



HAL
open science

Problème de Schrödinger, inégalités fonctionnelles et transport optimal

Giovanni Conforti

► **To cite this version:**

Giovanni Conforti. Problème de Schrödinger, inégalités fonctionnelles et transport optimal. Probability [math.PR]. Institut Polytechnique de Paris, 2021. tel-04438901

HAL Id: tel-04438901

<https://hal.science/tel-04438901v1>

Submitted on 5 Feb 2024

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.



INSTITUT POLYTECHNIQUE DE PARIS

École doctorale de mathématiques Hadamard (ED 574)

CMAP–UMR 7641
École Polytechnique

Mémoire présenté pour l'obtention du

Diplôme d'Habilitation à Diriger les Recherches

Discipline : Mathématiques

par

Giovanni Conforti

Problème de Schrödinger, inégalités fonctionnelles
et transport optimal.

Rapporteurs : THIERRY BODINEAU
JAN MAAS
GIUSEPPE SAVARÉ

Date de soutenance : 15 Decembre 2021

Composition du jury : THIERRY BODINEAU
ANA BELA CRUZEIRO
FRANÇOIS DELARUE
NATHAEL GOZLAN
ARNAUD GUILLIN
JAN MAAS
GIUSEPPE SAVARÉ

Résumé

Ce mémoire est une synthèse de la majorité des mes travaux depuis la fin de ma thèse. Mon domaine principal de recherche est la théorie des probabilités et ses liens avec la géométrie Riemannienne et certaines branches de l'analyse, notamment le transport optimal. Toutes ces thématiques de recherche trouvent un terrain commun dans l'étude du problème de Schrödinger, qui est au cœur de mon activité de recherche de ces dernières années. Il y a deux thèmes récurrents dans ce manuscrit. Le premier consiste à montrer que les ponts de Schrödinger satisfont une équation du second ordre, c'est-à-dire une équation qui s'exprime à l'aide d'un terme d'accélération. Ceci m'a amené à explorer différents formalismes et à chercher à établir de nouveaux liens entre eux. Le second thème consiste à quantifier, à l'aide d'inégalités fonctionnelles et d'estimés de dissipation d'entropie, la vitesse de convergence à l'équilibre et le comportement ergodique des processus de Markov, avec un intérêt particulier pour ceux qui sont solutions optimales d'un problème de contrôle stochastique. Le mémoire est articulé en six chapitres et trois parties. La première partie, composée par les quatre premiers chapitres, est consacrée au problème de Schrödinger. La deuxième partie porte sur les inégalités convexes de Sobolev pour les chaînes de Markov et se compose d'un seul chapitre, exactement comme la troisième partie, où je traite le problème de la construction des courbes splines pour interpoler des mesures de probabilité.

Summary

This document contains an overview of a large part of my work after the completion of my PhD Thesis. I have conducted most of my research in probability theory and the common ground it shares with Riemannian geometry and some branches of analysis, optimal transport in particular. All these research fields interact naturally when looking at the family of Schrödinger problems, whose study has been at the heart of my scientific interests over the past few years. There are two central themes in this manuscript: the first one is that of showing that Schrödinger bridges solve a second order equation, i.e. an equation involving an acceleration term. Developing this theme requires to address the old question of how to properly define the acceleration of a stochastic processes and permits to enrich the already strong connections between optimal transport, in particular Otto calculus, and stochastic analysis, in particular Itô calculus. The second recurrent theme is that of quantifying by means of functional inequalities and entropy dissipation estimates the trend to equilibrium and the ergodic behavior of Markov processes, with a particular emphasis on optimally controlled diffusion processes and Markov chains on countable state spaces. I have divided this manuscript in three parts and six chapters. The first four chapters constitute the first part and are devoted to the Schrödinger problem. The second is about convex Sobolev inequalities for Markov chains and is made of one single chapter as the third part, that concentrates on the problem of defining spline curves to interpolate between probability measures.

