{2}\right) &= g(x) \int f(y)(2\pi\epsilon t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2\epsilon t}\right) dy. \end{cases}$$

By taking f and g exponential functions of the type

$$e^{a_2 x^2 + a_1 x + a_0} \quad \text{with} \quad a_0, a_1, a_2 \in \mathbb{R}$$

we can solve the system explicitly by determining the coefficients of f and g .

One can also express the entropic interpolation through the *push-forward* notation as introduced at (3.16), $\mu_t^\epsilon = (\hat{x}_t^\epsilon)_\# \mu_0$ where

$$\hat{x}_t^\epsilon(x) = \sqrt{D_t^\epsilon}(x - x_0) + x_t.$$

Furthermore \hat{x}_t^ϵ satisfies the differential equation

$$\dot{\hat{x}}_t^\epsilon = v^{cu,\epsilon}(\hat{x}_t^\epsilon) \tag{3.41}$$

where $v^{cu,\epsilon}$ is the current velocity, and is given by

$$v^{cu,\epsilon} = \frac{\dot{D}_t^\epsilon}{2D_t^\epsilon}(x - x_t) + x_1 - x_0.$$

It can be finally verified that the entropic interpolation (3.40) satisfies the PDE

$$\dot{\mu}_t^\epsilon + \nabla \cdot (\mu_t^\epsilon v^{cu,\epsilon}) = 0. \tag{3.42}$$

Remark 3.7.1 *Let observe that if $x_0 = x_1$, μ_t^ϵ is not constant in time, unlike the McCann interpolation.*

- Denoting $P \in \mathcal{P}(\Omega)$ the path measure whose flow is given by (3.40) and that minimizes $H(\cdot|R)$, the entropic cost between μ_0, μ_1 as in (3.39) is

$$\mathcal{A}_u^\epsilon(\mu_0, \mu_1) = H(P|R).$$

The easiest way to compute this quantity is to use the Benamou-Brenier formulation in Section 3.5. The resulting formula has not a nice and interesting form, therefore we don't report it explicitly.

- In the dual formulation proved in Section 3.3, the supremum is reached by the function $\psi \in C_b(\mathbb{R}^n)$ given, up to a constant term, by

$$\psi_t(x) = -\frac{1}{2} \frac{\delta}{1 + \delta(1-t)} x^2 - \frac{1}{2} \frac{\gamma}{1 + \delta(1-t)} x \tag{3.43}$$

with δ as in (3.40) and $\gamma = 2[x_0(1 + \delta) - x_1]$.

- In the Benamou-Brenier formulation in Theorem 3.5.1 the minimizer vector field is

$$v^H = \nabla \psi_t.$$

where ψ_t is given by (3.43) and $\nabla \psi_t$ represents the forward velocity. It can be easily verified that the equation

$$\partial_t \mu_t + \nabla \cdot \left(\mu_t \left[\nabla \psi_t - \frac{\nabla \mu_t}{2\mu_t} \right] \right) = 0$$

is satisfied.

Ornstein-Uhlenbeck semigroup

As a second example, we consider on the state space \mathbb{R}^n the Ornstein-Uhlenbeck semigroup, that corresponds to the case $V = |x|^2/2$ for the Kolmogorov semigroup in Section 3.2, whose infinitesimal generator is given by $L = (\Delta - x \cdot \nabla)/2$ and the invariant measure is the standard Gaussian in \mathbb{R}^n . Here again we consider the kernel representation with a dilatation in time; in other words, for $\varepsilon > 0$, the kernel with respect to the Lebesgue measure is given by

$$p_t(x, y) = (2\pi(1 - e^{-\varepsilon t}))^{-n/2} \exp\left(-\frac{|y - xe^{-\varepsilon t/2}|^2}{2(1 - e^{-\varepsilon t})}\right)$$

i.e. $p_t(x, y) = \mathcal{N}(xe^{-\varepsilon t/2}, 1 - e^{-\varepsilon t})$.

- The entropic interpolation (3.12) is given by

$$\mu_t^\varepsilon(x) = \mathcal{N}(m_t, D_t^\varepsilon) \quad (3.44)$$

where $m_t = a_t[(e^{-\varepsilon t/2} - e^{-\varepsilon(1-t)/2})x_0 + (e^{-\varepsilon(1-t)/2} - e^{-\varepsilon(1+t)/2})x_1]$ with

$$a_t := \frac{1 + \delta - \delta e^{-\varepsilon}}{(1 - e^{-\varepsilon})[\delta(1 + \delta)(e^{-\varepsilon t} + e^{-\varepsilon(1-t)}) - 2\delta^2 e^{-\varepsilon}]}$$

with δ as in (3.45), and $D_t^\varepsilon : [0, 1] \rightarrow \mathbb{R}^+$ defined as

$$D_t^\varepsilon := -1 + 2(1 - e^{-\varepsilon})a_t$$

satisfying, as in the case of the Heat semigroup, $D_0^\varepsilon = D_1^\varepsilon = 1$.

Furthermore, we have $\mu_t^\varepsilon = (\hat{x}_t^\varepsilon)_\# \mu_0$ where

$$\hat{x}_t^\varepsilon := \sqrt{D_t^\varepsilon}(x - x_0) + m_t.$$

It can be verified that equations (3.42) and (3.41) hold true also in the Ornstein-Uhlenbeck case, with the current velocity given by

$$v_{cu}^\varepsilon = \frac{\dot{D}_t^\varepsilon}{2D_t^\varepsilon}(x - x_0) + \dot{m}_t$$

- The entropic cost between μ_0, μ_1 can be computed as in the Heat semigroup case by

$$\mathcal{A}_u^\varepsilon(\mu_0, \mu_1) = H(P|R)$$

where P is the path measure associated to the flow (3.44) which minimizes $H(\cdot|R)$.

- In the dual formulation at Section 3.3, the supremum is reached, up to a constant term, by the function

$$\psi_t(x) = -\frac{1}{2} \frac{\varepsilon \delta e^{-\varepsilon(1-t)}}{1 + \delta(1 - e^{-\varepsilon(1-t)})} x^2 + \frac{\varepsilon \gamma e^{-\varepsilon(1-t)/2}}{1 + \delta(1 - e^{-\varepsilon(1-t)})} x \quad (3.45)$$

where $\delta = (e^{-\varepsilon} - \sqrt{e^{-2\varepsilon} - e^{-\varepsilon} + 1})/(e^{-\varepsilon} - 1)$ and $\gamma = (x_0 e^{-\varepsilon/2} - x_1(1 + \delta - \delta e^{-\varepsilon}))/(1 - e^{-\varepsilon})$.

- In the Benamou-Brenier formulation (Theorem 3.5.1) the minimizer vector field is

$$v^{OU} = \nabla \psi_t.$$

Remark 3.7.2 *Let observe that both in the Heat and Ornstein-Uhlenbeck cases, if we take the limit $\varepsilon \rightarrow 0$ of the entropic interpolation, of the velocities v^H, v^{OU} and of the function ψ_t , we recover the respective results for the Wasserstein cost stated in Subsection 3.7.1.*

In the following figures we refer to the McCann interpolation with a dotted line, the Heat semigroup with a dashed line and the Ornstein-Uhlenbeck semigroup with a continuous line. We fix $\varepsilon = 1$ and consider marginal measures in one dimension.

- Figure 3.1 represents the variance of the three interpolations, independent from the initial and final means x_0, x_1 .
- Figures 3.2 and 3.3 correspond to the mean in the three cases respectively with the initial and final means symmetric w.r.t the origin, and for any means. It's worth to remark from these images that the McCann interpolation and the entropic interpolation in the case of the heat semigroup, have the same mean.
- Finally figures 3.4 and 3.5 represent the three interpolations at time $t = 0, 1/2, 1$ respectively with different marginal data, as before.

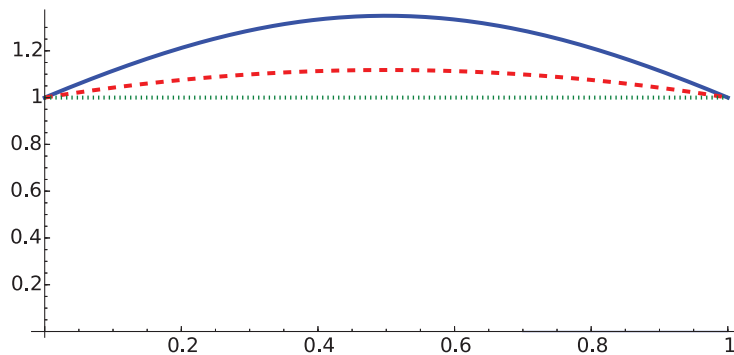


Figure 3.1: Variance, $\varepsilon = 1$

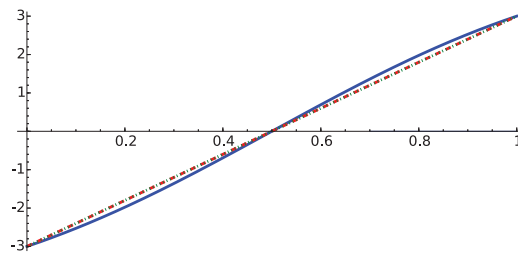
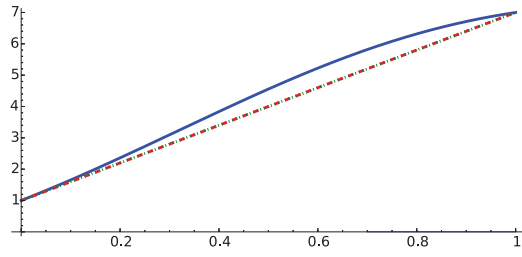
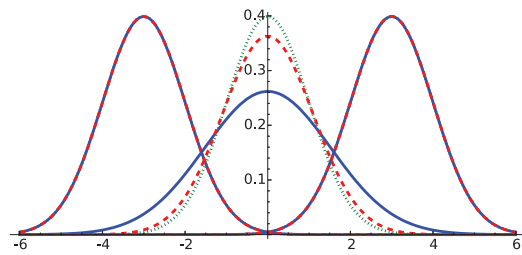
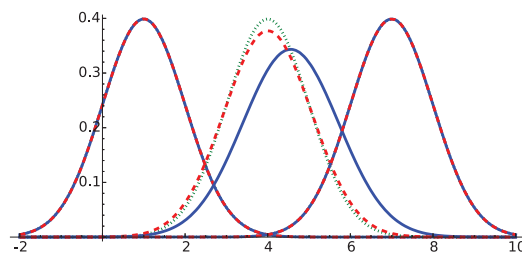


Figure 3.2: Mean with $x_0 = -3, x_1 = 3, \varepsilon = 1$

Figure 3.3: Mean $x_0 = 1, x_1 = 7, \varepsilon = 1$ Figure 3.4: Interpolations at time $t = 0, 1/2, 1, x_0 = -3, x_1 = 3, \varepsilon = 1$ Figure 3.5: Interpolations at time $t = 0, 1/2, 1, x_0 = 1, x_1 = 7, \varepsilon = 1$

3.7.3 Numerical examples

The question of determining explicit solutions f, g of the Schrödinger system (3.14) is not trivial. Explicit computations, as those presented in Section 3.7.2, are possible in very few cases. Just think to the trivial case of the heat semigroup as reference, and two uniform probability measures as marginals: despite trivial, we cannot determine the two functions f and g *by hands*. With the precious help of my PhD colleague Maxime Herda, we propose here two further examples via numerical computations. In both cases we choose the heat semigroup on the real line as reference measure and play with the choice of the marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$.

The implemented algorithm is the following. The first step is to discretize the Schrödinger system (3.14). Given a space interval $I \subset \mathbb{R}$, we consider the discretization $(x_i)_{i \in \{1, \dots, N\}}$ of I . We denote by $f_i \approx f(x_i)$ and $g_i \approx g(x_i)$ and the same for $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R})$. Then the Schrödinger system (3.14) writes as,

$$\mu_{0,i} = f_i \sum_{j=1}^N G_{ij} g_j \quad \mu_{1,i} = g_i \sum_{j=1}^N G_{ij} f_j \quad (3.46)$$

for all $i = 1, \dots, N$. Here $G_{ij} = \Delta x G(x_i - x_j)$ and G is the Gaussian kernel. System (3.46) can be reformulated as,

$$f_i = \frac{\mu_{0,i}}{\sum_{j=1}^N G_{ij} \left(\frac{\mu_{1,j}}{\sum_{k=1}^N G_{jk} f_k} \right)}, \quad g_i = \frac{\mu_{1,i}}{\sum_{j=1}^N G_{ij} \left(\frac{\mu_{0,j}}{\sum_{k=1}^N G_{jk} g_k} \right)}. \quad (3.47)$$

Note that (3.46) and (3.47) are equivalent up to a constant factor. Indeed for any solution (f, g) of (3.46), $(\lambda f, \lambda g)$ is a solution of (3.47) for any $\lambda \in \mathbb{R}$. Any confusion will be avoided by a re-normalization to unit mass of the terms on the right hand side in (3.46). It is a fixed point problem of the form $f = \mathcal{F}_{01}(f)$ and $g = \mathcal{F}_{10}(g)$, where the applications \mathcal{F}_{01} and \mathcal{F}_{10} are given at (3.47). The maps \mathcal{F}_{01} and \mathcal{F}_{10} are contractions on the positive cone \mathbb{R}_+^n with respect to the Hilber projective metric,

$$d_H(f, g) := \log \left(\frac{\max_i f_i/g_i}{\min_i f_i/g_i} \right).$$

The Sinkhorn algorithm [Cut13] consists in iterating \mathcal{F}_{01} and \mathcal{F}_{10} to converge to the solutions f and g of (3.46). Once the solutions f, g of the Schrödinger system are determined, the entropic interpolation is defined according to (2.12). Figures 3.6 and 3.8 show respectively two choices of marginal measures. In both cases we set $I = [-5, 5]$ and $N = 1000$. Moreover, by considering the Gaussian kernel with variance $\varepsilon > 0$, we can see in Figures 3.7 and 3.9 the behavior of the entropic interpolation when ε decreases to zero. We stopped at $\varepsilon = 0.05$ since the algorithm is very unstable when $\varepsilon > 0$ is too small as pointed out in [BCC⁺15].

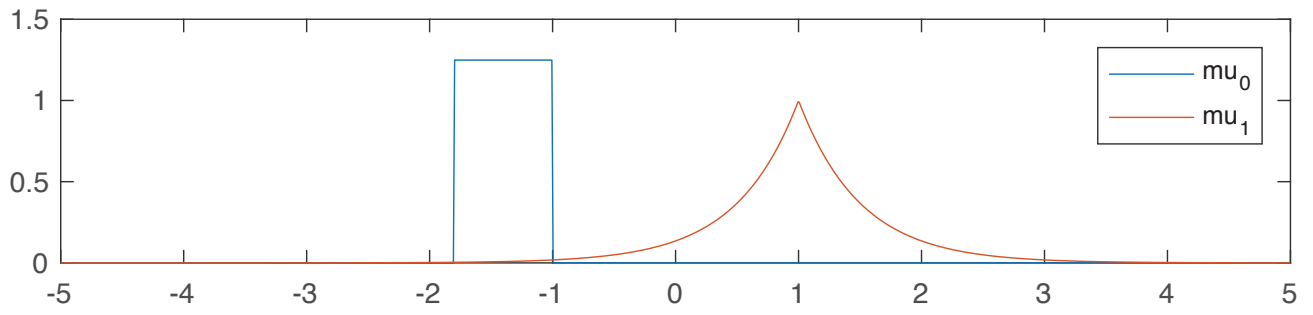


Figure 3.6: Marginals μ_0, μ_1

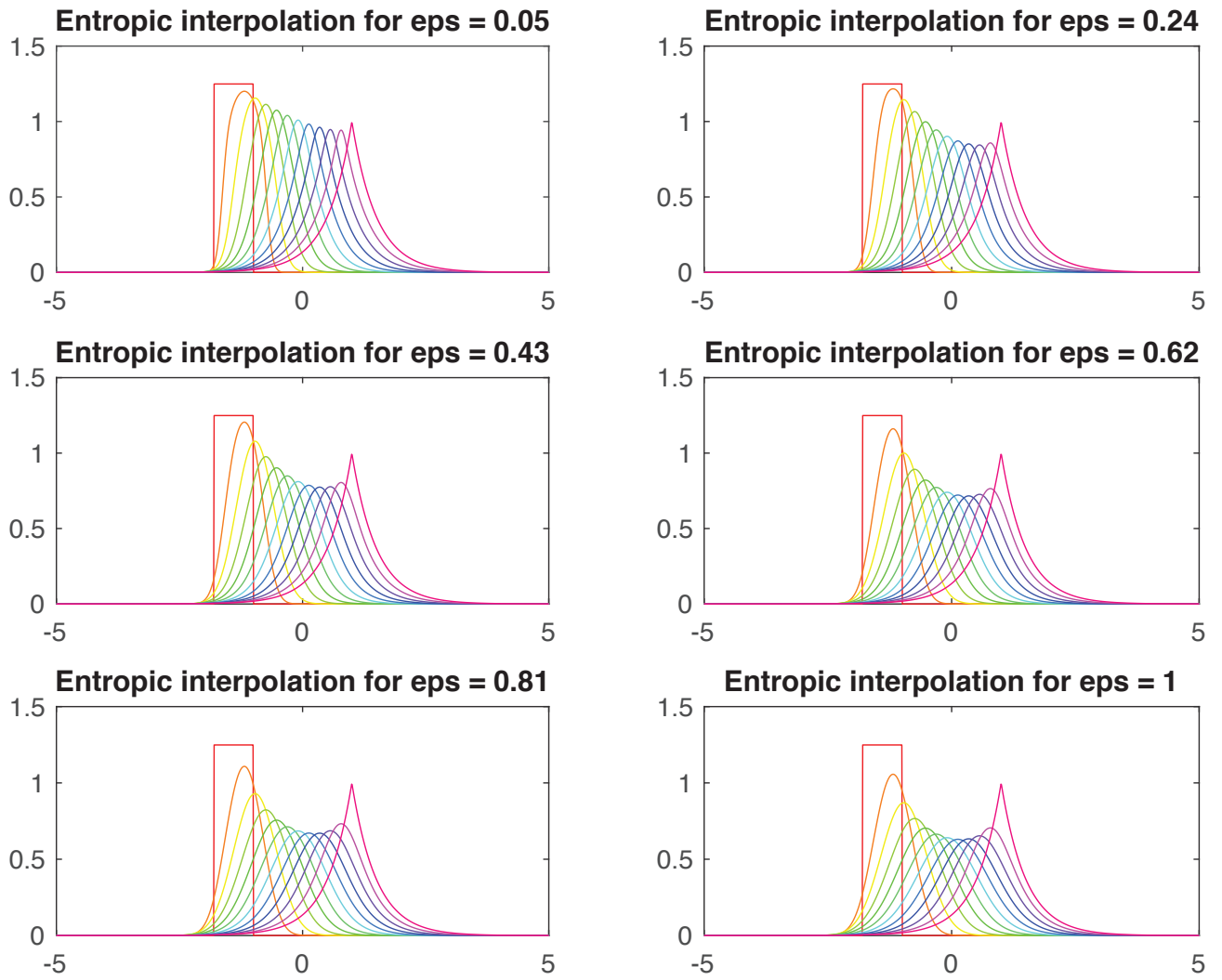
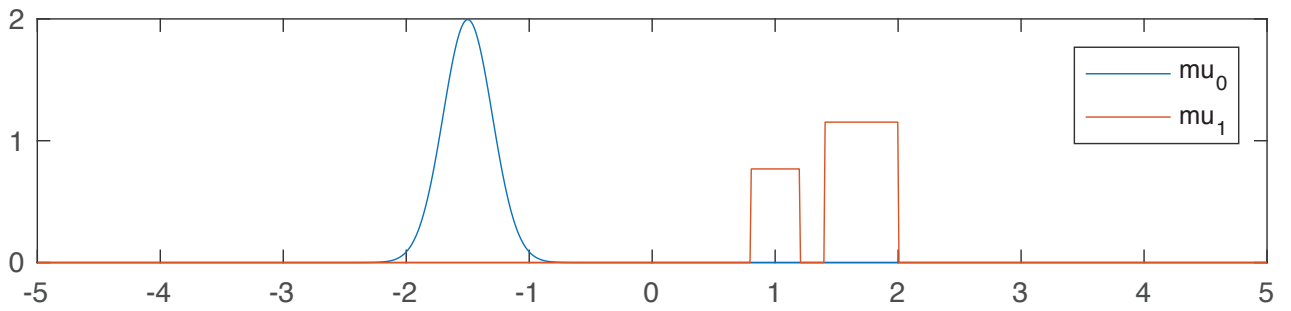
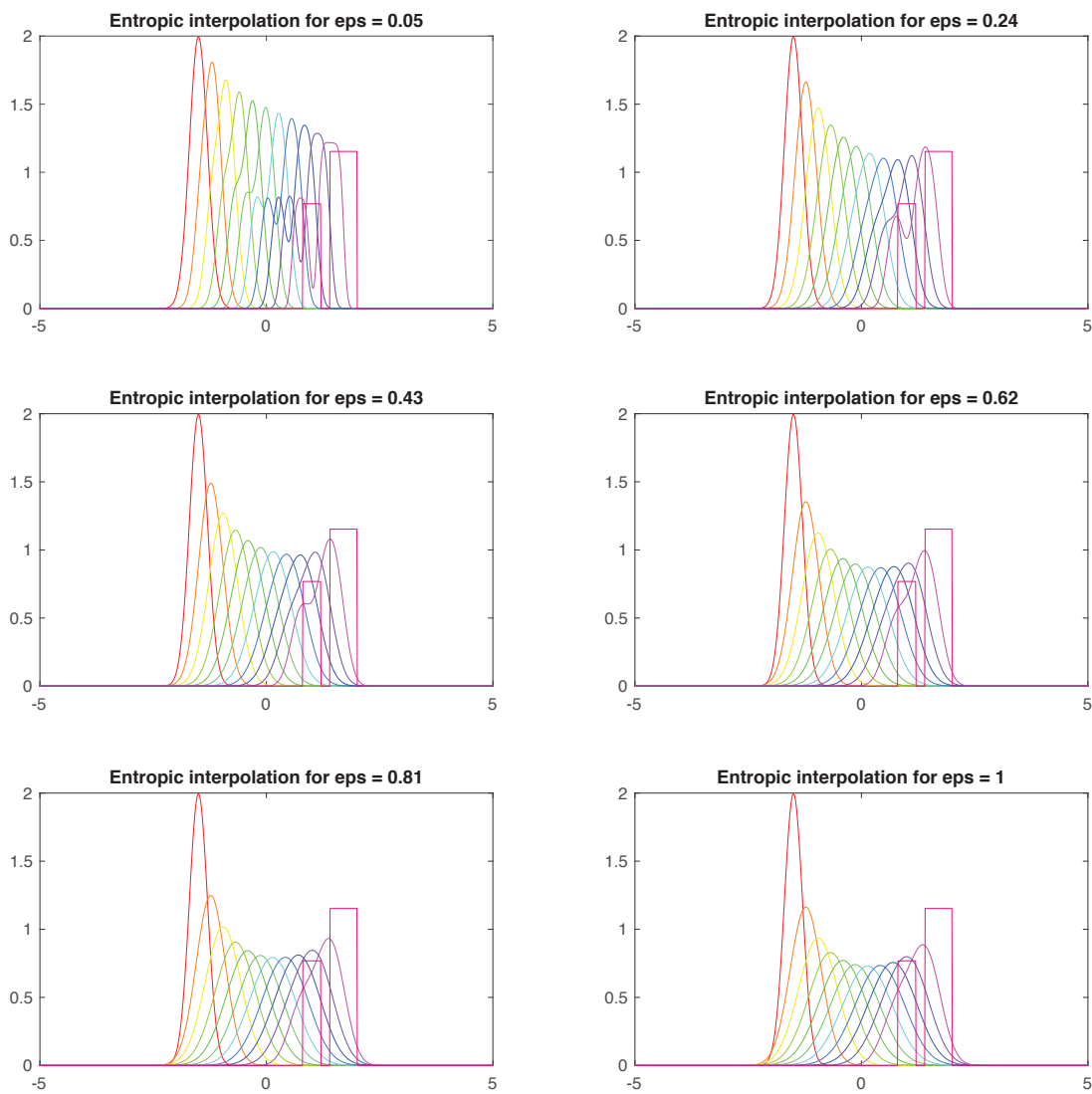


Figure 3.7: Interpolations for different values of ε

Figure 3.8: Marginals μ_0, μ_1 Figure 3.9: Interpolations at for different values of ε

Chapter 4

Convexity and regularity properties under $CD(0, n)$

In this chapter we report a slightly modified version of [Rip17]. Sections 4.1 and 4.2 contain repetitions of the previous chapters, hence the reader who already went over them can start over from Section 4.3 which contains the main results.

Abstract In this paper we prove a convexity property of the relative entropy along entropic interpolations (solutions of the Schrödinger problem), and a regularity property of the entropic cost along the heat flow. Then we derive a dimensional EVI inequality and a contraction property for the entropic cost along the heat flow. As a consequence, we recover the equivalent results in the Wasserstein space, proved by Erbar, Kuwada and Sturm

Résumé Dans cet article nous démontrons une propriété de convexité de l'entropie relative le long des interpolations entropiques (solutions du problème de Schrödinger), et une propriété de régularité du coût entropique le long du flot de la chaleur. Ensuite, nous en déduisons une inégalité EVI dimensionnelle et une propriété de contraction pour le coût entropique le long du flot de la chaleur. En conséquence, nous retrouvons les résultats équivalents dans l'espace de Wasserstein, démontrés par Erbar, Kuwada et Sturm.

4.1 Introduction

Convexity of the entropy along evolutionary equations is a powerful tool to prove regularity properties, asymptotic behavior, etc. We extend and compare some main and fruitful results around convexity of the entropy in the Wasserstein space, to the context of the Schrödinger problem.

For simplicity, results are presented in \mathbb{R}^n associated with the Lebesgue measure \mathcal{L} . Generalization can be stated in the context of a n -dimensional Riemannian manifold (M^n, g) .

We consider the relative entropy functional, loosely defined for any couple of positive measures μ, ν on \mathbb{R}^n as

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

whenever the integral is meaningful. A rigorous definition is stated at Section 4.2.

Context of the Wasserstein space

The quadratic Monge-Kantorovich distance between two probability measures $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n) = \{\mu \in \mathcal{P}(\mathbb{R}^n) : \int |x|^2 d\mu < \infty\}$, is defined as

$$W_2^2(\mu_0, \mu_1) := \inf_{\pi} \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \pi(dxdy) \right\},$$

where the infimum is running over all the couplings π of μ_0 and μ_1 , namely, all the probability measures $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals μ_0 and μ_1 , that is for any bounded measurable functions φ and ψ on \mathbb{R}^n , $\int [\varphi(x) + \psi(y)] \pi(dxdy) = \int \varphi d\mu_0 + \int \psi d\mu_1$.

The space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ is geodesic. This means that for any couple $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ there is a path $(\mu_s^{MC})_{s \in [0,1]}$ in $\mathcal{P}_2(\mathbb{R}^n)$ such that,

$$W_2(\mu_s^{MC}, \mu_t^{MC}) = |t - s| W_2(\mu_0, \mu_1) \quad \forall s, t \in [0, 1].$$

• The first result in this context is the convexity of the entropy along geodesics, i.e. if $(\mu_s^{MC})_{s \in [0,1]}$ is a geodesic in $\mathcal{P}_2(\mathbb{R}^n)$, then the map

$$[0, 1] \ni s \mapsto H(\mu_s^{MC} | \mathcal{L})$$

is convex. The entropy is then *displacement convex* in the sense introduced by McCann in [McC97]. This was a breakthrough, a starting point of the Lott-Sturm-Villani theory, who defined the (positive) curvature in a *metric measure space* (mms space), see [LV09, Stu06b]. This is usually noted the $CD(0, \infty)$ condition.

• Taking into account the dimension, the main progress is proposed by Erbar-Kuwada-Sturm in [EKS15] who proved the stronger result, under the same assumption, that the map

$$[0, 1] \ni s \mapsto N(\mu_s^{MC}) := \exp\left(-\frac{1}{n} H(\mu_s^{MC} | \mathcal{L})\right),$$

is concave. This condition can be used to define the $CD(0, n)$ condition in a mms space. This result has two main applications.

→ First a dimensional *evolution variational inequality* (EVI) for the quadratic Monge-Kantorovich distance, that writes for all $t \geq 0$ as,

$$\frac{d^+}{dt} W_2^2(T_t u \mathcal{L}, v \mathcal{L}) \leq n \left(1 - e^{-\frac{1}{n} [H(v|\mathcal{L}) - H(T_t u|\mathcal{L})]}\right), \quad (4.1)$$

for any $u, v \in \mathcal{P}_2(\mathbb{R}^n)$ where $(T_t)_{t \geq 0}$ is the heat semi-group in \mathbb{R}^n . This inequality is actually equivalent to say that the heat flow is the gradient flow associated to the entropy functional [AGS08].

→ Then, it provides a proof of the dimensional contraction with respect to the W_2 distance, for any $u, v \in \mathcal{P}_2(\mathbb{R}^n)$ and any $\tau > 0$,

$$W_2^2(T_\tau u \mathcal{L}, T_\tau v \mathcal{L}) \leq W_2^2(u, v) - 4n \int_0^\tau \sinh^2\left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{2n}\right) dt, \quad (4.2)$$

result proved in [BGG16, BGGK16].

Context of the Schrödinger problem

It will be properly defined in Section 4.2. Roughly speaking, the *entropic cost* associated to the Schrödinger problem, is, up to a constant term, the minimization problem,

$$\mathcal{A}(\mu_0, \mu_1) = \inf\{H(P|R) ; P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0 \text{ and } P_1 = \mu_1\} \in (-\infty, \infty], \quad (4.3)$$

for a fixed reference measure $R \in \mathcal{M}_+(C([0, 1], \mathbb{R}^n))$, that we can consider to be the Brownian motion on \mathbb{R}^n with reversing measure the Lebesgue measure, and marginals μ_0, μ_1 in some restriction of the set \mathcal{P}_2 , that will be defined later at Section 4.2. If we denote $\hat{P} \in \mathcal{P}(\Omega)$ the minimizer of (4.3), the *entropic interpolation* is defined as its marginal flow, i.e.

$$\mu_t := \hat{P}_t = (X_t)_\# \hat{P} \in P(\mathbb{R}^n), \quad \text{for any } 0 \leq t \leq 1.$$

• The first result on the subject is due to Léonard [Léo17] who proved that for any entropic interpolation $(\mu_s)_{s \in [0, 1]}$, the map

$$[0, 1] \ni s \mapsto H(\mu_s|\mathcal{L}),$$

is convex, where \mathcal{L} denotes the Lebesgue measure on \mathbb{R}^n .

One particular entropic geodesic is the map $\mu_s = T_s f$, $s \in [0, 1]$, where f is a smooth probability density. This is one of the starting point of the Bakry-Émery-Ledoux theory to prove regularity, asymptotic behaviour etc. of diffusion Markov generators (see [BGL14]).

In this context, adding the dimension, again it is more convenient to look at the exponential entropy. In 1985, Costa [Cos85] proved that the map

$$\mathbb{R}^+ \ni s \mapsto N(T_s f) = \exp\left(-\frac{2}{n} H(T_s f|\mathcal{L})\right) \quad (4.4)$$

is concave. This is an important result in information theory, and it is also useful to prove some functional inequalities, for instance the dimensional log-Sobolev inequality as it is reported in [ABC⁺00, Ch. 10].

All these results have their counterparts in the more general case of Ricci curvature bounded from below by some $\kappa \in \mathbb{R}$. We refer to references for general cases, Riemannian manifolds or mms spaces, see for instance [Con17].

The aim of the paper is to complete the picture concerning convexity and regularity for the Schrödinger problem. First we prove a Costa type result for the entropic interpolation, that is for any entropic interpolation $(\mu_s)_{s \in [0, 1]}$, the map

$$[0, 1] \ni s \mapsto \exp\left(-\frac{1}{n} H(\mu_s|\mathcal{L})\right),$$

is concave, cf. Theorem 4.3.1. Note the absence of the factor 2 with respect to (4.4). The relation between these two expressions will be done at Remark 4.3.2. Secondly, we prove that for any u and v , probability densities in some space that will be specified later, the map

$$\mathbb{R}^+ \ni t \mapsto \mathcal{A}(T_t u \mathcal{L}, v \mathcal{L}),$$

is differentiable and,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(T_t u \mathcal{L}, v \mathcal{L}) = -\frac{1}{2} \left. \frac{d}{ds} \right|_{s=1} H(\mu_s|\mathcal{L}),$$

where $(\mu_s)_{s \in [0,1]}$ is the entropic interpolation between u and v , cf. Theorem 4.3.4.

From these two results, one can deduce an EVI inequality for the entropic cost (Corollary 4.3.6)

$$\frac{d}{dt} \mathcal{A}(f \mathcal{L}, T_t g \mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(f|\mathcal{L}) - H(T_t g|\mathcal{L})]} \right),$$

for any $t > 0$ and a dimensional contraction inequality (Corollary 4.3.8) of the entropic cost along the heat flow,

$$\mathcal{A}(T_\tau u \mathcal{L}, T_\tau v \mathcal{L}) - \mathcal{A}(u \mathcal{L}, v \mathcal{L}) \leq -n \int_0^\tau \sinh^2 \left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{2n} \right) dt.$$

In conclusion, this approach provides an easy and rigorous proof of the analogous results for the Wasserstein distance. In fact, if we introduce a parameter $\varepsilon > 0$ in the entropic cost as follows,

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf \{ \varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0 \text{ and } P_1 = \mu_1 \} - \frac{\varepsilon}{2} [H(\mu_0|m) + H(\mu_1|m)]$$

where R^ε is the reference path measure associated to the generator $L^\varepsilon = \varepsilon \Delta/2$. As proved in [Léo12a, Mik04] we have the convergence property,

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{W_2^2(\mu_0, \mu_1)}{2}.$$

Thanks to this result, it is an immediate consequence to derive (4.1) and (4.2) as limits respectively of EVI and the contraction for the ε -entropic cost.

The paper is organized as follows. In the next section, we give in full details the setting of the Schrödinger problem. In Section 4.3 we present and prove our main results. First, in Theorem 4.3.1 we prove concavity of the exponential entropy along entropic interpolations. Then, in Theorem 4.3.4 we prove the regularity property of the entropic cost. We derive from these two results the EVI inequality for the entropic cost (Corollary 4.3.6), the dimensional contraction along the heat flow (Corollary 4.3.8) and an integral form of EVI (Proposition 4.3.10). Finally, we deduce contraction in Wasserstein distance (Remark 4.3.9) and the classical EVI for the Wasserstein distance (Corollary 4.3.11).

4.2 Setting

In this section we fix notations and recall some definition and property of the main objects of our framework.

First we extend the definition of relative entropy to measures that are not necessary finite. Let $r \in \mathcal{M}_+(Y)$ be σ -finite. Then, it exists at least a measurable non negative function W such that $z_W = \int e^{-W} dr < \infty$. By defining the probability measure $dr^W = e^{-W} dr / z_W$, we can write, $H(p|r) = H(p|r^W) - \int W dp - \log z_W$. And it is well defined for any $p \in \mathcal{P}(Y)$ such that $\int W dp < \infty$. Therefore the *relative entropy* of a probability measure $p \in \mathcal{P}(Y)$ such that $\int W dp < \infty$, with respect to a positive σ -measure $r \in \mathcal{M}_+(Y)$ is defined by,

$$(-\infty, \infty] \ni H(p|r) = \begin{cases} \int \log \frac{dp}{dr} dp & \text{if } p \ll r \\ +\infty & \text{otherwise.} \end{cases} \quad (4.5)$$

For more details about relative entropy, conditional expectation and disintegration for unbounded measures, see [Léo14a] or Section 1.1.1.

A *reference path measure* $R \in \mathcal{M}_+(\Omega)$, is a positive measure on the set $\Omega = C([0, 1], Y)$. We fix the state space $Y = \mathbb{R}^n$, equipped with its Borel σ -field, and the path space Ω , with the canonical σ -field $\sigma(X_t; 0 \leq t \leq 1)$ generated by the canonical process for all $t \in [0, 1]$,

$$X_t(\omega) := \omega_t \in \mathbb{R}^n, \quad \omega = (\omega_s)_{0 \leq s \leq 1} \in \Omega.$$

Moreover, for any measure $Q \in \mathcal{M}_+(\Omega)$ and any $t \in [0, 1]$, we denote

$$Q_t(\cdot) := Q(X_t \in \cdot) = (X_t)_\# Q \in \mathcal{M}_+(\mathbb{R}^n).$$

In this chapter the reversing and Lebesgue measure coincide, hence without ambiguity by abuse of notation we will not distinguish between density functions and measures.

The reference path measure

In the sequel, as reference path measure, we consider the reversible Brownian motion R on the state space $Y = \mathbb{R}^n$, with generator $L = \Delta/2$ and initial condition $R_0(dx) := \mathcal{L}$ the Lebesgue measure on \mathbb{R}^n . Note that it is an unbounded measure since R has the same mass as R_0 (See [Léo14b]). In particular, we denote $(T_t)_{t \geq 0}$ the Markov semigroup associated to the generator L , which is defined for any bounded measurable function f and any $t \geq 0$ and $x \in \mathbb{R}^n$ by,

$$T_t f(x) = \int f(y) \frac{e^{-|x-y|^2/2t}}{(2\pi t)^{n/2}} dy. \tag{4.6}$$

Reversibility will play a crucial role at different points of our proofs. We recall that it is equivalent to say that for any couple of functions $f, g \in C_c^\infty(\mathbb{R}^n)$,

$$\int f T_t g dx = \int g T_t f dx.$$

In other words, let us define the time reversal mapping $X_t^* := X_{1-t}$, $0 \leq t \leq 1$ and $R^* = (X^*)_\# R$ the time-reversed of R . Reversibility means that $R^* = R$. Moreover it implies that R is \mathcal{L} -stationary, that is, $R_t = \mathcal{L}$, for all $0 \leq t \leq 1$, and this is equivalent to say that the generator is symmetric, that is for $f, g \in C_c^\infty(\mathbb{R}^n)$,

$$\int f L g dx = \int g L f dx.$$

It can be easily verified that the *carré du champ* operator, defined for any couple of functions $f, g \in C_c^\infty(\mathbb{R}^n)$, by $\Gamma(f, g) = [L(fg) - fLg - gLf]/2$, when it is associated to $L = \Delta/2$, is given by $\Gamma(f, g) = \nabla f \cdot \nabla g/2$. It satisfies the integration by parts formula,

$$\int f \Delta g dx = - \int \nabla f \cdot \nabla g dx = -2 \int \Gamma(f, g) dx. \tag{4.7}$$

Moreover the iterated *carré du champ* operator Γ_2 , defined for $f \in C_c^\infty(\mathbb{R}^n)$ by,

$$\Gamma_2(f) := \frac{1}{2} L(|\nabla f|^2) - \nabla f \cdot \nabla L f$$

when the generator $L = \Delta/2$, writes as $\Gamma_2(f) = ||\text{Hess} f||_2^2/4$ and by the Cauchy-Schwarz inequality, it yields to,

$$\Gamma_2(f) \geq \frac{1}{4n}(\Delta f)^2 = \frac{1}{n}(Lf)^2. \quad (4.8)$$

Inequality (4.8) is known under the name of $CD(0, n)$ condition, introduced by Bakry and Emery in [BE85], see also [BGL14]. For further details about this subject we refer to Section 1.1.4.

Entropic cost \mathcal{A}

In order to define the entropic cost, we need to fix, in addition to a reference path measure, two marginal probability measures on the state space \mathbb{R}^n . In particular we assume that the marginals,

$$\mu_0, \mu_1 \in \Pi := \left\{ \mu \in P(\mathbb{R}^n) : H(\mu|\mathcal{L}) < \infty, \int |x|^2 d\mu < \infty \right\}. \quad (4.9)$$

Under this assumption, the *entropic cost* is defined as

$$\mathcal{A}(\mu_0, \mu_1) := \inf\{H(P|R); P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0 \text{ and } P_1 = \mu_1\} - \frac{1}{2}[H(\mu_0|\mathcal{L}) + H(\mu_1|\mathcal{L})] \quad (4.10)$$

In order to show that the entropic cost is well defined and finite, it is more convenient to use the equivalent static definition. Let us consider the joint law of the initial and final position of the reversible Brownian motion R , that is,

$$R_{01}(dxdy) = \frac{e^{-|x-y|^2/2}}{(2\pi)^{n/2}} dxdy.$$

It is shown in [Léo14b, Prop. 2.3] (see Theorem 2.1.3) that the entropic cost can be defined equivalently as,

$$\mathcal{A}(\mu_0, \mu_1) = \inf\{H(\pi|R_{01}); \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) \text{ s.t. } \pi_0 = \mu_0, \pi_1 = \mu_1\} - \frac{1}{2}[H(\mu_0|\mathcal{L}) + H(\mu_1|\mathcal{L})]$$

where $\pi_0 := \pi(\cdot \times \mathbb{R}^n)$, and $\pi_1 := \pi(\mathbb{R}^n \times \cdot)$. The assumption on the marginals μ_0, μ_1 to have second order moment finite in (4.9), implies that the relative entropy with respect to R_{01} is bounded from below. To see this, it is enough to choose $W(x, y) = |x|^2 + |y|^2$ in the definition of relative entropy (4.5). Moreover, the assumption of finite relative entropy in (4.9) together with the fact that $R_{01}(dxdy) \geq e^{-|x|^2 - |y|^2} dxdy$, makes sure that $H(\mu_0 \otimes \mu_1|R_{01}) < \infty$, therefore $\mathcal{A}(\mu_0, \mu_1)$ is bounded also from above (See [Léo14b, Lemma 2.a, Prop. 2.5] and Section 2.1 for more details and the general case).

However, in order to enunciate rigorously our results we need some more restrictive assumption on the marginals. In particular, we will assume μ_0, μ_1 to be smooth and compactly supported probability measures. As a consequence of this assumption we have that $\mu_0, \mu_1 \in \Pi$. A stronger result in Theorem 4.3.4 is stated under stronger assumptions that will be specified later at...

Remark 4.2.1 Note that unlike the W_2^2 , \mathcal{A} is not the square of a distance. Though, by the reversibility of the reference measure R , \mathcal{A} is symmetric, that is, for all suitable μ_0, μ_1 ,

$$\mathcal{A}(\mu_0, \mu_1) = \mathcal{A}(\mu_1, \mu_0).$$

Indeed, since the time reversing mapping X^* (defined above) is one-to-one, it holds $H(P|R) = H(P^*|R^*)$. Moreover, since R is reversible, it implies that, $H(P|R) = H(P^*|R)$. Thus, if P is a minimizer in (4.10), then P^* is the minimizer of the same Schrödinger problem with switched marginals.

We introduce here a fluctuation parameter, that will allow us to link the entropic cost to the quadratic Wasserstein distance. Let R^ε be the law of the reversible Brownian motion with infinitesimal generator

$$L^\varepsilon = \frac{\varepsilon}{2}\Delta, \quad \text{for any } \varepsilon > 0.$$

Note that it doesn't change the dynamics, but it corresponds to a simple dilatation in time. We define the ε -entropic cost as,

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon); P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0, P_1 = \mu_1\} - \frac{\varepsilon}{2}[H(\mu_0|\mathcal{L}) + H(\mu_1|\mathcal{L})]. \quad (4.11)$$

Note the rescaling factor ε in front of the entropy, in order \mathcal{A}^ε not to explode to infinity in the limit when ε vanishes. It is shown in [Léo12a] via large deviation arguments, see also Section 2.4 that the ε -entropic cost is a regular approximation of the square of the quadratic Wasserstein distance, namely,

$$\lim_{\varepsilon \rightarrow 0} \mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{W_2^2(\mu_0, \mu_1)}{2}. \quad (4.12)$$

We only use the ε -entropic cost \mathcal{A}^ε to recover the classic results for W_2 at Remark 4.3.9 and Corollary 4.3.11 at the end of the chapter.

Entropic interpolations

We assume from now on, that the reference path measure R is associated to the heat semigroup as introduced at the beginning of this section, and $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$. Under these assumptions, by [Léo14b, Thm. 2.12], a unique minimizer exists, is called *entropic bridge* and is characterized by the formula,

$$\hat{P} = f(X_0)g(X_1)R \in \mathcal{P}(\Omega), \quad (4.13)$$

where f, g positive, are unique solutions of the Schrödinger system (cf. [Föl88]),

$$\begin{cases} \mu_0 &= fT_1g \\ \mu_1 &= gT_1f. \end{cases} \quad (4.14)$$

Here T_1 is the heat semigroup (4.6) at time $t = 1$, and again by abuse of notation we denote by μ_i also the density of the probability measure μ_i with respect to the Lebesgue measure, for $i = 0, 1$.

Remark 4.2.2 *As proved in [GT17, Thm. 3.1], and in [Tam17] in a non compact setting, the assumption on the marginal measures to have densities with respect to the Lebesgue measure in $C_c^\infty(\mathbb{R}^n)$, makes sure that the functions f, g solution of (4.14) are positive and L^∞ then $T_t f, T_{1-t} g \in C^\infty(\mathbb{R}^n)$ for any $0 \leq t \leq 1$. Thus,*

$$f = \frac{\mu_0}{T_1 g} \in C_c^\infty(\mathbb{R}^n)$$

since $T_1 g$ is smooth and strictly positive and $\mu_0 \in C_c^\infty(\mathbb{R}^n)$. The same is valid for g .

Definition 4.2.3 (Entropic interpolation) *The R -entropic interpolation between μ_0 and μ_1 is defined as the marginal flow of the minimizer (4.13), that is $\mu_t := \hat{P}_t = (X_t)_\# \hat{P} \in P(\mathbb{R}^n)$ for $0 \leq t \leq 1$. In particular, it is characterized by the formula,*

$$d\mu_t = T_t f T_{1-t} g \mathcal{L}, \quad \forall 0 \leq t \leq 1,$$

or equivalently,

$$d\mu_t = e^{\varphi_t} + \psi_t \mathcal{L}, \quad \forall 0 \leq t \leq 1,$$

where for any $t \in [0, 1]$, $\varphi_t := \log T_t f$ and $\psi_t := \log T_{1-t} g$ with f, g solutions of (4.14) and $(T_t)_{t \geq 0}$ the heat semigroup (4.6). Note that since by hypothesis f, g are $C_c^\infty(\mathbb{R}^n)$, then μ_t is in $C^\infty(\mathbb{R}^n)$ for all $0 \leq t \leq 1$.