List of publications

Accepted journal articles

- Giovanni Conforti, Anna Kazeykina, and Zhenjie Ren. Game on random environment, mean-field Langevin system and neural networks. *arXiv preprint arXiv:2004.02457*, forthcoming in *Mathematics of Operations Research*, 2020
- ◊ Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. Long-time behaviour of entropic interpolations. *forthcoming in Potential Analysis*, 2020
- ◊ Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. On the variational interpretation of local logarithmic Sobolev inequalities. *forthcoming in Annales de la faculté des Sciences de Toulouse*, 2020
- Giovanni Conforti and Christian Léonard. Time reversal of Markov processes with jumps under a finite entropy condition. *forthcoming in Stochastic processes and Applications*, 2021

- ◇ Giovanni Conforti. A probabilistic approach to convex (ϕ) -entropy decay for Markov chains. *arXiv preprint arXiv:2004.10850, forthcoming in Annals of Applied Probability*, 2020
- ◇ Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):964–1008, 2021
- ◇ Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, 178(1):475–530, 2020
- Giovanni Conforti, Tetiana Kosenkova, and Sylvie Roelly. Conditioned point processes with application to lévy bridges. *Journal of Theoretical Probability*, 32(4):2111–2134, 2019
- ◇ Giovanni Conforti and Luigia Ripani. Around the entropic Talagrand inequality. *Bernoulli*, 26(2):1431–1452, 2020
- ◇ Yongxin Chen, Giovanni Conforti, and Tryphon Georgiou. Measure-valued spline curves: An optimal transport viewpoint. *SIAM Journal on Mathematical Analysis*, 50(6):5947–5968, 2018
- ◇ Giovanni Conforti and Michele Pavon. Extremal flows in Wasserstein space. *Journal of Mathematical Physics*, 59(6):063502, 2018
- ◇ Giovanni Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, 174(1-2):1–47, 2019
- Giovanni Conforti and Max Von Renesse. Couplings, gradient estimates and logarithmic Sobolev inequality for Langevin bridges. *Probability Theory and Related Fields*, 172(1):493–524, 2018
- Giovanni Conforti. Fluctuations of bridges, reciprocal characteristics and concentration of measure. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 54(3):1432–1463, 2018
- Giovanni Conforti. Bridges of Markov counting processes: quantitative estimates. *Electronic Communications in Probability*, 21, 2016
- Giovanni Conforti and Christian Léonard. Reciprocal classes of random walks on graphs. *Stochastic Processes and their Applications*, 127(6):1870–1896, 2017
- Giovanni Conforti and Sylvie Roelly. Bridge mixtures of random walks on an Abelian group. *Bernoulli*, 23(3):1518–1537, 2017
- Giovanni Conforti, Christian Léonard, Rüdiger Murr, and Sylvie Roelly. Bridges of markov counting processes. reciprocal classes and duality formulas. *Electronic Communications in Probability*, 20, 2015
- Giovanni Conforti, Paolo Dai Pra, and Sylvie Roelly. Reciprocal class of jump processes. *Journal of Theoretical probability*, 30(2):551–580, 2017

Conference proceedings and book chapters

- Giovanni Conforti, Stefano De Marco, and Jean-Dominique Deuschel. On small-noise equations with degenerate limiting system arising from volatility models. In *Large Deviations and Asymptotic Methods in Finance*, pages 473–505. Springer, 2015
- Yongxin Chen, Giovanni Conforti, Tryphon Georgiou, and Luigia Ripani. Multi-marginal Schrödinger bridges. In *International Conference on Geometric Science of Information*, pages 725–732. Springer, 2019
- ◇ Alberto Chiarini, Giovanni Conforti, and Luca Tamanini. Schrödinger problem for lattice gases: A heuristic point of view. In *International Conference on Geometric Science of Information*, pages 891–899. Springer, 2021

Articles under review

- ◇ Julio Backhoff-Veraguas, Mathias Beiglböck, and Giovanni Conforti. A non-linear monotonicity principle and applications to Schrödinger type problems. *arXiv preprint arXiv:2101.09975, in revision at Bulletin of the London Mathematical Society*, 2021
- Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021
- ◇ Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Zhenjie Ren. Entropic turnpike estimates for the kinetic schrödinger problem. *arXiv preprint arXiv:2108.09161*, 2021

In preparation

- ◇ Giovanni Conforti, Richard Kraaij, and Daniela Tonon. Hamilton Jacobi equation for controlled gradient flows: the comparison principle. *In preparation*
- ◇ Giovanni Conforti and Laurent Pfeiffer. A coupling approach to the turnpike property in stochastic control. *In preparation*

Disclaimer Only the articles marked with ◇ are discussed in this thesis. The other ones are excluded from the presentation, either because they are closely connected to my PhD Thesis or because they treat topics that are quite far from the core of the manuscript. Moreover, at various places I presented simplified versions of the results in order to increase readability. Should any doubt arise, I refer the interested reader to the published versions of the corresponding articles.