The two functions φ_t and ψ_t satisfy respectively the Hamilton-Jacobi-Bellman equations,

$$\begin{cases} \partial_t \varphi_t - \frac{\Delta \varphi_t}{2} - \frac{|\nabla \varphi_t|^2}{2} = 0, & 0 \leq t \leq 1, \\ \varphi_0 = \log f & t = 0 \end{cases} \quad \begin{cases} \partial_t \psi_t + \frac{\Delta \psi_t}{2} + \frac{|\nabla \psi_t|^2}{2} = 0, & 0 \leq t \leq 1, \\ \psi_1 = \log g, & t = 1. \end{cases} \quad (4.15)$$

In analogy to the Kantorovich potentials, φ_0 and ψ_1 are often referred to as *Schrödinger potentials*. By adding the PDEs for φ and ψ in (4.15), we deduce that the entropic interpolation is a smooth solution of the transport equation,

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (\mu_t \nabla \theta_t) = 0, & \forall t \in (0, 1] \\ \mu_0 = \mu_0, & t = 0, \end{cases} \quad (4.16)$$

where $\nabla \theta_t = \nabla(\psi_t - \varphi_t)/2$. We briefly recall here the definitions of forward, backward, osmotic and current velocity introduced by Nelson in [Nel67] and explained in Sections 1.1.2 and 2.2, and how they are related,

$$\begin{cases} v_t^{cu} := \nabla \theta_t = \frac{\nabla \psi_t - \nabla \varphi_t}{2} \\ v_t^{os} := \frac{1}{2} \nabla \log \mu_t = \nabla \psi_t + \nabla \varphi_t \end{cases} \quad \begin{cases} \vec{v}_t := \nabla \psi_t = \frac{1}{2} \nabla \log \mu_t + \nabla \theta_t \\ \overleftarrow{v}_t := \nabla \varphi_t = -\nabla \theta_t + \frac{1}{2} \nabla \log \mu_t. \end{cases}$$

Moreover, the definition of P^* , the time reversal of the minimizer in (4.10), implies that, $\mu_t^* = \mu_{1-t} = T_{1-t} f T_t g$, for any $0 \leq t \leq 1$, and it establishes the following relation between the backward and the forward velocities respectively associated to μ_t and its time reversal μ_t^* ,

$$\overleftarrow{v}_t(x) = \vec{v}_{1-t}^*(x), \quad 0 \leq t \leq 1, x \in \mathbb{R}^n. \quad (4.17)$$

Dual and Benamou-Brenier formulations

Finally, we recall two equivalent formulations of the entropic cost, that will be crucial in the proof of our main results. First, the dual formulation, in analogy with the Kantorovich formulation for the Monge problem ([MT06], [GLR17, Section 4] and Chapter 3).

Theorem 4.2.4 (Dual Kantorovich formulation) *For $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$ defined at (4.25), then*

$$\vec{\mathcal{A}}(\mu_0, \mu_1) = \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int \psi d\mu_1 - \int Q_1 \psi d\mu_0 \right\}. \quad (4.18)$$

Here, $Q_1 \psi := \log T_1 e^\psi$, where T_1 is the heat semigroup at time $t = 1$. The supremum is achieved by the Schrödinger potential $\psi_1 = \log g$ that appears in (4.15).

We recall from Section 2.3 that the *forward entropic cost* is defined by,

$$\vec{\mathcal{A}}(\mu_0, \mu_1) = \inf\{H(P|R) ; P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0 \text{ and } P_1 = \mu_1\} - H(\mu_0|\mathcal{L})$$

and is related to the entropic cost (4.10) by the relation

$$\mathcal{A}(\mu_0, \mu_1) = \vec{\mathcal{A}}(\mu_0, \mu_1) + \frac{1}{2}[H(\mu_0|\mathcal{L}) - H(\mu_1|\mathcal{L})]. \quad (4.19)$$

Then, the Benamou-Brenier formulation for the entropic cost (Theorem 3.5.1), in analogy with the one for the Wasserstein distance proved in [BB00]. In the case of the entropic cost this formulation is been proved for the Brownian motion in [CGP16] and for a general Kolmogorov semigroup in [GLR17, Section 5].

Theorem 4.2.5 (Benamou-Brenier formulation) *Let R be the Brownian motion on \mathbb{R}^n , $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$ defined at (4.25), then*

$$\mathcal{A}(\mu_0, \mu_1) = \frac{1}{2} \inf \int_0^1 \int_{\mathbb{R}^n} \left(|v_t(z)|^2 + \frac{1}{4} |\nabla \log \mu_t(z)|^2 \right) \mu_t(z) dz dt,$$

where the infimum runs over all the couples $(\mu_t, v_t)_{0 \leq t \leq 1}$ such that

$$\begin{cases} \mu_t \in P(\mathbb{R}^n), \forall 0 \leq t \leq 1, \\ \mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1, \\ \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0. \end{cases} \quad (4.20)$$

The infimum is achieved by the couple $(\mu_t, \nabla \theta_t)_{0 \leq t \leq 1} \in C^\infty(\mathbb{R}^n)$, where $(\mu_t)_{t \in [0,1]}$ is the entropic interpolation between μ_0 and μ_1 and $(\nabla \theta_t)_{t \in [0,1]}$ appears in (4.16).

4.3 Main results

The first result in this section is about the concavity of the exponential entropy along the entropic interpolation. For simplicity we state and prove our results in the Euclidian space \mathbb{R}^n , but it is still true in more general cases, like a smooth complete connected Riemannian manifold satisfying the $CD(0, n)$ condition, under lighter assumptions on the marginals μ_0, μ_1 . This result is a generalization of an older result known in information theory as the Costa's Theorem [Cos85, CT06], that establishes the concavity of the exponential entropy along the heat flow.

Theorem 4.3.1 (Concavity of exponential entropy) *Let $R \in \mathcal{P}(\Omega)$ be the reversible Brownian motion, $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$. Let $(\mu_s)_{s \in [0,1]}$ be the entropic interpolation between the probability measures μ_0 and μ_1 . Then the function,*

$$\Psi : [0, 1] \ni s \mapsto e^{-H(\mu_s|\mathcal{L})/n}$$

is concave.

Proof

◁ The assumptions on the marginals μ_0, μ_1 make sure that the function Ψ is smooth for all $0 \leq s \leq 1$. Thus, to prove the concavity we will show that the second derivative of Ψ is non positive. In the rest of the proof we use the shortest notation $\frac{d}{ds} H(\mu_s|\mathcal{L}) := h'(s)$. A double differentiation provides,

$$\Psi''(s) = \frac{e^{-H(\mu_s)/n}}{n} \left[\frac{1}{n} h'(s)^2 - h''(s) \right].$$

It remains to prove that

$$CD(0, n) \implies \frac{1}{n} h'(s)^2 - h''(s) \leq 0. \quad (4.21)$$

By definition of the entropy functional, and the transport equation (4.16), easy computations show that the first and second order derivatives of the entropy along the entropic interpolation write as, (see [Léo17])

$$h'(s) = \int \nabla \theta_s \cdot \nabla \mu_s dx \quad (4.22)$$

$$h''(s) = \int [4\Gamma_2(\theta_s) + \Gamma_2(\log \mu_s)] d\mu_s. \quad (4.23)$$

Explicit computations are in Section 2.5.

But the $CD(0, n)$ condition implies (4.8), therefore,

$$\begin{aligned} h''(s) &\geq \frac{1}{n} \int (\Delta \theta_s)^2 d\mu_s \\ &\stackrel{(i)}{\geq} \frac{1}{n} \left(\int \Delta \theta_s d\mu_s \right)^2 \\ &\stackrel{(ii)}{=} \frac{1}{n} \left(\int \nabla \theta_s \cdot \nabla \mu_s dx \right)^2 = \frac{1}{n} (h'(s))^2 \end{aligned}$$

where (i) follows from the Jensen's inequality and (ii) from integration by parts (4.7). Thus, it yields (4.21) and this completes the proof. \triangleright

Remark 4.3.2 Note that formally, if $\theta_s = -\log \mu_s/2$ in (4.16) (hence in (4.22) and (4.23)), then μ_s would be the heat flow and we would recover the stronger result obtained by Costa, namely, the concavity of the function Ψ^2 . Though, when μ_s is the McCann interpolation, concavity of Ψ is the best we can obtain, being equivalent to $CD(0, n)$ as proved in [EKS15]. Our result shows that it is still true for the entropic approximation of the McCann geodesics, namely the entropic interpolations.

Remark 4.3.3 An analogous result holds when we neglect the dimension and we add a drift to the generator, i.e. we consider $L = (\Delta - \nabla V \cdot \nabla)/2$, with reversing measure $dm = e^{-V} \mathcal{L}$ where the potential V is κ -convex for some non-negative κ .

The next theorem is about a differential property of the entropic cost. The stronger result holds under some more restrictive hypothesis on the marginal measures. To this aim let us introduce here the Schwartz space, that is the space of rapidly decreasing functions defined by,

$$\mathcal{S} = \{f : \mathbb{R}^n \rightarrow \mathbb{R}, \text{ s.t. } f \in C^\infty \text{ and } \|x^\alpha D^\beta f\|_\infty < \infty, \forall \alpha, \beta \in \mathbb{Z}_+^n\}. \quad (4.24)$$

It is well known that \mathcal{S} is closed under multiplication and convolution and for any $1 \leq p \leq \infty$, $\mathcal{S} \subset L^p(\mathbb{R}^n)$.

We consider all the couples of measures that admit a *Schrödinger decomposition* of the type (4.14) in which f and g are two positive functions in the Schwartz space (4.24) such that there exists some $\alpha > 0$ such that $f, g \geq c_\alpha e^{-\alpha|x|^2}$ for some positive constant c_α . In other words, we define the set

$$\Pi_{\mathcal{S}} = \left\{ (\mu_0, \mu_1) : \mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n) \text{ and } \exists f, g \text{ s.t. } \begin{cases} \mu_0 = fT_1g, \mu_1 = gT_1f; \\ f, g \in \mathcal{S} \text{ and } f, g > 0; \\ \exists \alpha, c_\alpha > 0 \text{ s.t. } f, g \geq c_\alpha e^{-\alpha|x|^2}. \end{cases} \right\} \quad (4.25)$$

It is immediate to see that if $(\mu_0, \mu_1) \in \Pi_{\mathcal{S}}$ then $\mu_0, \mu_1 \in \Pi$ defined at (4.9).

Theorem 4.3.4 (Regularity of the entropic cost) *Let u, v be two probability densities with respect to the Lebesgue measure on \mathbb{R}^n and let $(T_t)_{t \geq 0}$ denote the heat semigroup.*

(a) *If u, v are such that $u\mathcal{L}, v\mathcal{L} \in \Pi_{\mathcal{S}}$ as defined in (4.25), then the function,*

$$[0, \infty) \ni t \mapsto \mathcal{A}(u\mathcal{L}, T_tv\mathcal{L})$$

is differentiable. In particular, for $t = 0$ it holds,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u, T_tv) = - \left. \frac{1}{2} \frac{d}{ds} \right|_{s=1} H(\mu_s | \mathcal{L}) \quad (4.26)$$

where $(\mu_s)_{0 \leq s \leq 1}$ is the entropic interpolation between u and v .

(b) *If $u, v \in C_c^\infty(\mathbb{R}^n)$, then,*

$$\left. \frac{d^+}{dt} \right|_{t=0} \mathcal{A}(u, T_tv) \leq - \left. \frac{1}{2} \frac{d}{ds} \right|_{s=1} H(\mu_s | \mathcal{L})$$

Where we used the notation

$$\left. \frac{d^+}{dt} \right|_{t=0} f(t) = \limsup_{h \rightarrow 0^+} \frac{f(t+h) - f(t)}{h} \quad (4.27)$$

to denote the super derivative.

Proof

◁ We start by proving that

$$\limsup_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_tv) - \mathcal{A}(u, v)}{t} \leq - \left. \frac{1}{2} \frac{d}{ds} \right|_{s=1} H(\mu_s | \mathcal{L}).$$

for $u, v \in C_c^\infty(\mathbb{R}^n)$. This will prove (b). The same arguments are valid under the assumption $(u, v) \in \Pi_{\mathcal{S}}$, thus we will complete the proof of (a), by proving the converse inequality,

$$\liminf_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_tv) - \mathcal{A}(u, v)}{t} \geq - \left. \frac{1}{2} \frac{d}{ds} \right|_{s=1} H(\mu_s | \mathcal{L}). \quad (4.28)$$

Let $(\mu_s)_{s \in [0,1]}$ be the entropic interpolation between u and v , with associated vector field $\nabla \theta_s$, verifying equation (4.16), that is, $\partial_s \mu_s + \nabla \cdot (\mu_s \nabla \theta_s) = 0$. According to the Benamou-Brenier formulation at Theorem 4.2.5 the entropic cost between u and v can be written as,

$$\mathcal{A}(u, v) = \frac{1}{2} \int_0^1 \int_{\mathbb{R}^n} \left(|\nabla \theta_s|^2 + \frac{1}{4} |\nabla \log \mu_s|^2 \right) \mu_s \, dz ds. \quad (4.29)$$

Following a method already used in [DS08], we define a new path between u and $T_t v$, for all fixed $t \geq 0$, by

$$(\mu_s^t)_{s \in [0,1]} := (T_{st} \mu_s)_{s \in [0,1]}.$$

Let us compute the derivative with respect to $s > 0$, in order to see if it satisfies a transport equation,

$$\begin{aligned} \partial_s \mu_s^t &= \partial_s T_{st} \mu_s \\ &= \frac{t}{2} \Delta \mu_s^t - T_{st} \nabla \cdot (\mu_s \nabla \theta_s) \\ &= \nabla \cdot \left(\frac{t}{2} \nabla \mu_s^t - \mathbf{T}_{st} (\mu_s \nabla \theta_s) \right) \\ &= \nabla \cdot \left[\mu_s^t \left(\frac{t}{2} \nabla \log \mu_s^t - \frac{1}{\mu_s^t} \mathbf{T}_{st} (\mu_s \nabla \theta_s) \right) \right] \end{aligned}$$

where $(\mathbf{T}_t)_{t \geq 0}$ is, roughly speaking, the heat semigroup acting on \mathbb{R}^n -valued functions as a standard heat semigroup on each coordinate,

$$\mathbf{T}_t \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} T_t f_1 \\ \vdots \\ T_t f_n \end{pmatrix}$$

and the associated generator is $\Delta/2$, acting on \mathbb{R}^n -valued functions in similar way,

$$\Delta/2 \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \Delta/2 f_1 \\ \vdots \\ \Delta/2 f_n \end{pmatrix}$$

Therefore $(\mu_s^t)_{s \in [0,1]}$ satisfies the transport equation $\partial_s \mu_s^t + \nabla \cdot (\mu_s^t v_s^t) = 0$, with the vector field

$$v_s^t = -\frac{t}{2} \nabla \log \mu_s^t + \frac{\mathbf{T}_{st} (\mu_s \nabla \theta_s)}{\mu_s^t}. \quad (4.30)$$

Moreover, $\mu_0^t = u$ and $\mu_1^t = T_t v$ for all $t \geq 0$, and of course, for any $s > 0$, $(\mu_s^t)_{t \geq 0}$ is a probability on \mathbb{R}^n . Then the three conditions in (4.20) are satisfied, and, by the Benamou-Brenier formulation (Theorem 4.2.5), we can write,

$$\mathcal{A}(u, T_t v) \leq \int_0^1 \int \left(\frac{|v_s^t|^2}{2} \mu_s^t + \frac{1}{8} |\nabla \log \mu_s^t|^2 \mu_s^t \right) dx ds. \quad (4.31)$$

Taking the difference between (4.31) and (4.29),

$$\begin{aligned} \mathcal{A}(u, T_t v) - \mathcal{A}(u, v) &\leq \int_0^1 \int \left(\frac{|v_s^t|^2}{2} \mu_s^t + \frac{1}{8} |\nabla \log \mu_s^t|^2 \mu_s^t \right) dx ds \\ &\quad - \int_0^1 \int \left(\frac{|v_s|^2}{2} \mu_s + \frac{1}{8} |\nabla \log \mu_s|^2 \mu_s \right) dx ds. \end{aligned} \quad (4.32)$$

Here we denoted $v_s = v_s^0 = \nabla\theta_s$, by definition (4.30). And recall that the couple $(\mu_s, \nabla\theta_s)$ is optimal in Theorem 4.2.5 when $\mu_0 = u$ and $\mu_1 = v$. Dividing (4.32) by $t > 0$ and taking the superior limit for $t \rightarrow 0^+$, we obtain

$$\limsup_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_t v) - \mathcal{A}(u, v)}{t} \leq \frac{d}{dt} \Big|_{t=0} \left[\int_0^1 \int \frac{|v_s^t|^2}{2} \mu_s^t dx ds + \frac{1}{8} \int_0^1 \int |\nabla \log \mu_s^t|^2 \mu_s^t dx ds \right].$$

Note that on the right hand side, the superior limit is actually a limit since $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$. Indeed, all the three terms are differentiable. From Lemma 4.3.5 below we can conclude that,

$$\limsup_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_t v) - \mathcal{A}(u, v)}{t} \leq - \frac{1}{2} \frac{d}{ds} \Big|_{s=1} H(\mu_s | \mathcal{L}).$$

This concludes the proof of (b). These arguments are valid also for $(u, v) \in \Pi_S$, hence to conclude the proof of (a) let us show (4.28). Note that since $(u, v) \in \Pi_S$ then $u, T_t v \in \Pi$, therefore the entropic cost $\mathcal{A}(u, T_t v)$ is still finite. We will use the Kantorovich dual formulation (4.18) for the forward entropic cost. Indeed,

$$\vec{\mathcal{A}}(u, T_t v) - \vec{\mathcal{A}}(u, v) \geq \int \xi T_t v dx - \int Q_1 \xi u dx - \int \psi v dx + \int Q_1 \psi u dx$$

where $\xi \in C_b(\mathbb{R}^n)$ is any bounded continuous function and ψ is optimal in (4.18) when $\mu_0 = u$ and $\mu_1 = v$. By choosing $\xi = \psi$, we get

$$\vec{\mathcal{A}}(u, T_t v) - \vec{\mathcal{A}}(u, v) \geq \int \psi(T_t v - v) dx. \quad (4.33)$$

Note that since g is in the Schwartz space, $\psi = \log g$ is smooth but not bounded. Hence we are not actually allowed to take $\xi = \psi$ but we should approximate ψ by a sequence of bounded and continuous functions by standard arguments. For the sake of simplicity we avoid here the technical details and use the inexact shortcut $\xi = \psi$. Note also that the right hand side of (4.33) is bounded from below thanks to the third hypothesis on the functions f, g in the definition of Π_S , namely up to (bounded) constant factors,

$$\int \psi(T_t v - v) dx \geq - \int |x|^2 (T_t v - v) dx \geq -\infty.$$

The same is false in the case $f, g \in C_{\text{infty}}^1(\mathbb{R}^n)$ for which $\int \psi T_t v dx = -\infty$. On the other hand, by symmetry

$$\vec{\mathcal{A}}(u, T_t v) - \vec{\mathcal{A}}(u, v) = \vec{\mathcal{A}}(T_t v, u) - \vec{\mathcal{A}}(v, u) + H(T_t v | \mathcal{L}) - H(v | \mathcal{L}). \quad (4.34)$$

Again, by duality (4.18),

$$\vec{\mathcal{A}}(T_t v, u) - \vec{\mathcal{A}}(v, u) \geq \int \xi u dx - \int Q_1 \xi T_t v dx - \int \psi^* u dx + \int Q_1 \psi^* T_t v dx.$$

Here, as before, we choose $\xi = \psi^*$, where ψ^* is optimal in (4.18) in the reverse case, when $\mu_0 = v$ and $\mu_1 = u$. In this case we get,

$$\vec{\mathcal{A}}(T_t v, u) - \vec{\mathcal{A}}(v, u) \geq - \int Q_1 \psi^*(T_t v - v) dx. \quad (4.35)$$

Take the half-sum of (4.33) and (4.35) and recall relation (4.34),

$$\vec{\mathcal{A}}(u, T_t v) - \vec{\mathcal{A}}(u, v) + \frac{1}{2} [H(v|\mathcal{L}) - H(T_t v|\mathcal{L})] \geq \frac{1}{2} \left(\int \psi(T_t v - v) dx - \int Q_1 \psi^*(T_t v - v) dx \right).$$

According to (4.19) it reduces to,

$$\mathcal{A}(u, T_t v) - \mathcal{A}(u, v) \geq \frac{1}{2} \left(\int \psi(T_t v - v) dx - \int Q_1 \psi^*(T_t v - v) dx \right).$$

We divide both sides by t and take the inferior limit for $t \rightarrow 0^+$. By the assumptions on the marginals μ_0, μ_1 , the \liminf on the right hand side is actually a limit. By definition of the infinitesimal generator (that we recall being for any measurable bounded function f as $Lf := \lim_{t \rightarrow 0} (T_t f - f)/t$) we obtain,

$$\liminf_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_t v) - \mathcal{A}(u, v)}{t} \geq \frac{1}{2} \int \psi Lv dx - \frac{1}{2} \int Q_1 \psi^* Lv dx$$

and after integration by parts (4.7), it yields to

$$\liminf_{t \rightarrow 0^+} \frac{\mathcal{A}(u, T_t v) - \mathcal{A}(u, v)}{t} \geq -\frac{1}{2} \int \left(\frac{\nabla \psi - \nabla Q_1 \psi^*}{2} \right) \cdot \nabla v dx. \quad (4.36)$$

Note that since ψ^* is optimal, $Q_1 \psi^*$ coincides with the potential $\psi_t^* = \log T_{1-t} g^* = \log T_{1-t} f$ for $t = 0$. Therefore $\nabla Q_1 \psi^* = \nabla \psi_0^* = \vec{v}_0^* = \overleftarrow{v}_1 = \nabla \varphi_1$, where the middle equality is given by the relation (4.17). We can conclude that on the right hand side of (4.36) we have,

$$\nabla \theta_1 = \frac{\nabla \psi_1 - \nabla \varphi_1}{2}$$

and this proves (4.28).

We have proven that the inferior and supremum limits are bounded from above and below by the same quantity, and it proves (4.26). Moreover by the semigroup property, the differential property can be extended to any $t \geq 0$. \triangleright

Lemma 4.3.5

$$\left. \frac{d}{dt} \right|_{t=0} \int_0^1 \int \left(\frac{|v_s^t|^2}{2} + \frac{1}{8} |\nabla \log \mu_s^t|^2 \right) \mu_s^t dx ds = -\left. \frac{1}{2} \frac{d}{ds} \right|_{s=1} H(\mu_s|\mathcal{L}).$$

Proof

\triangleleft We denote for simplicity,

$$\begin{aligned} (i) &= \left. \frac{d}{dt} \right|_{t=0} \int_0^1 \int \frac{|v_s^t|^2}{2} \mu_s^t dx ds, \\ (ii) &= \left. \frac{d}{dt} \right|_{t=0} \int_0^1 \int \left(\frac{1}{8} |\nabla \log \mu_s^t|^2 \right) \mu_s^t dx ds. \end{aligned}$$

Let us compute separately, these two derivatives,

$$\begin{aligned}
(i) &= \int_0^1 \int v_s^t \frac{d}{dt} v_s^t \mu_s^t + \frac{|v_s^t|^2}{2} \frac{d}{dt} \mu_s^t dx ds \Big|_{t=0} \\
&= \int_0^1 \int v_s^t \cdot \left[-\frac{1}{2} \nabla \log \mu_s^t - \frac{t}{2} \nabla \left(\frac{s \Delta \mu_s^t}{2 \mu_s^t} \right) + \frac{1}{(\mu_s^t)^2} \left(s \frac{\Delta}{2} \mathbf{T}_{st}(\mu_s \nabla \theta_s) \mu_s^t \right. \right. \\
&\quad \left. \left. - \frac{s}{2} \mathbf{T}_{st}(\mu_s \nabla \theta_s) \Delta \mu_s^t \right) \right] \mu_s^t + \frac{|v_s^t|^2}{4} s \Delta \mu_s^t dx ds \Big|_{t=0} \\
&= \int_0^1 \int \nabla \theta_s \cdot \left[-\frac{1}{2} \nabla \log \mu_s + \frac{1}{(\mu_s)^2} \left(s \frac{\Delta}{2} (\mu_s \nabla \theta_s) \mu_s - \frac{s}{2} \mu_s \nabla \theta_s \Delta \mu_s \right) \right] \mu_s + \\
&\quad + \frac{|\nabla \theta_s|^2}{4} s \Delta \mu_s dx ds \\
&= \frac{1}{2} \int_0^1 \int -\nabla \theta_s \cdot \nabla \mu_s + s \nabla \theta_s \cdot \Delta (\mu_s \nabla \theta_s) - s \frac{|\nabla \theta_s|^2}{2} \Delta \mu_s dx ds \\
&= \frac{1}{2} \int_0^1 \int -\nabla \theta_s \cdot \nabla \mu_s + s \left[\nabla \Delta \theta_s \cdot \nabla \theta_s \mu_s - \frac{1}{2} \Delta (|\nabla \theta_s|^2) \mu_s \right] dx ds \\
&\stackrel{(4.22)}{=} \int_0^1 \left[-\frac{1}{2} \frac{d}{ds} H(\mu_s | \mathcal{L}) - 2s \int \Gamma_2(\theta_s) \mu_s dx \right] ds.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
(ii) &= \frac{1}{8} \int_0^1 \int 2 \nabla \log \mu_s^t \cdot \nabla \left(\frac{s \Delta \mu_s^t}{2 \mu_s^t} \right) \mu_s^t + \frac{1}{2} |\nabla \log \mu_s^t|^2 s \Delta \mu_s^t dx ds \Big|_{t=0} \\
&= \frac{1}{8} \int_0^1 \int s \nabla \log \mu_s \cdot \nabla \left(\frac{\Delta \mu_s}{\mu_s} \right) \mu_s + \frac{1}{2} |\nabla \log \mu_s|^2 s \Delta \mu_s dx ds \\
&= \frac{1}{8} \int_0^1 \int s (\nabla \log \mu_s \cdot \nabla (\Delta \log \mu_s + |\nabla \log \mu_s|^2) + \frac{1}{2} \Delta |\nabla \log \mu_s|^2) \mu_s dx ds \\
&= \frac{1}{8} \int_0^1 \int s \nabla \log \mu_s \cdot \nabla \Delta \log \mu_s \mu_s - \frac{1}{2} s \Delta |\nabla \log \mu_s|^2 \mu_s dx ds \\
&= -\frac{1}{2} \int_0^1 \int s \Gamma_2(\log \mu_s) \mu_s dx ds
\end{aligned}$$

Taking the sum of (i) and (ii),

$$\begin{aligned}
(i) + (ii) &= \int_0^1 \left[-\frac{1}{2} \frac{d}{ds} H(\mu_s | \mathcal{L}) - 2s \int \Gamma_2(\psi_s) \mu_s + \frac{1}{2} \Gamma_2(\log \mu_s) dx \right] ds \\
&\stackrel{(4.23)}{=} \int_0^1 -\frac{1}{2} \frac{d}{ds} H(\mu_s | \mathcal{L}) - \frac{s}{2} \frac{d^2}{ds^2} H(\mu_s | \mathcal{L}) ds \\
&= -\frac{1}{2} \int_0^1 \frac{d}{ds} \left(s \frac{d}{ds} H(\mu_s | \mathcal{L}) \right) ds \\
&= -\frac{1}{2} \frac{d}{ds} \Big|_{s=1} H(\mu_s | \mathcal{L}).
\end{aligned}$$

▷

The next theorem is an immediate consequence of Theorem 4.3.1 and Theorem 4.3.4. It establishes the Evolution Variational Inequality for the entropic cost, under the $CD(0, n)$ condition.

Corollary 4.3.6 (EVI for entropic cost) *Under the same hypothesis of Theorem 4.3.4 (a),*

$$\frac{d^+}{dt} \mathcal{A}(u \mathcal{L}, T_t v \mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right) \quad (4.37)$$

for any $t \geq 0$. If u, v satisfy the hypothesis of Theorem 4.3.4 (b), then the standard derivative on the left hand side has to be replaced by the super derivative.

Proof

◁ Without loss of generality, by the semigroup property, it is enough to prove (4.37) for $t = 0$, i.e.

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u \mathcal{L}, T_t v \mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right).$$

By Theorem 4.3.1, we have in particular that the concavity of the function Ψ implies that

$$\Psi'(1) \leq \Psi(1) - \Psi(0).$$

Taking into account the definition of Ψ , it is equivalent to say

$$e^{-H(v|\mathcal{L})/n} \left(-\frac{1}{n} \left. \frac{d}{ds} \right|_{s=1} H(\mu_s|\mathcal{L}) \right) \leq e^{-H(v|\mathcal{L})/n} - e^{-H(u|\mathcal{L})/n}$$

that after rearranging the terms gives,

$$-\left. \frac{d}{ds} \right|_{s=1} H(\mu_s|\mathcal{L}) \leq n \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right). \quad (4.38)$$

We conclude by applying (4.38) to (4.26), to obtain the claimed result,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u \mathcal{L}, T_t v \mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right).$$

▷

Remark 4.3.7 *We show later at Corollary 4.3.11 that the EVI inequality for the entropic cost provides an immediate and alternative proof of the EVI inequality for the Wasserstein distance under the $CD(0, n)$ condition.*

The evolution variational inequality has some nice consequence. The first one stated in the first Corollary below, is the contraction of the entropic cost along the heat flow. It is an improvement of Theorem 3.6.1 ([GLR17, Thm 6.6 (b)]) where dimensional contraction with respect to two different time variables $t, s \geq 0$ is shown.

Corollary 4.3.8 (Contraction) *Under the same hypothesis of Theorem 4.3.4 (a), it holds for any $t > 0$,*

$$\mathcal{A}(T_\tau u, T_\tau v) \leq \mathcal{A}(u, v) - n \int_0^\tau \sinh^2 \left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{2n} \right) dt. \quad (4.39)$$

Proof

◁ To derive contraction from EVI, we follow a standard strategy already used to deduce contraction in Wasserstein distance, see [AGS08].

Inequality (4.37) is true for any probability density u , thus it is true in particular for $T_s u$ for some fixed $s \geq 0$,

$$\frac{d}{dt} \mathcal{A}(T_s u, T_t v) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(T_s u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right). \quad (4.40)$$

We mentioned in Remark 4.2.1 (ii), that the entropic cost is symmetric with respect to the initial and final measures, thus,

$$\frac{d}{dt} \mathcal{A}(T_t v, u) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right)$$

and if we switch u and v , it yields to

$$\frac{d}{dt} \mathcal{A}(T_t u, v) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(v|\mathcal{L}) - H(T_t u|\mathcal{L})]} \right).$$

Since v is also arbitrary, we can replace it by $T_s v$ for some $s \geq 0$ fixed,

$$\frac{d}{dt} \mathcal{A}(T_t u, T_s v) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(T_s v|\mathcal{L}) - H(T_t u|\mathcal{L})]} \right). \quad (4.41)$$

We take the sum of (4.40) and (4.41),

$$\frac{d}{dt} \mathcal{A}(T_s u, T_t v) + \frac{d}{dt} \mathcal{A}(T_t u, T_s v) \leq \frac{n}{2} \left(2 - e^{-\frac{1}{n}[H(T_s u|\mathcal{L}) - H(T_t v|\mathcal{L})]} - e^{-\frac{1}{n}[H(T_s v|\mathcal{L}) - H(T_t u|\mathcal{L})]} \right).$$

We take $s = t$ and integrate in t between 0 and τ , for some $\tau > 0$,

$$\begin{aligned} \mathcal{A}(T_\tau u, T_\tau v) - \mathcal{A}(u, v) &\leq \frac{n}{2} \int_0^\tau 2 - \left(e^{-\frac{1}{n}[H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})]} + e^{\frac{1}{n}[H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right) dt \\ &= n \int_0^\tau 1 - \cosh \left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{n} \right) dt. \end{aligned}$$

Moreover, by recalling that $\sinh^2 x = (\cosh(2x) - 1)/2$, we can rewrite the contraction inequality as,

$$\mathcal{A}(T_\tau u, T_\tau v) - \mathcal{A}(u, v) \leq -2n \int_0^\tau \sinh^2 \left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{2n} \right) dt.$$

▷

Remark 4.3.9 *Contraction for the entropic cost, implies the analogue dimensional contraction for the quadratic Wasserstein distance along the heat flow [BGG16, BGGK16].*

$$W_2^2(T_\tau u, T_\tau v) \leq W_2^2(u, v) - 4n \int_0^\tau \sinh^2 \left(\frac{H(T_t u|\mathcal{L}) - H(T_t v|\mathcal{L})}{2n} \right) dt. \quad (4.42)$$

It can be seen, by considering (4.39) for the ε -entropic cost \mathcal{A}^ε defined in (4.11), take the limit for $\varepsilon \rightarrow 0$ and recall the convergence property (4.12).

Finally, we show an *integral* and equivalent form of (4.37). Through the integral form we will be able in Corollary 4.3.11 to deduce the classical EVI for the Wasserstein distance.

Proposition 4.3.10 *Under the same assumption as in Theorem 4.3.4, the following statements are equivalent:*

(i) For any $u, v \in \Pi_S$, and any $t \geq 0$,

$$\frac{d}{dt} \mathcal{A}(u, T_t v) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right); \quad (4.43)$$

(ii) For any $u, v \in \Pi_S$,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u, T_t v) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right);$$

(iii) For any $u, v \in \Pi_S$, and any $t \geq 0$,

$$\mathcal{A}(u, T_t v) - \mathcal{A}(u, v) \leq \frac{n}{2} t \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right). \quad (4.44)$$

Proof

\triangleleft (i) \Rightarrow (iii) We integrate (4.43) with respect to $0 \leq t \leq \tau$, for any fixed $\tau > 0$. Thus,

$$\begin{aligned} \mathcal{A}(u, T_\tau v) - \mathcal{A}(u, v) &\leq \frac{n}{2} \int_0^\tau 1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} dt \\ &\leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_\tau v|\mathcal{L})]} \right) \int_0^\tau dt \\ &= \frac{n}{2} \tau \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_\tau v|\mathcal{L})]} \right) \end{aligned}$$

where the second inequality is given by the fact that the entropy is decreasing along the heat flow.

(iii) \Rightarrow (ii) We divide by t both sides of (4.44), and take the limit for $t \rightarrow 0^+$. The limit exists on the left hand side by Theorem 4.3.4, and on the right hand side by continuity. Therefore we obtain,

$$\left. \frac{d}{dt} \right|_{t=0} \mathcal{A}(u|\mathcal{L}, T_t v|\mathcal{L}) \leq \frac{n}{2} \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right).$$

(ii) \Rightarrow (i) It is true by the semigroup property. \triangleright

Corollary 4.3.11 (EVI for the Wasserstein distance) *Under the same assumption of Theorem 4.3.4, EVI for the Wasserstein distance holds for any $t \geq 0$,*

$$\frac{d^+}{dt} W_2^2(u, T_t v) \leq n(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]}). \quad (4.45)$$

Proof

\triangleleft Let us consider (4.43) for the ε -entropic cost (4.11). As we just showed, it implies the integral form (4.44) for the ε -entropic cost,

$$\mathcal{A}^\varepsilon(u, T_t v) - \mathcal{A}^\varepsilon(u, v) \leq \frac{n}{2} t \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right).$$

Taking the limit for $\varepsilon \rightarrow 0$, and recalling (4.12) we obtain, for any $t > 0$,

$$W_2^2(u, T_t v) - W_2^2(u, v) \leq nt \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(T_t v|\mathcal{L})]} \right).$$

Finally, divide by t both sides and take the lim sup for $t \rightarrow 0^+$, to get

$$\left. \frac{d^+}{dt} W_2^2(u, T_t v) \right|_{t=0} \leq n \left(1 - e^{-\frac{1}{n}[H(u|\mathcal{L}) - H(v|\mathcal{L})]} \right).$$

Again, by the semigroup property, it implies (4.45) for any $t \geq 0^+$. Remark that at the limit $\varepsilon \rightarrow 0$ the differential property at Theorem 4.3.4 is no more satisfied, hence we cannot have better than the sup-derivative of the Wasserstein distance. \triangleright

Remark 4.3.12 *The absence of the factor 1/2 on the left hand side in (4.45) and (4.42), is due to the fact that we have chosen the heat semigroup associated to $L = \Delta/2$, that is a more natural choice in the framework of the Schrödinger problem. It is straightforward to see that they are equivalent to their analogues in [EKS15, BGG16] and [BGGK16], where $L = \Delta$.*

Chapter 5

Evolutionary variational inequality under $CD(\kappa, \infty)$

In this chapter we investigate the *Evolutionary variational inequality* under the $CD(\kappa, \infty)$ condition for some $\kappa \in \mathbb{R}$. Let $(T_t)_{t \geq 0}$ be the Markov semigroup associated with the generator $L = (\Delta - \nabla V \cdot \nabla)/2$ and reversing measure $\mathbf{m} = e^{-V} \mathcal{L}$. We recall that in this context to say that L satisfies a $CD(\kappa/2, \infty)$ condition is equivalent to say that the potential V is κ -convex, that is $|\nabla^2 V| \geq \kappa/2 Id$. The EVI inequality in Wasserstein distance under $CD(\kappa/2, \infty)$ writes as,

$$\frac{d^+}{dt} W_2^2(u, T_t v) + \frac{\kappa}{2} W_2^2(u, T_t v) \leq H(u|\mathbf{m}) - H(T_t v|\mathbf{m}), \quad \forall t \geq 0, \quad (5.1)$$

where d^+/dt is defined at (4.27). This inequality was first introduced in [AGS08] to define gradient flows in measure metric spaces and κ -convexity of functionals such as the relative entropy (See also [Vil09, Ch. 23]). In Chapter 4 we proved this inequality for the entropic cost in the flat dimensional case $CD(0, n)$. Here we show the counterpart of these results in the case $CD(\kappa, \infty)$. In particular we will see that, contrarily to the $CD(0, N)$ case, when we *replace* the ε -entropic cost to the Wasserstein distance in (5.1) we obtain a penalized inequality due to the diffusion phenomenon. Luckily at the limit for $\varepsilon \rightarrow 0^+$, the additional term vanishes, and we recover the standard results for the Wasserstein distance.

After briefly reviewing in Section 5.1 the basic notions and properties that we need in the proof, at Section 5.2 we present the main result of this Chapter, that will be proved at Section 5.3 via two different methods.

5.1 Setting

We start by recalling some essential facts already stated in Chapters 1, 2, 3. We denote $(T_t)_{t \geq 0}$ the Kolmogorov Markov semigroup associated with generator $L = (\Delta - \nabla V \cdot \nabla)/2$ and recall that it satisfies the $CD(\kappa/2, \infty)$ condition (1.32), that is equivalent to the gradient bound

$$|\nabla T_t f|^2 \leq e^{-\kappa t} T_t(|\nabla f|^2) \quad (5.2)$$

for any function $f \in C_c^\infty(\mathbb{R}^n)$, (Theorem 1.1.26).

We consider the Schrödinger problem with reference measure R^ε , associated to the generator $L^\varepsilon = \varepsilon(\Delta - \nabla V \cdot \nabla)/2$ for which we adopt the convention for the notation (2.11),

that we recall,

$$L^\varepsilon = \varepsilon L, \quad \Gamma^\varepsilon f = \varepsilon \Gamma f, \quad \Gamma_2^\varepsilon f = \varepsilon^2 \Gamma_2.$$

where

$$Lf = (\Delta f - \nabla V \cdot \nabla f)/2, \quad \Gamma f = \frac{|\nabla f|^2}{2}, \quad \Gamma_2 f = \frac{1}{2}L(|\nabla f|^2) - \nabla f \cdot \nabla Lf.$$

Let the marginals $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ be such that the hypothesis of Theorem 2.1.4 are satisfied, then the associated entropic cost is defined by,

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \inf\{\varepsilon H(P|R^\varepsilon) : P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\} - \frac{\varepsilon}{2}[H(\mu_0|\mathbf{m}) + H(\mu_1|\mathbf{m})].$$

See Chapter 2 for any further detail about the Schrödinger problem. In Chapter 3 we proved the dual Kantorovich formulation for the entropic cost (2.37), that we recall here.

Theorem 5.1.1 (Dual Kantorovich formulation) *For $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$, then*

$$\vec{\mathcal{A}}^\varepsilon(\mu_0, \mu_1) = \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int \psi d\mu_1 - \int \mathcal{Q}_1^\varepsilon \psi d\mu_0 \right\}. \quad (5.3)$$

Here, $\mathcal{Q}_t^\varepsilon \psi(x) := \varepsilon \log T_{\varepsilon t}(e^{\psi/\varepsilon})(x)$, $x \in \mathbb{R}^n$ and it satisfies the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t u - \frac{|\nabla u|^2}{2} - \frac{\varepsilon}{2} \Delta u = 0, & t > 0, \\ u(0, \cdot) = \psi. \end{cases} \quad (5.4)$$

We recall from Section 2.3 that the *forward entropic cost* is defined by,

$$\vec{\mathcal{A}}^\varepsilon(\mu_0, \mu_1) = \inf\{H(P|R) ; P \in \mathcal{P}(\Omega) \text{ s.t. } P_0 = \mu_0 \text{ and } P_1 = \mu_1\} - \varepsilon H(\mu_0|\mathcal{L})$$

and is related to the entropic cost (4.10) by the relation

$$\mathcal{A}(\mu_0, \mu_1) = \vec{\mathcal{A}}(\mu_0, \mu_1) + \frac{\varepsilon}{2}[H(\mu_0|\mathcal{L}) - H(\mu_1|\mathcal{L})]. \quad (5.5)$$

On the other hand the entropic cost of Definition 2.3.2 admits the Benamou-Brenier formulation, proved at Chapter 3, which states that

Theorem 5.1.2 (Benamou-Brenier formulation) *Let R be the Brownian motion on \mathbb{R}^n , $\mu_0, \mu_1 \in C_c^\infty(\mathbb{R}^n)$, then*

$$\mathcal{A}^\varepsilon(\mu_0, \mu_1) = \frac{1}{2} \inf \int_0^1 \int_{\mathbb{R}^n} \left(|\nabla \phi_s|^2 + \frac{\varepsilon^2}{4} |\nabla \log \rho_s|^2 \right) \mu_s(z) dx ds, \quad (5.6)$$

where we denoted $\rho_s = d\mu_s/d\mathbf{m}$ for all $0 \leq s \leq 1$, and the infimum runs over all the couples $(\mu_s, \nabla \phi_s)_{0 \leq s \leq 1}$

$$\begin{cases} \mu_s \in P(\mathbb{R}^n), \forall 0 \leq s \leq 1, \phi_s \in C^\infty([0, 1] \times \mathbb{R}^n) \\ \mu_0 = \mu_0, \mu_1 = \mu_1, \\ \partial_s \mu_t + \nabla \cdot (\mu_s \nabla \phi_s) = 0. \end{cases}$$

For the proof of these two results, in the case $\varepsilon = 1$, see Sections 3.4 and 3.5. We denote by $(\mu_s^\varepsilon)_{0 \leq s \leq 1}$ the entropic interpolation, solution of the Schrödinger problem (See Section 2.2), and recall that we denote the density of $(\mu_s^\varepsilon)_{0 \leq s \leq 1}$ with respect to the Lebesgue measure and to the equilibrium measure \mathbf{m} , respectively by,

$$\mu_s^\varepsilon = \frac{d\mu_s^\varepsilon}{dx}, \quad \rho_s^\varepsilon = \frac{d\mu_s^\varepsilon}{d\mathbf{m}},$$

where,

$$\rho_s^\varepsilon = e^{-\frac{\varphi_s^\varepsilon + \psi_s^\varepsilon}{\varepsilon}} \quad \mu_s^\varepsilon = e^{-\frac{\varphi_s^\varepsilon + \psi_s^\varepsilon}{\varepsilon}} e^{-V}.$$

5.2 Main result

Theorem 5.2.1 (EVI) *Let $u, v \in C_c^\infty(\mathbb{R}^n)$ be two probability densities with respect to \mathbf{m} on \mathbb{R}^n . Under $CD(\kappa/2, \infty)$ with $\kappa \in \mathbb{R}$ we have for all $t \geq 0$,*

$$\frac{d}{dt} \mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, T_t \mathbf{v}\mathbf{m}) + \frac{\kappa}{2} \mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m}) - \frac{\varepsilon^2}{8} |\kappa| \int_0^1 I_{\mathbf{m}}(\rho_s) ds \leq \frac{1}{2} [H(u|\mathbf{m}) - H(T_t v|\mathbf{m})] \quad (5.7)$$

where $(\rho_s)_{0 \leq s \leq 1}$ is the density of the entropic interpolation between $\mathbf{u}\mathbf{m}$ and $\mathbf{v}\mathbf{m}$, and $I_{\mathbf{m}}(\rho_s)$ is a short notation for the Fisher information of ρ_s with respect to the reference measure $d\mathbf{m}$,

$$I_{\mathbf{m}}(\rho) = \int |\nabla \log \rho|^2 \rho d\mathbf{m}. \quad (5.8)$$

Remarks 5.2.2 *Let us point out that:*

- we obtained a penalized inequality, due to the presence of the additional term,

$$-\frac{\varepsilon^2}{8} |\kappa| \int_0^1 I_{\mathbf{m}}(\rho_s) ds$$

which corresponds to $|\kappa|$ times the osmotic action introduced at Section 2.3.

- the case $\kappa = 0$ is trivial. It can be seen as a consequence of the case $\kappa > 0$, or as a consequence of Corollary 4.3.6 when $N \rightarrow \infty$.