Contents

I	Schrödinger problem	1
1	Schrödinger problem(s)	3
1.1	Schrödinger's thought experiment	3
1.1.1	Towards an abstract formulation	5
1.2	A gallery of examples	5
1.2.1	A first encounter with the toy model	8
1.2.2	Two questions	8
2	Constructing Schrödinger bridges	11
2.1	The fg -decomposition for the static problem	11
2.1.1	Optimal transport and cyclical monotonicity	12
2.1.2	The dual problems	12
2.1.3	A non linear monotonicity principle for Schrödinger type problems	13
2.1.4	Shape theorems	15
2.2	Newton's law for the entropic interpolation	17
2.2.1	A second encounter with the toy model	17
2.2.2	Otto calculus and the geometry of optimal transport	17
2.2.2.1	Heuristics	17
2.2.2.2	Some rigorous constructions	19
2.2.3	Newton's law	23
2.3	McKean Vlasov FBSDE for the mean field Schrödinger problem	25
2.3.1	Mean Field Schrödinger bridges	25
2.3.2	On the relation between Newton's laws and FBSDEs	27
2.4	Perspectives	28
2.4.1	Mean Field Schrödinger problem	28
2.4.2	Lattice gases	28
3	Hamilton Jacobi equations for controlled gradient flows	31
3.1	Introduction	31
3.1.1	A third encounter with the toy model	31
3.2	Gradient flows in EVI formulation and the Tataru distance	33
3.2.1	Evolutional Variational Inequality	34
3.2.2	The Tataru distance	34
3.2.3	Development of a proper lower and upper bound	34
3.3	Two comparison principles	36
3.3.1	More regular Hamiltonians	37
3.3.2	Some examples	39
3.4	Perspectives	40
4	Turnpike property and functional inequalities	41
4.1	Fourth encounter with the toy model	42
4.2	Convexity of the entropy along Schrödinger bridges	43
4.3	The total energy and long time behavior of the entropic cost	46
4.3.1	The conserved total energy and the energy-transport inequality	46
4.3.2	Long time behavior of the entropic cost	47
4.3.3	Short-time behavior of the entropic cost	47

4.4	Entropic turnpike estimates	48
4.4.1	Classical Schrödinger problem	48
4.4.1.1	Curvature dimension condition	48
4.4.1.2	Variational interpretation of local logarithmic Sobolev inequalities	50
4.4.2	Mean Field Schrödinger problem	51
4.4.2.1	Exponential convergence to equilibrium and the turnpike property	51
4.4.3	Kinetic Schrödinger problem	52
4.5	Turnpike estimates in Wasserstein distance for non uniformly convex potentials	53
4.6	Perspectives	56
II Markov chains		57
5	Convex entropy decay for Markov chains	59
5.1	Functional inequalities for discrete Markov chains	59
5.1.1	Trying to imitate Bakry-Émery	60
5.1.2	Discrete convex Sobolev inequalities	61
5.2	Coupling rates and convex entropy decay	63
5.2.1	Setup and main assumptions	63
5.2.2	Coupling rates and second derivative of the entropy	64
5.3	A sufficient condition for interacting random walks on \mathbb{N}^d	65
5.3.1	Equivalence with Wasserstein contraction	69
5.3.2	Examples	69
5.4	Hardcore models	71
5.5	Perspectives	72
5.5.1	Overcoming the limitations of convex entropy decay	72
5.5.2	Relations with entropic curvature and $CD(\kappa, +\infty)$	72
III Interpolation in the space of measures		75
6	Splines	77
6.1	Motivation and problem formulation	78
6.2	Problem formulation	78
6.2.1	Splines in \mathbb{R}^d	78
6.2.2	Splines in $\mathcal{P}_2(\mathbb{R}^d)$	79
6.3	The structure of measure-valued splines	79
6.3.1	Decomposition of optimal solutions	80
6.3.2	Formulation of the problem in phase space	80
6.3.3	Monge solutions in phase space	82
6.4	Fluid dynamical formulation of (6.2.3) and connections with Otto calculus	83
6.4.1	A fluid dynamic formulation for (6.2.3)	83
6.4.2	A Riemannian approach	83
6.5	Perspectives and numerical experiments	85
Frequently used notation		89
Bibliography		91