Corollary 5.2.3 (EVI under $CD(\kappa, \infty)$ for W_2) *Under the same assumptions as Theorem 5.2.1, we have*

$$\frac{d^+}{dt} W_2^2(u, T_t v) + \frac{\kappa}{2} W_2^2(u, T_t v) \leq H(T_t v|\mathbf{m}) - H(u|\mathbf{m}), \quad \forall t \geq 0.$$

Proof

◁ Consider an integrated form of the EVI inequality (5.7).

$$\mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, T_t \mathbf{v}\mathbf{m}) - \mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m}) + \frac{\kappa}{2} t \mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m}) + \frac{\varepsilon^2}{8} |\kappa| t \int_0^1 I_{\mathbf{m}}(\mu_s) ds \leq \frac{1}{2} \int_0^t H(u|\mathbf{m}) - H(T_\tau v|\mathbf{m})$$

Taking the limit for $\varepsilon \rightarrow 0^+$, we have from (2.41) and (2.50),

$$\frac{W_2^2(\mathbf{u}\mathbf{m}, T_t \mathbf{v}\mathbf{m})}{2} - \frac{W_2^2(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m})}{2} + \kappa t \frac{W_2^2(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m})}{4} \leq \frac{1}{2} \int_0^t H(u|\mathbf{m}) - H(T_\tau v|\mathbf{m}).$$

It suffices to divide by $t > 0$ and take the superior limit for $t \rightarrow 0^+$, to obtain the claimed result. ▷

5.3 Proofs

5.3.1 First method

Here we present one proof of Theorem 5.2.1, following the strategy explained by Bakry Gentil and Ledoux in [BGL15] to prove EVI for the Wasserstein distance in the particular case $\kappa = 0$. This proof is based on the Kantorovich and Benamou-Brenier formulations, with the trick of running along the entropic interpolation $(\mu_s)_{0 \leq s \leq 1}$, not linearly in s but through a smooth bijective function $\zeta(s) : [0, 1] \rightarrow [0, 1]$ "carrying" the curvature κ .

First we prove the following auxiliary result.

Lemma 5.3.1 *For any couple $(\mu_s, v_s)_{0 \leq s \leq 1}$ such that conditions (i)-(iii) in (5.6) are satisfied, for all function $H \in L^2(\mathbf{m})$ and each smooth function $\zeta : [0, 1] \mapsto [0, 1]$, it holds,*

$$\int_0^1 \int \dot{\zeta} \dot{\mu}_{\zeta(s)} H - \dot{\zeta} K \mu_{\zeta(s)} \frac{|\nabla H|^2}{2} ds dx \leq \frac{1}{2K} \int_0^1 \int |v_s|^2 \mu_s ds dx$$

for all $K > 0$.

Proof

◁ Consider the transport equation (iii) in (5.6) at time $t = \zeta(s)$,

$$\begin{aligned} \int_0^1 \int \dot{\zeta} \dot{\mu}_{\zeta(s)} H ds dx &= \int_0^1 \int -\dot{\zeta} H \nabla \cdot (v_{\zeta(s)} \mu_{\zeta(s)}) ds dx \\ &= \int_0^1 \int \dot{\zeta} (\nabla H \cdot v_{\zeta(s)}) \mu_{\zeta(s)} ds dx \\ &\leq \int_0^1 \int \dot{\zeta} \left(\frac{K}{2} |\nabla H|^2 + \frac{1}{2K} |v_{\zeta(s)}|^2 \right) \mu_{\zeta(s)} ds dx \end{aligned}$$

for all $K > 0$. After the change of variable $t = \zeta(s)$ in the last term on the right hand side and after rearranging the terms, we get the desired result. ▷

Proof of Theorem 5.2.1

◁ Thanks to the semigroup property, we can prove inequality (5.7) in $t = 0$, without loss of generality. For $\varepsilon > 0$ and all $t \geq 0$, let us define

$$B := \int \psi T_t v d\mathbf{m} - \int \mathcal{Q}_1^\varepsilon \psi u d\mathbf{m} + \left(\frac{t - \varepsilon}{2} \right) \int (T_t v \log T_t v - u \log u) d\mathbf{m}$$

where ψ is the optimal function in the Kantorovich dual formulation (5.3) when $\mu_0 = u\mathbf{m}$ and $\mu_1 = T_t v\mathbf{m}$. Let $(\mu_s)_{0 \leq s \leq 1}$ be the entropic interpolation between $u\mathbf{m}$ and $v\mathbf{m}$, let $(\rho_s)_{0 \leq s \leq 1}$ be its density with respect to the equilibrium measure \mathbf{m} and let $\zeta(s)$ be a smooth function from $[0, 1]$ to $[0, 1]$ such that $\zeta(0) = 0$ and $\zeta(1) = 1$. Then, it can be easily checked that B can be written as,

$$B = \int_0^1 \frac{d}{ds} \int \mathcal{Q}_{1-s}^\varepsilon \psi T_{st} \rho_{\zeta(s)} + \left(\frac{t - \varepsilon}{2} \right) T_{st} \rho_{\zeta(s)} \log T_{st} \rho_{\zeta(s)} ds d\mathbf{m}$$

Let us compute explicitly the derivative with respect to $s \in [0, 1]$. Recall the Hamilton Jacobi Bellman equation (5.4) satisfied by the semigroup \mathcal{Q}^ε , and together with the properties in Proposition 1.1.22 we obtain,

$$\begin{aligned} B &= \int_0^1 \int \left[-\varepsilon L \mathcal{Q}_{1-s}^\varepsilon \psi - \frac{|\nabla \mathcal{Q}_{1-s}^\varepsilon \psi|^2}{2} \right] T_{st} \rho_{\zeta(s)} + \mathcal{Q}_{1-s}^\varepsilon \psi \left[t L T_{st} \rho_{\zeta(s)} + \dot{\zeta} T_{st} \dot{\rho}_{\zeta(s)} \right] \\ &\quad + \left(\frac{t - \varepsilon}{2} \right) [t L T_{st} \rho_{\zeta(s)} + \dot{\zeta} T_{st} \dot{\rho}_{\zeta(s)}] \log T_{st} \rho_{\zeta(s)} ds d\mathbf{m} \end{aligned}$$

We regroup the terms with factor $\dot{\zeta} \dot{\rho}_{\zeta(s)}$ and integrate by parts according to the identity,

$$\int f L g d\mathbf{m} = -\frac{1}{2} \int \nabla f \cdot \nabla g d\mathbf{m}.$$

We get,

$$B = \int_0^1 \int \dot{\zeta} \dot{\rho}_{\zeta(s)} \left[T_{st} \mathcal{Q}_{1-s}^\varepsilon \psi + \frac{t-\varepsilon}{2} T_{st} \log T_{st} \rho_{\zeta(s)} \right] + T_{st} \rho_{\zeta(s)} \left[\frac{t-\varepsilon}{2} \nabla \mathcal{Q}_{1-s}^\varepsilon \psi \cdot \nabla \log T_{st} \rho_{\zeta(s)} - \frac{|\nabla \mathcal{Q}_{1-s}^\varepsilon \psi|^2}{2} - \frac{t(t-\varepsilon)}{4} |\nabla \log T_{st} \rho_{\zeta(s)}|^2 \right] ds d\mathbf{m}.$$

We set $H := T_{st} \mathcal{Q}_{1-s}^\varepsilon \psi + \left(\frac{t-\varepsilon}{2}\right) T_{st} \log T_{st} \rho_{\zeta(s)}$ and rearrange the terms in the square brackets to obtain the square of a sum,

$$B = \int_0^1 \int \dot{\zeta} \dot{\rho}_{\zeta(s)} H - \frac{1}{2} T_{st} \rho_{\zeta(s)} \left| \nabla \mathcal{Q}_{1-s}^\varepsilon \psi + \left(\frac{t-\varepsilon}{2}\right) \nabla \log T_{st} \rho_{\zeta(s)} \right|^2 + \frac{\varepsilon^2}{8} T_{st} \rho_{\zeta(s)} |\nabla \log T_{st} \rho_{\zeta(s)}|^2 ds d\mathbf{m}.$$

For what concerns the second term,

$$\begin{aligned} & \int_0^1 \int -\frac{1}{2} T_{st} \rho_{\zeta(s)} |\nabla \mathcal{Q}_{1-s}^\varepsilon \psi + \left(\frac{t-\varepsilon}{2}\right) \nabla \log T_{st} \rho_{\zeta(s)}|^2 ds d\mathbf{m} \\ = & \int_0^1 \int -\frac{1}{2} \rho_{\zeta(s)} T_{st} \left(|\nabla \mathcal{Q}_{1-s}^\varepsilon \psi + \left(\frac{t-\varepsilon}{2}\right) \nabla \log T_{st} \rho_{\zeta(s)}|^2 \right) ds d\mathbf{m} \\ \leq & \int_0^1 \int -\frac{1}{2} e^{\kappa st} \rho_{\zeta(s)} \left(|\nabla T_{st} \mathcal{Q}_{1-s}^\varepsilon \psi + \left(\frac{t-\varepsilon}{2}\right) \nabla T_{st} \log T_{st} \rho_{\zeta(s)}|^2 \right) ds d\mathbf{m} \\ = & \int_0^1 \int -\frac{1}{2} e^{\kappa st} \rho_{\zeta(s)} |\nabla H|^2 ds d\mathbf{m}, \end{aligned}$$

where the inequality is given by the commutation property between the semigroup and the gradient (5.2). Thus we obtained,

$$B \leq \int_0^1 \int \dot{\zeta} \dot{\rho}_{\zeta(s)} H - \frac{1}{2} \rho_{\zeta(s)} e^{\kappa st} |\nabla H|^2 + \frac{\varepsilon^2}{8} T_{st} \rho_{\zeta(s)} |\nabla \log T_{st} \rho_{\zeta(s)}|^2 ds d\mathbf{m}. \quad (5.9)$$

Now, we use Lemma 5.3.1 applied to the couple $(\mu_s, \nu_s)_{0 \leq s \leq 1}$ given by the entropic interpolation between $u\mathbf{m}$ and $v\mathbf{m}$, and the associated current velocity, with $\dot{\zeta} K = e^{\kappa st}$. Together with the boundary conditions $\zeta(0) = 0$ and $\zeta(1) = 1$, it yields to,

$$\zeta(s) := \frac{1}{K} \frac{e^{\kappa st} - 1}{\kappa t}, \quad \text{where } K = \frac{e^{\kappa t} - 1}{\kappa t} > 0, \quad \forall t > 0. \quad (5.10)$$

This provides the following upper bound for the first two terms on the right hand side of (5.9),

$$B \leq \int_0^1 \int \frac{1}{2K} |v_s|^2 \rho_s + \frac{\varepsilon^2}{8} T_{st} \rho_{\zeta(s)} |\nabla \log T_{st} \rho_{\zeta(s)}|^2 ds d\mathbf{m} \quad (5.11)$$

According to the Benamou-Brenier formulation (5.6) the first term on the right hand side is equal to

$$\int_0^1 \int \frac{|v_s|^2}{2K} \rho_s ds d\mathbf{m} = \frac{1}{K} \left[\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) - \frac{\varepsilon^2}{8} \int_0^1 \int |\nabla \log \rho_s|^2 \rho_s ds d\mathbf{m} \right]. \quad (5.12)$$

On the other hand, given the Kantorovich dual formulation (5.3), B corresponds to

$$B = \vec{\mathcal{A}}^\varepsilon(u\mathbf{m}, T_t v\mathbf{m}) + \left(\frac{t-\varepsilon}{2}\right) [H(T_t v|\mathbf{m}) - H(u|\mathbf{m})],$$

which reduces, according to (5.5) to,

$$B = \mathcal{A}^\varepsilon(u\mathbf{m}, T_t v\mathbf{m}) + t[H(T_t v|\mathbf{m}) - H(u|\mathbf{m})]/2. \quad (5.13)$$

Finally the estimation (5.11) of B together with (5.12) and (5.13) yields to,

$$\begin{aligned} \mathcal{A}^\varepsilon(u\mathbf{m}, T_t v\mathbf{m}) + \frac{t}{2}[H(T_t v|\mathbf{m}) - H(u|\mathbf{m})] &\leq \frac{1}{K}\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) - \\ &\frac{\varepsilon^2}{8K} \int_0^1 I(\rho_s|\mathbf{m}) ds + \frac{\varepsilon^2}{8} \int_0^1 \int T_{st}\rho_{\zeta(s)} |\nabla \log T_{st}\rho_{\zeta(s)}|^2 ds d\mathbf{m} \end{aligned}$$

The $CD(\kappa/2, \infty)$ condition implies (see [BGL14, Section 5.7]),

$$I_{\mathbf{m}}(T_t \rho_s) \leq e^{-\kappa t} I_{\mathbf{m}}(\rho_s).$$

Moreover, dividing by $t > 0$,

$$\begin{aligned} \frac{1}{t}\mathcal{A}^\varepsilon(u\mathbf{m}, T_t v\mathbf{m}) - \frac{1}{Kt}\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) &\leq \frac{1}{2}[H(u|\mathbf{m}) - H(T_t v|\mathbf{m})] + \\ &\frac{\varepsilon^2}{8} \left[\int_0^1 \frac{e^{-\kappa t s}}{t} I_{\mathbf{m}}(\rho_{\zeta(s)}) ds - \frac{1}{Kt} \int_0^1 I_{\mathbf{m}}(\rho_s) ds \right]. \quad (5.14) \end{aligned}$$

For small positive t , by the definition of K in (5.10), a Taylor expansion of $1/K$ gives $1/K = 1 - \kappa/2t + o(t)$. For what concerns the terms inside the square brackets, they can be written as,

$$\int_0^1 I_{\mathbf{m}}(\rho_s) \underbrace{\left(\frac{1}{t} - \frac{1}{Kt}\right)}_{(*)} ds + \int_0^1 I_{\mathbf{m}}(\rho_{\zeta(s)}) \underbrace{\frac{e^{-\kappa s t} - 1}{t}}_{(**)} ds + \int_0^1 \frac{1}{t} \underbrace{[I_{\mathbf{m}}(\rho_{\zeta(s)}) - I_{\mathbf{m}}(\rho_s)]}_{(***)} ds. \quad (5.15)$$

Let us compute the three terms separately. Taylor expansions provide $(*) = (K-1)/Kt = \kappa/2 + o(t)$ and $(**) = (e^{-\kappa s t} - 1)/t = -\kappa s + o(t)$. After the change of variable $\zeta(s) = \sigma$, with $ds = 1/(1 + \kappa t \sigma K) d\sigma$ for the first term in $(***)$, we can write it as,

$$(***) = \int_0^1 I_{\mathbf{m}}(\mu_s) \left(\frac{K}{t + \kappa t^2 s K} - \frac{1}{t} \right) ds$$

A Taylor expansion of the term into the square brackets gives,

$$(***) = \int_0^1 (\kappa/2 - \kappa s + o(t)) I_{\mathbf{m}}(\rho_s) ds$$

Hence we obtained that (5.15) is equal to

$$\int_0^1 \kappa(1 - 2s + o(t)) I_{\mathbf{m}}(\rho_s) ds.$$

Raplacing it in the square brackets in (5.14), and taking the limit for $t \rightarrow 0^+$, yields,

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \mathcal{A}^\varepsilon(u \mathbf{m}, T_t v \mathbf{m}) - \frac{\kappa}{2} \mathcal{A}^\varepsilon(u \mathbf{m}, u \mathbf{m}) &\leq \frac{1}{2} [H(u|\mathbf{m}) - H(v|\mathbf{m})] \\ &\quad + \frac{\varepsilon^2}{8} \kappa \int_0^1 (1 - 2s) I_{\mathbf{m}}(\rho_s) ds \end{aligned}$$

Since $0 \leq s \leq 1$, it holds $\kappa(1 - 2s) \leq |\kappa|$, hence we obtain the desired result,

$$\frac{d}{dt} \Big|_{t=0} \mathcal{A}^\varepsilon(u \mathbf{m}, T_t v \mathbf{m}) - \frac{\kappa}{2} \mathcal{A}^\varepsilon(u \mathbf{m}, u \mathbf{m}) \leq \frac{1}{2} [H(u|\mathbf{m}) - H(v|\mathbf{m})] + \frac{\varepsilon^2}{8} |\kappa| \int_0^1 I_{\mathbf{m}}(\rho_s) ds.$$

▷

5.3.2 Second method

Here we prove Theorem 5.2.1 following the method proposed by Daneri and Savaré in [DS08] for the Wasserstein distance based on the approach of [OW05] and [Ott01]. We will show in particular that the same arguments are still valid for the entropic cost with some proper adaptation.

Again, we denote $(\mu_s)_{0 \leq s \leq 1}$ the entropic interpolation between $u \mathbf{m}$ and $v \mathbf{m}$ and $(\rho_s)_{0 \leq s \leq 1}$ its density with respect to the reversing measure \mathbf{m} . We consider the following perturbation of the entropic interpolation through the Kolmogorov semigroup, for all $t \geq 0$

$$(\rho_s^t)_{0 \leq s \leq 1} := (T_{st} \mu_s)_{0 \leq s \leq 1} \tag{5.16}$$

Let ϕ_s^t be defined by

$$\partial_s \mu_s^t + \nabla \cdot (\mu_s^t \nabla \phi_s^t) = 0 \tag{5.17}$$

We denote the action of the perturbed path as,

$$A_s^t := \int |\nabla \phi_s^t|^2 \rho_s^t d\mathbf{m} + \frac{\varepsilon^2}{4} \int |\nabla \log \rho_s^t|^2 \rho_s^t d\mathbf{m} \tag{5.18}$$

We present here some preliminary results that will be crucial in the following proof of Theorem 5.2.1.

Lemma 5.3.2 *Let A_s^t, ρ_s^t and ϕ_s^t be respectively like in (5.18), (5.16) and (5.17). Then the following formula holds*

$$\frac{1}{2} \frac{\partial}{\partial t} A_s^t + \frac{1}{2} \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) = -2s \int [\Gamma_2(\phi_s^t) + \frac{\varepsilon^2}{4} \Gamma_2(\log \rho_s^t)] \rho_s^t d\mathbf{m} \tag{5.19}$$

Furthermore, under $CD(\kappa/2, \infty)$

- if $\kappa = 0$,

$$\frac{\partial}{\partial t} A_s^t + \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) \leq 0 \tag{5.20}$$

- if $\kappa \neq 0$,

$$\frac{\partial}{\partial t} A_s^t + \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) \leq -\kappa s A_s^t \tag{5.21}$$

Proof

\triangleleft The proof of (5.19) is essentially the same of the proof of Lemma 4.3.5. For simplicity in the proof we assume $\varepsilon = 1$. Let us start by computing (i) and (ii) separately in

$$\frac{1}{2} \frac{\partial}{\partial t} A_s^t = \underbrace{\frac{1}{2} \frac{\partial}{\partial t} \int |\nabla \phi_s^t|^2 \rho_s^t \, d\mathbf{m}}_{(i)} + \underbrace{\frac{\varepsilon^2}{8} \frac{\partial}{\partial t} \int |\nabla \log \rho_s^t|^2 \rho_s^t \, d\mathbf{m}}_{(ii)} \quad (5.22)$$

We recall that the computation of (i) can be found in a more general setting in [DS08], but we decided to report it for completeness.

$$\begin{aligned} (i) &= \int \nabla \phi_s^t \cdot \frac{\partial}{\partial t} \nabla \phi_s^t \rho_s^t \, d\mathbf{m} + \frac{1}{2} \int |\nabla \phi_s^t|^2 \frac{\partial}{\partial t} \rho_s^t \, d\mathbf{m} \\ &= \int -\nabla \cdot \left(\rho_s^t \frac{\partial}{\partial t} \nabla \phi_s^t \right) \phi_s^t \, d\mathbf{m} + \frac{s}{2} \int |\nabla \phi_s^t|^2 L \rho_s^t \, d\mathbf{m} \\ &= \int \left[\frac{\partial^2}{\partial t \partial s} \rho_s^t + \nabla \cdot \left(\nabla \phi_s^t \frac{\partial}{\partial t} \rho_s^t \right) \right] \phi_s^t \, d\mathbf{m} + \frac{s}{2} \int L(|\nabla \phi_s^t|^2) \rho_s^t \, d\mathbf{m} \\ &= \int L \rho_s^t \phi_s^t - s L \nabla \cdot (\rho_s^t \nabla \phi_s^t) \phi_s^t + \nabla \cdot (s L \rho_s^t \nabla \phi_s^t) \phi_s^t + \frac{s}{2} L(|\nabla \phi_s^t|^2) \rho_s^t \, d\mathbf{m} \\ &= \int \rho_s^t L \phi_s^t - s \nabla \cdot (\rho_s^t \nabla \phi_s^t) L \phi_s^t - s L \rho_s^t |\nabla \phi_s^t|^2 + \frac{s}{2} L(|\nabla \phi_s^t|^2) \rho_s^t \, d\mathbf{m} \\ &= \int \rho_s^t L \phi_s^t - s \nabla \rho_s^t \cdot \nabla \phi_s^t L \phi_s^t - s (\Delta \phi_s^t)^2 \rho_s^t - s \rho_s^t L(|\nabla \phi_s^t|^2) + \frac{s}{2} L(|\nabla \phi_s^t|^2) \rho_s^t \, d\mathbf{m} \\ &= \int \rho_s^t L \phi_s^t - s \nabla \rho_s^t \cdot \nabla \phi_s^t L \phi_s^t - s \rho_s^t (L \phi_s^t)^2 - \frac{s}{2} \rho_s^t L(|\nabla \phi_s^t|^2) \, d\mathbf{m} \\ &= \int -\frac{1}{2} \nabla \rho_s^t \cdot \nabla \phi_s^t + s \rho_s^t [(L \phi_s^t)^2 + \nabla \phi_s^t \cdot \nabla L \phi_s^t] - \frac{s}{2} \rho_s^t (L \phi_s^t)^2 - \frac{s}{2} L(|\nabla \phi_s^t|^2) \rho_s^t \, d\mathbf{m} \\ &= \int -\frac{1}{2} \nabla \rho_s^t \cdot \nabla \phi_s^t \, d\mathbf{m} + s \int \left(\nabla \phi_s^t \cdot \nabla L \phi_s^t - \frac{1}{2} L(|\nabla \phi_s^t|^2) \right) \rho_s^t \, d\mathbf{m} \end{aligned}$$

We remark now that,

- By definition of Γ_2 operator 1.1.21

$$\nabla \phi_s^t \cdot \nabla L \phi_s^t - \frac{1}{2} L(|\nabla \phi_s^t|^2) = -2\Gamma_2(\phi_s^t). \quad (5.23)$$

- On the other hand it follows by Proposition 2.5.1 that

$$\int \frac{1}{2} \nabla \phi_s^t \cdot \nabla \rho_s^t \, d\mathbf{m} = \frac{1}{2} \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}).$$

Thanks to these two considerations, we can conclude that

$$(i) = -\frac{1}{2} \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) - 2s \int \Gamma_2(\phi_s^t) \rho_s^t \, d\mathbf{m}$$

Let us compute now the second part of (5.22), which does not appear in the case of the Wasserstein distance.

$$\begin{aligned}
(ii) &= \frac{\varepsilon^2}{8} \frac{\partial}{\partial t} \int |\nabla \log \rho_s^t|^2 \rho_s^t d\mathbf{m} \\
&= \frac{\varepsilon^2}{8} \int 2\nabla \log \rho_s^t \cdot \frac{\partial}{\partial t} \nabla \log \rho_s^t \rho_s^t + |\nabla \log \rho_s^t|^2 \frac{\partial}{\partial t} \rho_s^t d\mathbf{m} \\
&= \frac{\varepsilon^2}{4} \int \nabla \log \rho_s^t \cdot \left(\frac{\nabla(sL\rho_s^t)\rho_s^t - \nabla\rho_s^t(sL\rho_s^t)}{\rho_s^t} \right) d\mathbf{m} + \frac{\varepsilon^2}{8} s \int |\nabla \log \rho_s^t|^2 L\rho_s^t d\mathbf{m} \\
&= \frac{s\varepsilon^2}{4} \left[\int \nabla \log \rho_s^t \cdot \nabla L\rho_s^t - \nabla \log \rho_s^t \cdot \frac{\nabla\rho_s^t}{\rho_s^t} L\rho_s^t d\mathbf{m} + \frac{1}{2} \int |\nabla \log \rho_s^t|^2 L\rho_s^t d\mathbf{m} \right] \\
&= \frac{s\varepsilon^2}{4} \int \nabla \log \rho_s^t \cdot \nabla L\rho_s^t - |\nabla \log \rho_s^t|^2 L\rho_s^t + \frac{1}{2} |\nabla \log \rho_s^t|^2 L\rho_s^t d\mathbf{m} \\
&= \frac{s\varepsilon^2}{4} \int \nabla \log \rho_s^t \cdot \nabla L \log \rho_s^t \rho_s^t - \frac{1}{2} L(|\nabla \log \rho_s^t|^2) \rho_s^t d\mathbf{m} \\
&= -\frac{s\varepsilon^2}{2} \int \Gamma_2(\log \rho_s^t) \rho_s^t d\mathbf{m}
\end{aligned}$$

where the last equality is given again by (5.23). This concludes the proof of (5.19). The two inequalities (5.20) and (5.21) follow directly from the Bakry-Emery condition (1.32). Let us assume $\kappa \neq 0$,

$$\begin{aligned}
\frac{1}{2} \frac{\partial}{\partial t} A_s^t + \frac{1}{2} \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) &= -2s \int [\Gamma_2(\phi_s^t) + \frac{\varepsilon^2}{4} \Gamma_2(\log \rho_s^t)] \rho_s^t d\mathbf{m} \\
&\leq -\frac{\kappa}{2} s \int |\nabla \phi_s^t|^2 + \frac{\varepsilon^2}{4} |\nabla \log \rho_s^t|^2 \rho_s^t d\mathbf{m} \\
&= -\frac{\kappa}{2} s A_s^t
\end{aligned}$$

The case $\kappa = 0$ is trivial. \triangleright

The next Lemma is a generalization to the entropic setting of [DS08, Lem. 5.1].

Lemma 5.3.3 *For every function $f \in C^\infty([0, 1])$, and $(\mu_s, \phi_s)_{0 \leq s \leq 1}$ satisfying conditions (i)-(iii) in (4.2.5),*

$$\mathcal{A}^\varepsilon(um, vm) - \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{L_f f(s)} |\nabla \log \rho_s|^2 ds d\mathbf{m} \leq \frac{L_f}{2} \int_0^1 f(s) A_s ds,$$

where $L_f = \int_0^1 f(s)^{-1} ds$.

Proof

\triangleleft We consider a smooth increasing map $r : [0, 1] \rightarrow [0, 1]$ defined by

$$r(s) := L_f^{-1} \int_0^s \frac{1}{f(s)} ds, \quad L_f := \int_0^1 \frac{1}{f(s)} ds.$$

We define s the inverse function of r .

$$s(r) = r^{-1}(r) = r^{-1}(r(s)) = s(r(s)).$$

So,

$$s'(r(s)) = L_f f(s).$$

We consider the new reparametrized curve

$$\bar{\rho}_r(x) := \rho_{S(r)}(x), \quad \bar{\phi}_r(x) := s'(r)\phi_{S(r)}(x).$$

It can be easily verified that $(\bar{\rho}_r, \bar{\phi}_r)$ satisfies conditions (i)-(iii) in (5.6) thus, the Benamou-Brenier formulation implies that

$$\mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m}) \leq \frac{1}{2} \int_0^1 \bar{A}_r dr, \quad (5.24)$$

where,

$$\begin{aligned} \bar{A}_r &= \int |\nabla \bar{\phi}_r|^2 \bar{\rho}_r(x) d\mathbf{m} + \frac{\varepsilon^2}{4} \int |\nabla \log \bar{\rho}_r|^2 \bar{\rho}_r d\mathbf{m} \\ &= \int |\nabla \phi_{S(r)}|^2 \rho_{S(r)} s'(r)^2 d\mathbf{m} + \frac{\varepsilon^2}{4} \int |\nabla \log \rho_{S(r)}|^2 \rho_{S(r)} d\mathbf{m} \\ &= s'(r)^2 A_{S(r)} + \frac{\varepsilon^2}{4} (1 - s'(r)^2) \int |\nabla \log \rho_{S(r)}|^2 \rho_{S(r)} d\mathbf{m}. \end{aligned}$$

Integrating with respect to $r \in [0, 1]$,

$$\begin{aligned} \int_0^1 \bar{A}_r dr &= \int_0^1 s'(r)^2 A_{S(r)} dr + \frac{\varepsilon^2}{4} \int_0^1 (1 - s'(r)^2) \int |\nabla \log \rho_{S(r)}|^2 \rho_{S(r)} d\mathbf{m} \\ &= \int_0^1 s'(r(s)) A_s ds + \frac{\varepsilon^2}{4} \int_0^1 \frac{1 - s'(r(s))^2}{s'(r(s))} \int |\nabla \log \rho_s|^2 \rho_s ds d\mathbf{m} \\ &= L_f \int_0^1 f(s) A_s ds + \frac{\varepsilon^2}{4} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{L_f f(s)} |\nabla \log \rho_s|^2 \rho_s ds d\mathbf{m}, \end{aligned}$$

where the last equality is obtained with the change of variable $s(r) = s$. Together with (5.24) it follows that

$$\mathcal{A}^\varepsilon(\mathbf{u}\mathbf{m}, \mathbf{v}\mathbf{m}) \leq \frac{L_f}{2} \int_0^1 f(s) A_s ds + \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{L_f f(s)} |\nabla \log \rho_s|^2 \rho_s ds d\mathbf{m}$$

that is the claimed result. \triangleright

We have now all the elements to prove the main result of this section.

Proof of Theorem 5.2.1

\triangleleft Let us remark that thanks to the semigroup property, we can prove EVI only for $t = 0$ without loss of generality.

Let us start by multiplying the inequality (5.21) by $e^{\kappa st}$,

$$\frac{1}{2} \frac{\partial}{\partial t} A_s^t e^{\kappa st} + \frac{\kappa}{2} s A_s^t e^{\kappa st} + \frac{1}{2} \frac{\partial}{\partial s} H(\mu_s^t | \mathbf{m}) e^{\kappa st} \leq 0,$$

that can be written also as,

$$\frac{1}{2} \frac{\partial}{\partial t} (A_s^t e^{\kappa st}) + \frac{1}{2} \frac{\partial}{\partial s} (H(\mu_s^t | \mathbf{m}) e^{\kappa st}) \leq \frac{\kappa}{2} t e^{\kappa st} H(\mu_s^t | \mathbf{m}).$$

We integrate with respect to $s \in [0, 1]$,

$$\frac{1}{2} \frac{\partial}{\partial t} \int_0^1 e^{\kappa st} A_s^t ds + \frac{1}{2} e^{\kappa t} H(\mu_t^1 | \mathbf{m}) - \frac{1}{2} H(\mu_t^0 | \mathbf{m}) \leq \int_0^1 \frac{\kappa}{2} t e^{\kappa st} H(\mu_s^t | \mathbf{m}) ds.$$

By definition of ρ_s^t (5.16) we have $\mu_t^0 = u$, $\mu_t^1 = T_t v$. Then, we integrate with respect to $t \in [0, \tau]$ for some $\tau \geq 0$,

$$\frac{1}{2} \int_0^1 e^{\kappa s \tau} A_\tau^s ds - \frac{1}{2} \int_0^1 \tilde{A}_0^s ds - \frac{\tau}{2} H(u | \mathbf{m}) + \frac{1}{2} \int_0^\tau e^{\kappa t} H(T_t v | \mathbf{m}) dt \leq \int_0^\tau \int_0^1 \frac{\kappa}{2} t e^{\kappa st} H(\mu_s^t | \mathbf{m}) dt ds. \quad (5.25)$$

Since the entropy is decreasing along the semigroup $(T_t)_{t \geq 0}$,

$$\int_0^\tau e^{\kappa t} H(T_t v | \mathbf{m}) dt \geq H(T_\tau v | \mathbf{m}) E_\kappa(\tau),$$

where we denoted,

$$E_\kappa(\tau) := \int_0^\tau e^{\kappa t} dt = \frac{e^{\kappa \tau} - 1}{\kappa}.$$

By the Benamou-Brenier formulation (5.6) and the fact that $(\mu_s^0)_{0 \leq s \leq 1}$ coincides with the entropic interpolation between $u\mathbf{m}$ and $v\mathbf{m}$,

$$-\frac{1}{2} \int_0^1 A_s^0 ds = -\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}).$$

In order to control the first term on the left hand side of (5.25), we apply Lemma 5.3.3 with

$$f(s) = e^{\kappa s \tau}, \quad \text{hence} \quad L_f(\tau) = \frac{1 - e^{-\kappa \tau}}{\kappa \tau}, \quad \forall \tau > 0. \quad (5.26)$$

So from (5.25) we get,

$$\begin{aligned} \frac{1}{L_f(\tau)} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) - \mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) - \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{L_f^2 f(s)} |\nabla \log \rho_\tau^s|^2 \rho_\tau^s ds d\mathbf{m} \\ - \frac{\tau}{2} H(u | \mathbf{m}) + \frac{E_\kappa(\tau)}{2} H(T_\tau v | \mathbf{m}) \leq \int_0^\tau \int_0^1 \frac{\kappa}{2} t e^{\kappa st} H(\mu_s^t | \mathbf{m}) dt ds. \end{aligned} \quad (5.27)$$

• We consider first the case $\kappa \leq 0$. Being negative, we can then neglect the term on the right hand side.

We divide (5.27) by $\tau > 0$ and let $\tau \rightarrow 0^+$. Since $L_f(0) = 1$, (5.27) gives,

$$\begin{aligned} \frac{d}{d\tau} \left[\frac{1}{L_f(\tau)} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \right]_{\tau=0} - \lim_{\tau \rightarrow 0} \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{t L_f^2 f(s)} |\nabla \log \rho_s^\tau|^2 \rho_s^\tau ds d\mathbf{m} \\ \leq \frac{1}{2} [H(u | \mathbf{m}) - H(v | \mathbf{m})]. \end{aligned} \quad (5.28)$$

Let us study now each term separately.

The first term:

$$\begin{aligned} \frac{d}{d\tau} \left[\frac{1}{L_f(\tau)} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \right]_{\tau=0} &= \frac{d}{d\tau} \frac{1}{L_f(\tau)} \Big|_{\tau=0} \mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) + \frac{1}{L_f(0)} \frac{d}{d\tau} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \Big|_{\tau=0} \\ &= \frac{\kappa}{2} \mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) + \frac{d}{d\tau} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \Big|_{\tau=0} \end{aligned}$$

It can be easily checked indeed that the derivative of $(L_f(\tau))^{-1}$ in $\tau = 0$ is equal to $\kappa/2$.

The second term:

$$\lim_{\tau \rightarrow 0} \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{\tau L_f^2 f(s)} |\nabla \log \rho_s^\tau|^2 \rho_s^\tau ds d\mathbf{m}.$$

By definition of f and L_f (5.26),

$$\begin{aligned} \frac{1 - L_f^2 f(s)^2}{L_f^2 f(s)} &= \frac{1 - \left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2 f(s)^2}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2 f(s)} \\ &= \frac{e^{-\kappa s\tau}}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - e^{\kappa s\tau} \\ &\leq \frac{e^{-\kappa\tau}}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - e^{\kappa\tau} \end{aligned}$$

where the inequality follows from the fact that $\kappa \leq 0$ and $0 \leq s \leq 1$. With the Taylor expansion we can conclude that

$$\begin{aligned} \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left(\frac{e^{-\kappa\tau}}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - e^{\kappa\tau} \right) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \left(\frac{-\kappa\tau + 3(\kappa\tau)^2/4 + o(\tau^2)}{1 - \kappa\tau + (\kappa\tau)^2/4 + o(\tau^2)} \right) \\ &= -\kappa. \end{aligned}$$

And finally we get,

$$\begin{aligned} \lim_{\tau \rightarrow 0} \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{\tau L_f^2 f(s)} |\nabla \log \rho_s^\tau|^2 \rho_s^\tau ds d\mathbf{m} &\leq \lim_{\tau \rightarrow 0} \frac{\varepsilon^2}{8} \frac{1}{\tau} \left(\frac{e^{-\kappa\tau}}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - e^{\kappa\tau} \right) \int_0^1 I_{\mathbf{m}}(\rho_s^\tau) ds \\ &= -\frac{\varepsilon^2}{8} \kappa \int_0^1 I_{\mathbf{m}}(\rho_s) ds. \end{aligned}$$

Coming back to (5.28), it reduces to

$$\frac{d}{d\tau} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \Big|_{\tau=0} + \kappa \left[\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) + \frac{\varepsilon^2}{8} \int_0^1 I_{\mathbf{m}}(\rho_s) ds \right] \leq \frac{1}{2} [H(u|\mathbf{m}) - H(v|\mathbf{m})].$$

• Let us consider now the case $\kappa \geq 0$. Since the entropy is decreasing along the heat flow, and convex along the entropic interpolation, the term on the right hand side of (5.25) is bounded by,

$$\begin{aligned} \int_0^\tau \int_0^1 \frac{\kappa}{2} t e^{\kappa t s} H(\mu_s^t | \mathbf{m}) ds dt &\leq \int_0^\tau \int_0^1 \frac{\kappa}{2} t e^{\kappa t s} H(\mu_s | \mathbf{m}) ds dt \\ &\leq \int_0^\tau \int_0^1 \frac{\kappa}{2} t e^{\kappa t s} [(1-s)H(u|\mathbf{m}) + sH(v|\mathbf{m})] ds dt \\ &\leq \frac{\kappa}{2} \tau^2 e^{\kappa\tau} [H(v|\mathbf{m}) + H(u|\mathbf{m})]. \end{aligned}$$

For what concerns the bounds of the terms of the left hand side of (5.25), the first one remains unchanged, while the second one changes as follows,

$$\begin{aligned}
\frac{1 - L_f^2 f(s)^2}{L_f^2 f(s)} &= \frac{1 - \left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2 f(s)^2}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2 f(s)} \\
&= \frac{e^{-\kappa s\tau}}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - e^{\kappa s\tau} \\
&\leq \frac{1}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{\tau \rightarrow 0} \frac{1}{\tau} \left(\frac{1}{\left(\frac{1 - e^{-\kappa\tau}}{\kappa\tau}\right)^2} - 1 \right) &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \frac{\kappa\tau - (\kappa\tau)^2/4 + o(\tau^2)}{1 - \kappa\tau + (\kappa\tau)^2/4 + o(\tau^2)} \\
&= \kappa.
\end{aligned}$$

So this time,

$$\lim_{\tau \rightarrow 0} \frac{\varepsilon^2}{8} \int_0^1 \int \frac{1 - L_f^2 f(s)^2}{\tau L_f^2 f(s)} |\nabla \log \mu_s^\tau|^2 \mu_s^\tau ds d\mathbf{m} \leq \kappa \int_0^1 \int |\nabla \log \mu_s|^2 \mu_s ds, d\mathbf{m}$$

which has opposite sign with respect to the previous case $\kappa \leq 0$. So, as before,

$$\frac{d}{d\tau} \mathcal{A}^\varepsilon(u\mathbf{m}, T_\tau v\mathbf{m}) \Big|_{\tau=0} + \kappa \left[\mathcal{A}^\varepsilon(u\mathbf{m}, v\mathbf{m}) - \frac{\varepsilon^2}{8} \int_0^1 I_{\mathbf{m}}(\rho_s) ds \right] \leq \frac{1}{2} [H(u|\mathbf{m}) - H(v|\mathbf{m})].$$

that is the claimed result. \triangleright

Chapter 6

Functional inequalities when $\varepsilon = 1$

In the previous chapters we used the Schrödinger problem to prove functional inequalities by recovering the optimality of McCann geodesics at the limit $\varepsilon \rightarrow 0^+$. In this chapter we prove that the Schrödinger problem is a useful tool itself for a fixed $\varepsilon = 1$. We will show the classical Gaussian concentration and Prékopa-Leindler inequalities by using only the statistical physical nature of the Schrödinger problem, together with a simple coupling technique.

6.1 Concentration inequality

In the last fifty years the study of the concentration measure phenomenon concerned a wide number of mathematicians from different fields, such as geometry, functional analysis, discrete mathematics, complexity theory and probability theory. See the textbook of Ledoux [Led01] for a complete description of the subject from all the different points of view.

In an abstract metric measure space (\mathcal{X}, d, m) equipped with a Borel probability measure m and distance d , a concentration inequality writes as

$$m(A^r) \geq 1 - C_p(r), \quad \forall r \geq 0, \forall A \subset \mathcal{X} : m(A) \geq p$$

where A is a Borel subset, $0 < p < 1$ and $A^r := \{x \in \mathcal{X}; d(x, A) \leq r\}$ is the r -enlargement of the subset A . The function C_p is called *concentration profile* of the measure m with respect to the distance d . It was Marton for the first time in 1986 in [Mar86] to provide a link between transport inequalities and concentration. Through an easy and elegant argument she proved that if a measure satisfies a transport-entropy inequality of the form $\alpha(\mathcal{T}_d) \leq H$ then it satisfies a concentration inequality with profile $C_p(r) = \exp\{-\alpha(r-r_0)\}$ for any bijection $\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, any measurable subset A and any $r \geq r_0 = \alpha^{-1}(\log 2)$. See for instance [GL10] for more details.

Here we propose a new proof of the concentration inequality of statistical mechanics nature. We will consider the measure of a set as the relative entropy of the uniform measure on the set, with respect to a prescribed reference measure, and then derive the result from the dynamics of the particles system that is optimal for a particular Schrödinger problem. The setting of our proof is the following.

We consider the state space $\mathcal{X} = \mathbb{R}^n$ and denote as usual $\Omega = C([0, 1], \mathbb{R}^n)$ the set of continuous paths. We fix the reference path measure $R \in \mathcal{P}(\Omega)$, with infinitesimal generator L defined for all functions $u : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}$, by

$$Lu = \Delta u/2 + b \cdot \nabla u \tag{6.1}$$

where b is a locally Lipschitz vector field and satisfies

$$[b(t, y) - b(t, x)] \cdot (y - x) \leq -\kappa|x - y|^2/2, \quad \forall x, y \in \mathbb{R}^n, t \geq 0 \quad (6.2)$$

for some $\kappa \in \mathbb{R}$. We denote the transition probability kernel $m^{x,t}(\cdot) := R^x(X_t \in \cdot) = R(X_t \in \cdot | X_0 = x) \in P(\mathbb{R}^n)$ for all $t > 0$.

Theorem 6.1.1 (Concentration inequality) *Let b be a locally Lipschitz vector field, verifying hypothesis (6.2); for any $0 < p < 1$, $t > 0$, $x \in \mathbb{R}^n$ and any Borel measurable subset $A \subset \mathbb{R}^n$ such that $m^{x,t}(A) \geq p$, we have*

$$m^{x,t}(A^r) \geq 1 - \exp \left\{ -\frac{[r - r(p, \kappa, t)]^2}{2\tau(\kappa, t)} \right\}, \quad \forall r \geq r(p, \kappa, t)$$

where $A^r := \{x \in \mathbb{R}^n; d(x, A) \leq r\}$ is the r -enlargement of the subset A , and

$$\tau(\kappa, t) = \begin{cases} (1 - e^{-\kappa t})/\kappa & \text{if } \kappa \neq 0 \\ t & \text{if } \kappa = 0, \end{cases} \quad (6.3)$$

and $r(p, \kappa, t) := \sqrt{2\tau(\kappa, t) \log(1/p)}$.

Remarks 6.1.2 *We denote,*

$$C_{p,t}^\kappa(r) = \exp \left\{ -\frac{[r - r(p, \kappa, t)]^2}{2\tau(\kappa, t)} \right\}.$$

- (a) *For any fixed $r > 0$, $0 < p < 1$ and $\kappa \geq 0$, $t \mapsto C_{p,t}^\kappa(r)$, $t \mapsto \tau(\kappa, t)$ and $t \mapsto r(p, \kappa, t)$ are increasing functions. This reflects the fact that $m^{x,t}$ starting from the Dirac measure $m^{x,0} = \delta_x$ loses concentration as time increases.*
- (b) *For any fixed $r > 0$, $0 < p < 1$ and $t > 0$, $\kappa \mapsto C_{p,t}^\kappa(r)$, $t \mapsto r(p, \kappa, t)$ are decreasing functions. This reflects the fact that increasing curvature enforces concentration.*
- (c) *In the flat case $\kappa = 0$ we have $\tau(o, t) = t$ that is the usual Brownian concentration.*

We prove the theorem after presenting some preliminary results about the two main tools that we will need in the proof: the *coupling* technique and a *light* version of the Schrödinger entropy minimization problem.

In the following, we consider an abstract probability space, $(\Theta, \mathcal{F}, \mathbb{P})$ and build a process X such that $X_{\#}\mathbb{P} = R$ where R is the reference path measure that we introduced above. Under hypothesis (6.2), R is the law of the \mathbb{R}^n -valued random process X , with SDE,

$$dX_t = b(t, X_t) dt + dW_t, \quad t \geq 0 \quad (6.4)$$

where W is a $(\mathbb{P}, \mathcal{F})$ -Brownian motion. In addition, let P and Q be two other path measures in $\mathcal{P}(\Omega)$, such that $P_0 = Q_0 = R_0$ and $H(P|R), H(Q|R) < \infty$. On the same abstract probability space $(\Theta, \mathcal{F}, \mathbb{P})$, there exists two \mathbb{R}^n -valued random processes Y and Z such that $Y_{\#}\mathbb{P} = P$ and $Z_{\#}\mathbb{P} = Q$, and satisfy respectively the SDEs,

$$\begin{aligned} dY_t &= [b(t, Y_t) + u_t] dt + dW_t, & t \geq 0 \\ dZ_t &= [b(t, Z_t) + v_t] dt + dW_t, & t \geq 0 \end{aligned}$$

for the same vector field $b(t, \cdot)$ and the same $(\mathbb{P}, \mathcal{F})$ -Brownian motion W as in (6.4), with initial conditions $Y_0 = Z_0 = X_0$, \mathbb{P} -a.e., where u_t and v_t are two predictable vector fields given by Girsanov's Theorem 1.1.14.

Lemma 6.1.3 For all $t > 0$,

$$|Z_t - Y_t|^2 \leq \tau(\kappa, t) \int_0^t |v_s - u_s|^2 ds < \infty$$

for the same constant $\tau(\kappa, t)$ as in (6.3).

Proof

◁ By Ito's formula,

$$\begin{aligned} d|Z_s - Y_s|^2 &= 2(Z_s - Y_s) \cdot d(Z_s - Y_s) \\ &= 2(Z_s - Y_s) \cdot [b(s, Z_s) - b(s, Y_s) + v_s - u_s] ds \\ &\leq -\kappa |Z_s - Y_s|^2 ds + 2(Z_s - Y_s) \cdot (v_s - u_s) ds. \end{aligned}$$

An integration in $0 \leq s \leq t$ and the initial condition $Z_0 = Y_0 = x$, lead to

$$|Z_t - Y_t|^2 \leq \int_0^t -\kappa |Z_s - Y_s|^2 + 2(Z_s - Y_s) \cdot (v_s - u_s) ds.$$

We can use here the following nonlinear Gronwall's type inequality, which states that for $c \geq 0$, and $a(s), b(s)$ continuous functions, if for $t \geq 0$,

$$u(t) \leq c + \int_0^t [a(s)u(s) + b(s)u(s)^{1/2}] ds,$$

then

$$u(t) \leq \left\{ c^{1/2} \exp \left[\frac{1}{2} \int_0^t a(s) ds \right] + \frac{1}{2} \int_0^t b(s) \exp \left[(1/2) \int_s^t a(r) dr \right] ds \right\}^2.$$

By applying this result to $u(t) = |Z_t - Y_t|^2$, $a(s) = -\kappa$, $b(s) = 2(v_s - u_s)$ and $c = 0$, we can conclude that

$$\begin{aligned} |Z_t - Y_t|^2 &\leq \left[\int_0^t (v_s - u_s) \exp \left\{ -\frac{\kappa}{2} \int_s^t dr \right\} ds \right]^2 \\ &\leq \left[\int_0^t (v_s - u_s) \exp \left\{ -\frac{\kappa}{2} (t - s) \right\} ds \right]^2 \\ &\leq \left[\left(\int_0^t |v_s - u_s|^2 ds \right)^{1/2} \left(\int_0^t e^{-\kappa(t-s)} ds \right)^{1/2} \right]^2 \\ &= \frac{1 - e^{-\kappa t}}{\kappa} \int_0^t |v_s - u_s|^2 ds, \end{aligned}$$

that is the desired result with $\tau(\kappa, t) := (1 - e^{-\kappa t})/\kappa$. ▷

The following Lemma is about a particular case of the entropy minimization problem (2.1). Let $R \in \mathcal{P}(\Omega)$ be associated to the generator (6.1) with some $R_0 \in \mathcal{P}(\mathbb{R}^n)$ and $R_T = m \in \mathcal{P}(\mathbb{R}^n)$ for $T > 0$.

Lemma 6.1.4 *Given the following entropy minimization problem,*

$$\inf\{H(Q|R); Q \in \mathcal{P}(\Omega), \text{ s.t. } Q_T(E) = 1\} \quad (6.5)$$

with instant $T > 0$ fixed, E a Borel subset of \mathbb{R}^n and R the same reference path measure as above. The unique solution of this problem is $Q^E = \mathbb{1}(X_T)/m^{x,T}(E)R := \rho_E(X_T)R \in \mathcal{P}(\Omega)$. The minimal value of the relative entropy is,

$$H(Q^E|R) = -\log m^{x,T}(E).$$

Remark 6.1.5 *It is indeed a Schrödinger problem in which the infimum runs among all the path measures Q , that concentrate on the subset E at time $T > 0$, whereas the initial constraint on Q_0 is forced by the initial condition on R , being the Dirac measure $R_0 = \delta_x$ and by the condition $Q_0 \ll R_0$.*

Proof

\triangleleft In the following, since $x \in \mathbb{R}^n$ and $T > 0$ are fixed, we will drop the index and use the lighter notation $m = m^{x,T}$. Let $Q \in \mathcal{P}(\Omega)$ be any path measure satisfying the marginal constraint in (6.5),

$$\begin{aligned} H(Q|R) &= E_Q \log \frac{dQ}{dR} = E_Q \log \left(\frac{dQ}{dQ^E} \frac{dQ^E}{dR} \right) = H(Q|Q^E) + E_Q \log \frac{dQ^E}{dR} \\ &\geq E_Q \log \rho^E(X_T) = E_{Q^E} \log \rho^E(X_T) = H(Q^E|R), \end{aligned}$$

where the inequality is a consequence of the non negativity of the relative entropy. The uniqueness of the solution is given by the strictly convex nature of the Schrödinger problem. It follows that the minimal value achieved is,

$$\begin{aligned} H(Q^E|R) &= E_{Q^E} \log \frac{dQ^E}{dR} = E_{Q^E} \log \frac{\mathbb{1}(X_T)}{m(E)} \\ &= \int_E \frac{1}{m(E)} \log \frac{1}{m(E)} dm = -\log m(E). \end{aligned} \quad (6.6)$$

And this completes the proof. \triangleright

We are ready now to prove our main result of this section, the concentration inequality.

Proof of Theorem 6.1.1

\triangleleft Consider problem (6.5) respectively for the subset A and $B = \mathbb{R}^n \setminus A^c$. We denote $P^{A,T}$ and $P^{B,T}$ the minimizers, and \mathbf{u}_t^A and \mathbf{u}_t^B the two predictable vector fields whose existence is assured by the Girsanov theory, and such that

$$H(P^{A,T}|R) = E_{P^{A,T}} \int_0^T \frac{|\mathbf{u}_t^A|^2}{2} dt, \quad H(P^{B,T}|R) = E_{P^{B,T}} \int_0^T \frac{|\mathbf{u}_t^B|^2}{2} dt. \quad (6.7)$$

This ensures in particular that $\int_0^\infty |\mathbf{u}_t^A|^2, \int_0^\infty |\mathbf{u}_t^B|^2 < \infty$. As we have already done in Lemma 6.1.3, we construct on an abstract probability space $(\Theta, \mathcal{F}, \mathbb{P})$ two random \mathbb{R}^n -valued processes Y^A and Y^B such that

$$Y_\#^A \mathbb{P} = P^A, \quad Y_\#^B \mathbb{P} = P^B$$

associated with the SDEs,

$$\begin{aligned} dY_t^A &= (b(t, Y_t^A) + u_t^A) dt + dW_t, & t \geq 0, \\ dY_t^B &= (b(t, Y_t^B) + u_t^B) dt + dW_t, & t \geq 0, \end{aligned}$$

and the initial constraint translates to the initial condition $X_0 = Y_0^A = Y_0^B = x$, \mathbb{P} -a.e. By the choice of the sets A and B it is a trivial observation that $r \leq |Y_T^A - Y_T^B|$; on the other hand by Lemma 6.1.3 it holds that,

$$r^2 \leq |Y_T^A - Y_T^B|^2 \leq \tau(\kappa, T) \int_0^T |u_t^A - u_t^B|^2 dt. \quad (6.8)$$

By the basic inequality, $-2ab \leq (as)^2 + (b/s)^2$, valid for all $s > 0$, we can bound (6.8) on the right hand side by,

$$\begin{aligned} r^2 &\leq |Y_T^A - Y_T^B|^2 \leq \tau(\kappa, T) \left[(1+s) \mathbb{E} \int_0^T |u_t^A|^2 dt + \left(1 + \frac{1}{s}\right) \mathbb{E} \int_0^T |u_t^B|^2 \right] \\ &\stackrel{(6.7)}{=} 2\tau(\kappa, T) \left[(1+s) H(P^{A,T}|R) + \left(1 + \frac{1}{s}\right) H(P^{B,T}|R) \right] \\ &\stackrel{(6.6)}{=} -2\tau(\kappa, T) \left[(1+s) \log m(A) + \left(1 + \frac{1}{s}\right) \log m(B) \right] \\ &\leq -2\tau(\kappa, T) \left[(1+s) \log p + \left(1 + \frac{1}{s}\right) \log m(B) \right] \end{aligned}$$

where the last inequality follows from the hypothesis $m(A) \geq p$. After rearranging the terms, and replacing $m(B) = 1 - m(A^r)$, we get

$$1 - m(A^r) \leq \exp \left\{ -\frac{1}{2\tau(\kappa, T)} \left[r^2 \frac{s}{1+s} + 2s \log p \tau(\kappa, T) \right] \right\}, \quad \forall s > 0$$

Optimizing in s , it can be easily verified that the best value is $\tilde{s} = r^2 / \sqrt{2\tau(\kappa, T) \log(1/p)} - 1$ providing,

$$m(A^r) \geq 1 - \exp \left\{ -\frac{(r - r(T, \kappa, p))^2}{2\tau(\kappa, T)} \right\}$$

where $r(T, \kappa, p) := \sqrt{2\tau(\kappa, T) \log(1/p)}$. \triangleright

Remark 6.1.6 *The same argument can be used also to provide a proof of the Talagrand inequality. Let $R \in \mathcal{P}(\Omega)$ be the same reference measure as before, associated to the SDE (6.4) and same initial condition $R_0 = \delta_x$ for some $x \in \mathbb{R}^n$. Consider the Schrödinger problem in Lemma 6.1.4 with marginal constraint at time $T > 0$ given by $P_T = fR_T$ for some real function f on \mathbb{R}^n , such that $\int f dR_T = 1$. By the Girsanov theory, there exists a predictable vector field u_t such that the minimizer $P \in \mathcal{P}(\Omega)$ is the law of the random process with SDE*

$$dY_t = (b(t, Y_t) + u_t) dt + dW_t, \quad \forall t > 0$$

and initial condition $P_0 = \delta_x$. By Lemma 6.1.3, it holds that $|\mathbf{Y}_T - \mathbf{X}_T|^2 \leq \tau(\kappa, T) \int_0^T |\mathbf{u}_t|^2 dt$. It follows that

$$W_2^2(P_T^x, R_T^x) \leq \mathbb{E}(|\mathbf{Y}_T - \mathbf{X}_T|^2) \leq \tau(\kappa, T) \mathbb{E} \int_0^T |\mathbf{u}_t|^2 dt = 2\tau(\kappa, T) H(P_T^x | R_T^x).$$

Letting $T \rightarrow \infty$ and thanks to the ergodic property we obtain the classic Talagrand inequality,

$$W_2^2(P, R) \leq \frac{2}{\kappa} H(P|R).$$

6.2 Prékopa-Leindler inequality

The Prékopa-Leindler inequality is the functional form of the Brunn-Minkowski inequality. It is been used in [BL00, Gen08] to prove other functional inequalities such as the Brascamp-Lieb or the logarithmic Sobolev inequality. Known from the 70s [Pré71], among all the proofs of this inequality the two main contributions were given by Borell [Bor00] who proposes a proof via Brownian motion, and, independently, by McCann and Barthe (in the Gaussian case), see for instance [Vil03], who use optimal transport.

The proof we propose here is in the same line of the proof of the concentration inequality, via coupling techniques. Our proof is very close to the works of Lehec [Leh13], [Leh14] (see also [Leh15] for the Riemannian setting) and the paper of Cordero-Erasquin and Maurey [CEM17]. In both cases Borell's formula is used to prove functional inequalities, among which the Prékopa-Leindler inequality. The Borell formula states that,

$$\log \left(\int_{\mathbb{R}^n} e^f d\gamma \right) = \sup_u \left[\mathbb{E} \left(f \left(W_1 + \int_0^1 u_s ds \right) - \frac{1}{2} \int_0^1 |u_s|^2 ds \right) \right]$$

where γ denotes the standard Gaussian measure on \mathbb{R}^n , W is the standard Brownian motion and f is a bounded real function on \mathbb{R}^n . Roughly speaking, the Borell formula is obtained by the variational formula for the relative entropy (1.2) and the Girsanov Theorem 1.1.14, which are also the main ingredients of our proof.

Theorem 6.2.1 (Prékopa-Leindler inequality) *Let θ, ψ, ψ' be three measurable $[-\infty, \infty)$ -valued functions on \mathbb{R}^n satisfying for each $0 < \lambda < 1$*

$$\theta((1-\lambda)x + \lambda x') \geq (1-\lambda)\psi(x) + \lambda\psi'(x') \quad \forall x, x' \in \mathbb{R}^n. \quad (6.9)$$

Then, for any $T > 0$,

$$\log \int_{\mathbb{R}^n} e^\theta d\gamma_T \geq (1-\lambda) \log \int_{\mathbb{R}^n} e^\psi d\gamma_T + \lambda \log \int_{\mathbb{R}^n} e^{\psi'} d\gamma_T,$$

where $\gamma_T := \mathcal{N}(0, TId)$.

The proof of this theorem is in the same spirit of the one of Theorem 6.1.1, and we will use similar tools. Moreover, a key role will be played by the following variational formula for the log-Laplace transform of the probability R .

Proposition 6.2.2 *Let Σ be a measurable space, and $R \in P(\Sigma)$; for any $f : \Sigma \rightarrow [-\infty, \infty)$,*

$$\log \int_{\Sigma} e^f dR = \sup \left\{ \int_{\Sigma} f dQ - H(Q|R) : Q \in P(\Sigma) \int_{\Sigma} f_+ dQ < \infty \right\}. \quad (6.10)$$

Proof

◁ The proof is very close to the one of the variational formula for the relative entropy (1.2) but we recover it here for completeness. The starting point is the elementary inequality $st \leq t \log t - t + e^s$ for all $s \in \mathbb{R}$ and $t \geq 0$ with equality if and only if $t = e^s$. We choose $s = f(x)$ and $t = dQ/dR$ for some $Q \in \mathcal{M}_+(\Omega)$ and integrate with respect to R to obtain

$$\int e^f dR \geq \int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega).$$

It implies,

$$\int e^f dR \geq \sup \left\{ \int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega); Q \in \mathcal{M}_+(\Omega), \int f_+ dQ < \infty \right\}.$$

Hence, the case of equality $e^f = dQ/dR$ is the limit of the sequence $Q_n = e^{\min\{f, n\}} R$, it provides,

$$\int e^f dR = \sup \left\{ \int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega); Q \in \mathcal{M}_+(\Omega), \int f_+ dQ < \infty \right\}.$$

We replace Q by aQ for some $a \geq 0$, leading

$$\int e^f dR = \sup a \left\{ \int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega) \log a; Q \in \mathcal{M}_+(\Omega), \int f_+ dQ < \infty \right\}.$$

Since,

$$\begin{aligned} \sup_{a \geq 0} \left\{ a \left(\int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega) \log a \right) \right\} \\ = Q(\Omega) \exp \left\{ \int f dQ/Q(\Omega) - \int \log \frac{dQ}{dR} dQ/Q(\Omega) \right\} \end{aligned}$$

we can write,

$$\begin{aligned} \int e^f dR &= \sup \left\{ a \left(\int f dQ - \int \log \frac{dQ}{dR} dQ + Q(\Omega) \log a \right); a \geq 0; Q \in \mathcal{M}_+(\Omega), \int f_+ dQ < \infty \right\} \\ &= \exp \sup \left\{ \int f dQ/Q(\Omega) - \int \log \frac{dQ/Q(\Omega)}{dR} dQ/Q(\Omega); Q \in \mathcal{M}_+(\Omega), \int f_+ dQ < \infty \right\} \end{aligned}$$

and this completes the proof. ▷

Proof of Theorem 6.2.1

◁ We denote $\Omega := C([0, 1], \mathbb{R}^n)$ the space of the \mathbb{R}^n -valued continuous paths defined for $t \in [0, 1]$; as reference path measure $R \in P(\Omega)$ we fix the Wiener measure with $R_0 = \delta_0$ and for any $T > 0$, $R_T = \gamma_T := \mathcal{N}(0, TId)$, the standard Gaussian measure on \mathbb{R}^n . For $f = \psi(X_T)$, the formula (6.10) when R is the Wiener measure, writes as,

$$\log \int_{\mathbb{R}^n} e^\psi d\gamma_T = \sup_{Q_T \in P(\mathbb{R}^n)} \left\{ \int_{\mathbb{R}^n} \psi dQ_T - H(Q_T|\gamma_T) \right\}. \quad (6.11)$$

It can be verified that the maximizer of this problem is given by $\hat{Q}_T = e^\psi \gamma_T / \mathcal{Z}$, $\in P(\mathbb{R}^n)$, where $\mathcal{Z} = \int e^\psi d\gamma_T$ is a renormalization factor. Indeed

$$\begin{aligned}
\int_{\mathbb{R}^n} \psi d\hat{Q}_T - H(\hat{Q}_T|\gamma_T) &= \frac{1}{\mathcal{Z}} \int \psi e^\psi d\gamma_T - \int \log\left(\frac{e^\psi}{\mathcal{Z}}\right) \frac{e^\psi}{\mathcal{Z}} d\gamma_T \\
&= \log \mathcal{Z} \\
&= \log \int_{\mathbb{R}^n} e^\psi d\gamma_T.
\end{aligned}$$

In addition we observe that (6.11) is equivalent to the following entropy minimization problem,

$$\inf \left\{ H(P|R); P \in P(\Omega) \text{ s.t. } P_T = \hat{Q}_T = \frac{e^\psi}{\mathcal{Z}} \gamma_T \right\}. \quad (6.12)$$

Remark that as in (6.5) it is a Schrödinger problem, where the final constraint in $T > 0$ is explicit, whereas the initial condition is forced by $R_0 = \delta_0$ and the condition $P_t \ll R_t$ for all $t \geq 0$. We denote $\hat{P} \in P(\Omega)$ the minimizer in (6.12), and see that

$$\begin{aligned}
H(\hat{P}|R) &= E_{\hat{P}} \left(\log \frac{d\hat{P}}{dR} \right) = E_{\hat{P}} \left(\log \frac{e^{\psi(X_T)}}{\mathcal{Z}} \right) = \int_{\mathbb{R}^n} \log \left(\frac{e^{\psi(x)}}{\mathcal{Z}} \right) d\hat{P}_T \\
&= \int_{\mathbb{R}^n} \log \left(\frac{d\hat{P}_T}{d\gamma_T} \right) d\hat{P}_T = H(\hat{P}_T|\gamma_T).
\end{aligned}$$

As in the case of the concentration inequality, we construct an abstract space $(\Theta, \mathcal{F}, \mathbb{P})$. By Girsanov's Theorem 1.1.14, there exists a predictable drift \mathbf{u}_t such that \hat{P} will be the law of the process with stochastic equation,

$$dY^u = \mathbf{u}_t dt + dW_t, \quad \forall t > 0,$$

and,

$$H(P|R) = H(P_T|\gamma_T) = \mathbb{E} \left(\int_0^\infty \frac{|\mathbf{u}_t|^2}{2} \right). \quad (6.13)$$

Let us consider (6.11) and (6.12) with a different function $\psi'(X_T)$. The optimal $P' \in \mathcal{P}(\Omega)$ then is the law of the process $Y^{u'}$ with drift u' and verifies (6.13).

Let θ be a measurable function on \mathbb{R}^n satisfying for $0 < \lambda < 1$

$$\theta((1-\lambda)x + \lambda x') \geq (1-\lambda)\psi(x) + \lambda\psi'(x') \quad \forall x, x' \in \mathbb{R}^n. \quad (6.14)$$

By the Borell's formula (6.10),

$$\begin{aligned}
\log \int e^\theta d\gamma_T &= \sup \left\{ \int \theta dQ_T - H(Q|\gamma_T) : \int \theta_+ dQ_T < \infty \right\} \\
&\geq \int \theta d\hat{Q}_T - H(\hat{Q}_T|\gamma_T),
\end{aligned} \quad (6.15)$$

for a specific $\hat{Q} \in \mathcal{P}(\Omega)$. We consider in particular \hat{Q} being the law of the process given by the linear combination of Y^u and $Y^{u'}$,

$$Z_t = (1-\lambda)Y_t^u + \lambda Y_t^{u'}.$$

We provide a lower bound for the two terms in (6.15), separately. By definition of \hat{Q} ,

$$\begin{aligned} \int \theta d\hat{Q}_T &= \mathbb{E}(\theta(Z_T)) = \mathbb{E}[\theta((1-\lambda)Y_T^u + \lambda Y_T^{u'})] \stackrel{(6.14)}{\geq} \mathbb{E}[(1-\lambda)\psi(Y_T^u) + \lambda\psi'(Y_T^{u'})] \\ &= (1-\lambda)\mathbb{E}(\psi(Y_T^u)) + \lambda\mathbb{E}(\psi'(Y_T^{u'})) = (1-\lambda) \int_{\Omega} \psi(X_T) dP + \lambda \int_{\Omega} \psi'(X_T) dP'. \end{aligned}$$

Again by definition, \hat{Q} has drift $\mathbf{v}_t = (1-\lambda)\mathbf{u}_t + \lambda\mathbf{u}'_t$, hence

$$\begin{aligned} H(\hat{Q}_T|\gamma_T) &= \mathbb{E} \int_0^\infty \frac{1}{2} |\mathbf{v}_t|^2 dt = \mathbb{E} \int_0^\infty \frac{1}{2} |(1-\lambda)\mathbf{u} + \lambda\mathbf{u}'|^2 dt \\ &\leq \frac{1}{2} \mathbb{E} \int_0^\infty (1-\lambda)|\mathbf{u}_t|^2 + \lambda|\mathbf{u}'_t|^2 dt \\ &= (1-\lambda)H(P_T|\gamma_T) + \lambda H(P'_T|\gamma_T). \end{aligned}$$

Therefore (6.15) can be controlled by,

$$\log \int e^\theta d\gamma_T \geq (1-\lambda) \int \psi dP_T + \lambda \int \psi' dP'_T - (1-\lambda)H(P_T|\gamma_T) - \lambda H(P'_T|\gamma_T).$$

But since P and P' are the maximizer of (6.10) respectively for ψ and ψ' , we can conclude that,

$$\log \int_{\mathbb{R}^n} e^\theta d\gamma_T \geq (1-\lambda) \log \int_{\mathbb{R}^n} e^\psi d\gamma_T + \lambda \log \int_{\mathbb{R}^n} e^{\psi'} d\gamma_T.$$

▷

We can actually generalize this result to any log-concave measure m . Let $\mathbf{m} = e^{-V} \in M(\mathbb{R}^n)$, where the potential $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is C^2 and satisfies

$$\nabla^2 V \geq \kappa \text{Id}, \quad \text{for some } \kappa \in \mathbb{R}. \quad (6.16)$$

Corollary 6.2.3 *Let θ, ψ, ψ' be three measurable $[-\infty, \infty)$ -valued functions on \mathbb{R}^n satisfying for each $0 < \lambda < 1$*

$$\theta((1-\lambda)x + \lambda x') \geq (1-\lambda)\psi(x) + \lambda\psi'(x') - \kappa' \frac{\lambda(1-\lambda)}{2} |x' - x|^2,$$

for all $x, x' \in \mathbb{R}^n$ and some $\kappa' < \kappa$. Then,

$$\log \int_{\mathbb{R}^n} e^\theta d\mathbf{m} \geq (1-\lambda) \log \int_{\mathbb{R}^n} e^\psi d\mathbf{m} + \lambda \log \int_{\mathbb{R}^n} e^{\psi'} d\mathbf{m}.$$

Proof

◁ Let us consider $\tilde{\theta} := \theta - V + Q_T$, $\tilde{\psi} := \psi - V + Q_T$, and $\tilde{\psi}' := \psi' - V + Q_T$, where $Q_T = -\log(d\gamma_T/dx)$. Hypothesis (6.9) in Theorem 6.2.1 for $\tilde{\theta}, \tilde{\psi}, \tilde{\psi}'$, for any $T > 0$, is equivalent to,

$$\begin{aligned} \theta((1-\lambda)x + \lambda x') &\geq (1-\lambda)\psi(x) + \lambda\psi'(x') \\ &\quad + V((1-\lambda)x + \lambda x') - (1-\lambda)V(x) - \lambda V(x') \\ &\quad - Q_T((1-\lambda)x + \lambda x') + (1-\lambda)Q_T(x) + \lambda Q_T(x'). \end{aligned} \quad (6.17)$$

Since (6.16) is equivalent to

$$(1-\lambda)V(x) + \lambda V(x') - V((1-\lambda)x + \lambda x') \geq \kappa \frac{\lambda(1-\lambda)}{2} |x' - x|^2,$$

and by definition of Q_T ,

$$\begin{aligned} (1-\lambda)Q_T(x) + \lambda Q_T(x') - Q_T((1-\lambda)x + \lambda x') &= (1-\lambda)\frac{|x|^2}{2T} + \lambda\frac{|x'|^2}{2T} - \frac{|(1-\lambda)x + \lambda x'|^2}{2T} \\ &= \frac{\lambda(1-\lambda)}{2T} |x' - x|^2. \end{aligned}$$

Therefore from (6.17) it follows that,

$$\theta((1-\lambda)x + \lambda x') \geq (1-\lambda)\psi(x) + \lambda\psi'(x') - \left(\kappa - \frac{1}{T}\right) \frac{\lambda(1-\lambda)}{2} |x' - x|^2.$$

So if θ, ψ, ψ' satisfy (6.16), then $\tilde{\theta}, \tilde{\psi}, \tilde{\psi}'$ satisfy (6.9), and by Theorem 6.2.1,

$$\log \int_{\mathbb{R}^n} e^{\tilde{\theta}} d\gamma_T \geq (1-\lambda) \log \int_{\mathbb{R}^n} e^{\tilde{\psi}} d\gamma_T + \lambda \log \int_{\mathbb{R}^n} e^{\tilde{\psi}'} d\gamma_T.$$

Finally, by definition of $\tilde{\theta}, \tilde{\psi}, \tilde{\psi}'$ and the arbitrary of T , we conclude the proof. \triangleright

To conclude, we show that by an appropriate choice of the functions θ, ψ and ψ' , we can recover the geometric counterpart of the Prékopa-Leindler inequality, that is the Brunn-Minkowski inequality.

Corollary 6.2.4 *Let $A, B \subset \mathbb{R}^n$ be two bounded measurable sets, such that $(1-\lambda)A + \lambda B$ is also measurable, then*

$$|(1-\lambda)A + \lambda B| \geq |A|^{1-\lambda} |B|^\lambda \quad (6.18)$$

where $|\cdot|$ denotes the volume of the set.

Proof

\triangleleft It is enough to consider the functions θ, ψ, ψ' in Theorem 6.2.1 such that, respectively

$$e^\theta = \mathbb{1}_{((1-\lambda)A + \lambda B)}, \quad e^\psi = \mathbb{1}_A, \quad e^{\psi'} = \mathbb{1}_B.$$

And by the properties of the logarithmic function, (6.18) is immediately derived.

\triangleright

Physical interpretation

We stress here the statistical mechanics flavour of this proof, given by the Schrödinger problem. We consider a large number N of independent and non interacting particles. We associate to each particle a random variable X_t^i defined on \mathbb{R}^n , for all $t \geq 0$. The initial

conditions $R_0 = \delta_0$, means that at time $t = 0$, they are all concentrated in the origin; and for each $t > 0$, each of them, is distributed as a Normal Gaussian random variable $\mathcal{N}(0, t\text{Id})$. By the law of large numbers, we know that for each $t > 0$, the empirical measure converges as $N \rightarrow \infty$ to the law $\mathcal{N}(0, t\text{Id})$. We assume to observe the system at time T uniformly distributed (and normalized) on a set A . We get,

$$H(Q|R) = H(Q_T|R_T) = -\log \gamma_T(A).$$

According to (2.69) the measure of the set gives an estimation of the probability to end up to the uniform distribution of the same set: the largest the set A is (w.r.t to Gaussian measure) the most likely (but still unexpected!) is to observe the system uniformly distributed on A .

Conclusions and perspectives

In this manuscript we studied the Schrödinger entropy minimization problem and some of its applications.

We proved some similarities between the solutions of the Schrödinger problem and the Monge-Kantorovich problem, respectively the entropic and the displacement interpolations, and their minimal values, respectively the entropic cost and the quadratic Wasserstein distance. Among them, the most useful in the applications are the **Kantorovich-Schrödinger dual formulation** and the **Benamou-Brenier-Schrödinger formulation**.

On the other hand, we saw that the Schrödinger problem naturally involves Markov semigroups and the Γ_2 -operator, hence, it lend itself to be studied under suitable curvature-dimension conditions. As a consequence of these features, we derived some functional property and inequality. We proved that **convexity of the relative entropy is conserved** along entropic interpolations under $CD(0, n)$ both for finite and infinite dimension and a **regularity** result of the entropic cost along the heat flow.

In order to prove functional inequalities we followed two strategies: one is to give proofs through the genuine Schrödinger problem, that is $\varepsilon = 1$, without exploiting the convergence to optimal transport; the other is to give proofs via the Schrödinger problem with $\varepsilon > 0$, and then take advantage of the limit $\varepsilon \rightarrow 0$ to recover the standard results like for instance, the **contraction** and **EVI** inequalities. The first method is very performing, providing easy and fast proofs for instance for Prékopa-Leindler or **concentration** inequalities. For what concerns the second method, what we can observe is that in some case, the inequalities that we obtain for $\varepsilon > 0$ are penalized by some additional term, due to the diffusion phenomenon, see for instance (5.7). On the other hand the advantage of this method is to be able to perform rigorous computations thanks to the regularity of entropic interpolations, without an additional effort. This is the main gain of working with the Schrödinger entropy minimization problem, as an approximation of the Monge-Kantorovich problem.

Far from being exhaustive, our work leaves many open questions, that hopefully will be answered soon. Let us mention some of them.

- A short term question is to study the convexity of the entropy along entropic interpolations and the EVI inequality under the **$CD(\kappa, n)$ condition**, with finite $n > 1$ and non-zero κ .
- **Characterize curvature-dimension conditions via entropic properties.** In [Sv05] it is proved that $CD(\kappa, \infty)$ is equivalent to contraction in Wasserstein distance of the heat flow. We proved that $CD(\kappa, \infty)$ implies the contraction of the heat semigroup with respect to the entropic cost; can we prove the reverse implication? Of course, as

we have shown, contraction in entropic cost, implies contraction in Wasserstein distance, but could we have a direct equivalence between curvature-dimension conditions and the Schrödinger problem without recurring to the optimal transport setting? On the same line, the κ -EVI inequality is strictly related to displacement κ -convexity of the entropy functional. Does the κ -EVI inequality for the entropic cost give rise to an analogue notion of *entropic* κ -convexity?

- Another natural extension of our work is to replace the Brownian motion on \mathbb{R}^n with a random walk on a **discrete graph**. This approach was initiated in [Léo16] to provide a notion of displacement interpolation in the discrete setting. In particular the discrete analogue of our results might give an hint for a notion of κ -convexity and curvature-dimension conditions on a discrete setting.
- **Numerical applications.** The Schrödinger problem revealed itself to be very performing in numerical computations, [Cut13, BCC⁺15]. It is therefore used as a regular approximation to solve optimal transport related problems. For instance in [CDPS17] it is studied the discretization of the JKO scheme [JKO98] where the Wasserstein distance is replaced by its entropy approximation (the equivalent of our entropic cost). This approach is relatively few explored and might be extended to many other optimal transport based questions.

Bibliography

- [ABC⁺00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000.
- [ABLZ17] M. Arnaudon, A. Bela Cruzeiro, C. Léonard, and J.-C. Zambrini. An entropic interpolation problem for incompressible viscous fluids. *ArXiv e-prints*, April 2017.
- [AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, second edition, 2008.
- [Bak94] D. Bakry. L’hypercontractivité et son utilisation en théorie des semigroupes. In *Ecole d’Eté de Probabilités de Saint-Flour (1992)*, number 1581 in Lecture Notes in Mathematics. Springer-Verlag, 1994.
- [BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
- [BBGM12] D. Bakry, F. Bolley, I. Gentil, and P. Maheux. Weighted Nash inequalities. *Rev. Mat. Iberoam.*, 28(3):879–906, 2012.
- [BCC⁺15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.
- [BE85] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités*, number 1123 in Lecture Notes in Mathematics. Springer-Verlag, 1985.
- [Ber32] S. Bernstein. Sur les liaisons entre les grandeurs aléatoires. *Verhand. Internat. Math. Kongr. Zürich*, Band I, 1932.
- [Beu60] A. Beurling. An automorphism of product measures. *Ann. of Math.*, 72:189–200, 1960.
- [BGG16] F. Bolley, I. Gentil, and A. Guillin. Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp-Lieb inequalities. *To appear in Annals of probability*, 2016.
- [BGGK16] F. Bolley, I. Gentil, A. Guillin, and K. Kuwada. Equivalence between dimensional contractions in Wasserstein distance and the curvature-dimension condition. *To appear in Annali della Scuola Normale Superiore di Pisa*, 2016.
- [BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2014.
- [BGL15] D. Bakry, I. Gentil, and M. Ledoux. On Harnack inequalities and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 14(3):705–727, 2015.
- [BL00] S. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.
- [BL06] D. Bakry and M. Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.*, 22(2):683–702, 2006.
- [BLN94] J.M. Borwein, A.S. Lewis, and R.D. Nussbaum. Entropy minimization, DAD problems and doubly stochastic kernels. *J. Funct. Anal.*, 123:264–307, 1994.
- [Bor00] C. Borell. Diffusion equations and geometric inequalities. *Pot. Anal.*, 12:49–71, 2000.

- [Caf92] L. Caffarelli. The regularity of mappings with convex potentials. In *Partial differential equations and related subjects (Trento, 1990)*, volume 269 of *Pitman Res. Notes Math. Ser.*, pages 53–58. Longman Sci. Tech., Harlow, 1992.
- [CDPS17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.*, 49(2):1385–1418, 2017.
- [CEM17] D. Cordero-Erausquin and B. Maurey. Some extensions of the Prékopa-Leindler inequality using Borell’s stochastic approach. *Studia Math.*, 238(3):201–233, 2017.
- [CGP16] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.*, 169(2):671–691, 2016.
- [CGP17] Y. Chen, T. T. Georgiou, and M. Pavon. Optimal transport over a linear dynamical system. *IEEE Trans. Automat. Control*, 62(5):2137–2152, 2017.
- [Con17] G. Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *ArXiv e-prints*, April 2017.
- [Cos85] M. H. M. Costa. A new entropy power inequality. *IEEE Trans. Inform. Theory*, 31(6):751–760, 1985.
- [CT06] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.
- [Cut13] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc., 2013.
- [CWZ00] A.B. Cruzeiro, L. Wu, and J.-C. Zambrini. Bernstein processes associated with a Markov process. In R. Rebolledo, editor, *Stochastic analysis and mathematical physics, ANESTOC’98. Proceedings of the Third International Workshop*, Trends in Mathematics, pages 41–71, Boston, 2000. Birkhäuser.
- [CZ91] A.B. Cruzeiro and J.-C. Zambrini. Malliavin calculus and Euclidean quantum mechanics, I. *J. Funct. Anal.*, 96(1):62–95, 1991.
- [CZ08] K.L. Chung and J.-C. Zambrini. *Introduction to Random Time and Quantum Randomness*. World Scientific, 2008.
- [Doo53] J.L. Doob. *Stochastic Processes*. Wiley, 1953.
- [Doo57] J.L. Doob. Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France*, 85:431–458, 1957.
- [DS08] S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.
- [DZ98] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition*. Applications of Mathematics 38. Springer Verlag, 1998.
- [Ein05] A. Einstein. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Annalen der Physik*, 17:549–560, 1905.
- [EKS15] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.
- [Eva98] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Series in Mathematics*. American Mathematical Society, 1998.
- [Föll86] H. Föllmer. Time reversal on Wiener space. In *Stochastic Processes - Mathematic and Physics*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, 1986.
- [Föll88] H. Föllmer. *Random fields and diffusion processes, in École d’été de Probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics*. Springer, Berlin, 1988.

- [For40] R. Fortet. Résolution d'un système d'équations de M. Schrödinger. *J. Math. Pures Appl.*, 9:83, 1940.
- [Gen08] I. Gentil. From the Prékopa-Leindler inequality to modified logarithmic Sobolev inequality. *Ann. Fac. Sci. Toulouse Math. (6)*, 17(2):291–308, 2008.
- [GL10] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Processes and Related Fields*, 16:635–736, 2010.
- [GLR17] Ivan Gentil, Christian Léonard, and Luigia Ripani. About the analogy between optimal transport and minimal entropy. *Ann. Fac. Sci. Toulouse Math. (6)*, 26(3):569–601, 2017.
- [GS09] A. Galichon and B. Salanie. Matching with trade-offs: Revealed preferences over competing characteristics. *SSRN-1487307*, 2009.
- [GT17] N. Gigli and L. Tamanini. Second order differentiation formula on compact $RCD^*(K, N)$ spaces. *ArXiv e-prints*, January 2017.
- [IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., second edition, 1989.
- [Jam74] B. Jamison. Reciprocal processes. *Z. Wahrsch. verw. Geb.*, 30:65–86, 1974.
- [Jam75] B. Jamison. The Markov processes of Schrödinger. *Z. Wahrsch. verw. Geb.*, 32(4):323–331, 1975.
- [JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [Kan42a] L. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS*, 37:199–201, 1942.
- [Kan42b] L. Kantorovich. On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS*, 37:199–201, 1942.
- [Kuw15] K. Kuwada. Space-time Wasserstein controls and Bakry-Ledoux type gradient estimates. *Calc. Var. Partial Differential Equations*, 54(1):127–161, 2015.
- [Led01] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [Leh13] J. Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):885–899, 2013.
- [Leh14] J. Lehec. Short probabilistic proof of the Brascamp-Lieb and Barthe theorems. *Canad. Math. Bull.*, 57(3):585–597, 2014.
- [Leh15] J. Lehec. Borell's formula for a Riemannian manifold and applications. *ArXiv e-prints*, December 2015.
- [Léo01a] C. Léonard. Convex conjugates of integral functionals. *Acta Math. Hungar.*, 93(4):253–280, 2001.
- [Léo01b] C. Léonard. Minimization of energy functionals applied to some inverse problems. *J. Appl. Math. Optim.*, 44:273–297, 2001.
- [Léo01c] C. Léonard. Minimizers of energy functionals. *Acta Math. Hungar.*, 93(4):281–325, 2001.
- [Léo12a] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *J. Funct. Anal.*, 262(1879-1920), 2012.
- [Léo12b] C. Léonard. Girsanov theory under a finite entropy condition. In *Séminaire de probabilités de Strasbourg, vol. 44.*, pages 429–465. Lecture Notes in Mathematics 2046. Springer, 2012.
- [Léo14a] C. Léonard. Some properties of path measures. In *Séminaire de probabilités de Strasbourg, vol. 46.*, pages 207–230. Lecture Notes in Mathematics 2123. Springer., 2014.
- [Léo14b] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, 34(4):1533–1574, 2014.

- [Léo16] C. Léonard. Lazy random walks and optimal transport on graphs. *Ann. Probab.*, 44(3):1864–1915, 2016.
- [Léo17] C. Léonard. On the convexity of the entropy along entropic interpolations. In Nicola Gigli, editor, *Measure Theory in Non-Smooth Spaces*, Partial Differential Equations and Measure Theory. De Gruyter Open, 2017.
- [LRZ14] C. Léonard, S. Roelly, and J.-C. Zambrini. Reciprocal processes. A measure-theoretical point of view. *Probab. Surv.*, 11:237–269, 2014.
- [LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.
- [Mar86] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3):445–446, 1986.
- [McC97] R. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.
- [McC01] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.
- [Mik04] T. Mikami. Monge’s problem with a quadratic cost by the zero-noise limit of h -path processes. *Probab. Theory Relat. Fields*, 129:245–260, 2004.
- [Mon81] G. Monge. Mémoire sur la théorie des déblais et des remblais. In *Histoire de l’Académie Royale des Sciences de Paris*, pages 666–704, 1781.
- [MT06] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stoch. Proc. Appl.*, 116:1815–1835, 2006.
- [Nag64] M. Nagasawa. Time reversions of markov processes. *Nagoya Math. J.*, 24:177–204, 1964.
- [Nel67] E. Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, 1967.
- [Nel88] E. Nelson. *Stochastic mechanics and random fields*, in *Ecole d’Eté de Probabilités de Saint-Flour XV-XVII-1985-87*, volume 1362 of *Lecture Notes in Mathematics*. Springer, Berlin, 1988.
- [Ott01] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [OV00] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [OW05] F. Otto and M. Westdickenberg. Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Anal.*, 37(4):1227–1255, 2005.
- [Pré71] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32:301–316, 1971.
- [Rip17] L. Ripani. Convexity and regularity properties for entropic interpolations. *ArXiv e-prints*, November 2017.
- [Roc97] R.T. Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics. Princeton University Press, Princeton, N.J., 1997. First published in the Princeton Mathematical Series in 1970.
- [Roy99] G. Royer. *Une initiation aux inégalités de Sobolev logarithmiques*, volume 5 of *Cours spécialisés*. Société Mathématique de France, 1999.
- [RT93] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I -projections. *Statist. Probab. Lett.*, 17:369–375, 1993.
- [San15] F. Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
- [Sch31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

- [Sch32] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932.
- [Stu06a] K-T. Sturm. On the geometry of metric measure spaces, I. *Acta Math*, 196:65–131, 2006.
- [Stu06b] K-T. Sturm. On the geometry of metric measure spaces, I,II. *Acta Math.*, 196:65–177, 2006.
- [Sv05] K-T. Sturm and M-K. von Renesse. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.
- [SvR09] K-T. Sturm and M. von Renesse. Entropic measure and Wasserstein diffusion. *Ann. Probab.*, 37(3):1114–1191, 2009.
- [Tam17] L Tamanini. *Analysis and geometry of RCD spaces via the Schrödinger problem*. PhD thesis, UNIVERSITÉ PARIS NANTERRE, 2017.
- [Vil03] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.
- [Vil09] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.
- [Wan05] F.Y. Wang. *Functional inequalities, Markov Semigroup and Spectral Theory*. Chinese Sciences Press, Beijing, New York, 2005.
- [Zam86] J.-C. Zambrini. Variational processes and stochastic versions of mechanics. *J. Math. Phys.*, 27:2307–2330, 1986.

Abstract

In the past 20 years the optimal transport theory revealed to be an efficient tool to study the asymptotic behavior for diffusion equations, to prove functional inequalities, to extend geometrical properties in extremely general spaces like metric measure spaces, etc. The curvature-dimension of the Bakry-Émery theory appears as the cornerstone of those applications. Just think to the easier and most important case of the quadratic Wasserstein distance W_2 : contraction of the heat flow in W_2 characterizes uniform lower bounds for the Ricci curvature; the transport Talagrand inequality, comparing W_2 to the relative entropy is implied and implies via the HWI inequality the log-Sobolev inequality; McCann geodesics in the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ allow to prove important functional properties like convexity, and standard functional inequalities, such as isoperimetry, measure concentration properties, the Prékopa-Leindler inequality and so on. However the lack of regularity of optimal maps, requires non-smooth analysis arguments.

The Schrödinger problem is an entropy minimization problem with marginal constraints and a fixed reference process. From the Large deviation theory, when the reference process is driven by the Brownian motion, its minimal value \mathcal{A} converges to W_2 when the *temperature* goes to zero. The entropic interpolations, solutions of the Schrödinger problem, are characterized in terms of Markov semigroups, hence computation along them naturally involves Γ_2 -computations and the curvature-dimension condition. Dating back to the 1930s, and neglected for decades, the Schrödinger problem recently enjoys an increasing popularity in different fields, thanks to this relation to optimal transport, smoothness of solutions and other well performing properties in numerical computations.

The aim of this work is twofold. First we study some analogy between the Schrödinger problem and optimal transport providing new proofs of the dual Kantorovich and the dynamic Benamou-Brenier formulations for the entropic cost \mathcal{A} . Secondly, as an application of these connections we derive some functional properties and inequalities under curvature-dimensions conditions. In particular, we prove the concavity of the exponential entropy along entropic interpolations under the curvature-dimension condition $CD(0, n)$ and regularity of the entropic cost along the heat flow. We also give different proofs the Evolutionary Variational Inequality for \mathcal{A} and contraction of the heat flow in \mathcal{A} , recovering as a limit case the classical results in W_2 , under $CD(\kappa, \infty)$ and also in the flat dimensional case. Finally we propose an easy proof of the Gaussian concentration property via the Schrödinger problem as an alternative to classical arguments as the Marton argument which is based on optimal transport.

Key words: Schrödinger problem, Optimal Transport, Functional inequalities, curvature-dimension.

Le problème de Schrödinger et ses liens avec le transport optimal et les inégalités fonctionnelles

Résumé : Au cours des 20 dernières années, la théorie du transport optimal s'est révélée être un outil efficace pour étudier le comportement asymptotique dans le cas des équations de diffusion, pour prouver des inégalités fonctionnelles et pour étendre des propriétés géométriques dans des espaces extrêmement généraux comme des espaces métriques mesurés, etc. La condition de courbure-dimension de la théorie Bakry-Emery apparaît comme la pierre angulaire de ces applications. Il suffit de penser au cas le plus simple et le plus important de la distance quadratique de Wasserstein W_2 : la contraction du flux de chaleur en W_2 caractérise les bornes inférieures uniformes pour la courbure de Ricci ; l'inégalité de Talagrand du transport, comparant W_2 à l'entropie relative est impliquée et implique, par l'inégalité HWI, l'inégalité log-Sobolev ; les géodésiques de McCann dans l'espace de Wasserstein $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ permettent de prouver des propriétés fonctionnelles importantes comme la convexité, et des inégalités fonctionnelles standards telles que l'isopérimétrie, des propriétés de concentration de mesure, l'inégalité de Prékopa-Leindler et ainsi de suite. Néanmoins, le manque de régularité des plans minimisation nécessite des arguments d'analyse non lisse.

Le problème de Schrödinger est un problème de minimisation de l'entropie avec des contraintes marginales et un processus de référence fixes. À partir de la théorie des grandes déviations, lorsque le processus de référence est le mouvement Brownien, sa valeur minimale \mathcal{A} converge vers W_2 lorsque la température est nulle. Les interpolations entropiques, solutions du problème de Schrödinger, sont caractérisées en termes de semigroupes de Markov, ce qui implique naturellement les calculs Γ_2 et la condition de courbure-dimension. Datant des années 1930 et négligé pendant des décennies, le problème de Schrödinger connaît depuis ces dernières années une popularité croissante dans différents domaines, grâce à sa relation avec le transport optimal, à la régularité de ses solutions, et à d'autres propriétés performantes dans des calculs numériques.

Le but de ce travail est double. D'abord, nous étudions certaines analogies entre le problème de Schrödinger et le transport optimal fournissant de nouvelles preuves de la formulation duale de Kantorovich et de celle, dynamique, de Benamou-Brenier pour le coût entropique \mathcal{A} . Puis, en tant qu'application de ces connexions, nous dérivons certaines propriétés et inégalités fonctionnelles sous des conditions de courbure-dimension. En particulier, nous prouvons la concavité de l'entropie exponentielle le long des interpolations entropiques sous la condition de courbure-dimension $CD(0, n)$ et la régularité du coût entropique le long du flot de la chaleur. Nous donnons également différentes preuves de l'inégalité variationnelle évolutionnaire pour \mathcal{A} et de la contraction du flux de la chaleur en \mathcal{A} , en retrouvant comme cas limite, les résultats classiques en W_2 , sous $CD(\kappa, \infty)$ et $CD(0, n)$. Enfin, nous proposons une preuve simple de la propriété de concentration gaussienne via le problème de Schrödinger comme alternative aux arguments classiques tel que l'argument de Marton basé sur le transport optimal.

Mots-clés : Transport Optimal ; Problème de Schrödinger ; Entropie relative ; Bakry-Emery ; Inégalités fonctionnelles.

Image en couverture : Dessin réalisé lors de la formation doctorale "Comment présenter sa recherche à tout public".