Part I

Schrödinger problem

Schrödinger problem(s)

Contents

1.1	Schrödinger's thought experiment	3
1.1.1	Towards an abstract formulation	5
1.2	A gallery of examples	5
1.2.1	A first encounter with the toy model	8
1.2.2	Two questions	8

1.1 Schrödinger's thought experiment

In a seminal work, E.Schrödinger asked the following question [198]:

*Imaginez que vous observez un **systeme de particules en diffusion**, qui soient en **équilibre thermodynamique**. Admettons qu'à un instant donné 0 vous les ayez trouvées en **répartition à peu près uniforme** et qu'à T vous ayez trouvé un **écart spontané et considérable par rapport à cette uniformité**. On vous demande de quelle manière cet écart s'est produit. Quelle en est la **manière la plus probable** ?*

A rigorous formulation of Schrödinger's question can be obtained arguing on the basis of Sanov's Theorem, as I now briefly explain. Consider a system of independent Brownian particles $(X_t^1, \dots, X_t^N)_{t \in [0, T]}$

$$\begin{cases} dX_t^i = dB_t^i, \\ X_0^i \sim \mathbf{m}, \quad i = 1, \dots, N, \end{cases}$$

where \mathbf{m} is some probability measure, and consider their empirical path distribution μ^N

$$\mu^N := \frac{1}{N} \sum_{i=1}^N \delta_{X^i},$$

that is a random element of $\mathcal{P}(\mathcal{C}([0, T]; \mathbb{R}^d))$ putting mass $1/N$ on the trajectory followed by each of the particles. Given a time interval $[0, T]$ and two probability measures μ and ν on \mathbb{R}^d representing the two observations in Schrödinger's thought experiment, (the *répartition à peu près uniforme* and the *écart spontané et considérable*) the natural way of interpreting the large deviations principle for $(\mu^N)_{N \geq 1}$ is by saying that for any possible evolution ρ we have, as $N \rightarrow +\infty$,

$$\text{Prob} \left[\mu^N \approx \rho \mid \mu_0^N \approx \mu, \mu_T^N \approx \nu \right] \asymp \exp(-N\mathcal{H}(\rho|R)). \tag{1.1.1}$$

In the above formula, μ_0^N and μ_T^N are the empirical marginal distributions, i.e. the marginals of μ^N at times $t = 0, T$, R is the law of the Brownian motion started at \mathbf{m} and $\mathcal{H}(\cdot|R)$ is the relative entropy functional

$$\mathcal{H}(\rho|R) = \begin{cases} \int \log \left(\frac{d\rho}{dR} \right) d\rho, & \text{if } \rho \ll R, \\ +\infty & \text{otherwise.} \end{cases}$$

The message of (1.1.1) is that relative entropy can be interpreted as a sort of negative log-likelihood in the many particles limit. Therefore, the problem of finding the most likely evolution conditionally on the observations $\{\mu_0^N \approx \mu, \mu_T^N \approx \nu\}$ is formulated as follows

$$\inf\{\mathcal{H}(P|R) : P \in \mathcal{P}(\Omega), (X_0)_\#P = \mu, (X_T)_\#P = \nu\}. \quad (1.1.2)$$

In the above, and in the rest of this manuscript, I shall denote by $(X_t)_{t \in [0, T]}$ the canonical process on $\Omega := \mathcal{C}([0, T]; \mathbb{R}^d)$ and by $\#$ the push forward of probability measures. The constrained entropy minimization problem (1.1.2) is the simplest and most classical instance of the *Schrödinger problem*. Besides (1.1.2), that I shall hereafter call the *large deviations formulation*, there exist three other equivalent formulations: a *static* formulation, a *stochastic control* formulation and a *fluid dynamic* formulation. Let me now spend some time introducing each of these different equivalent perspectives on Schrödinger's thought experiment.

Static formulation This formulation is obtained through the entropy decomposition formula, which reads as

$$\forall P \in \mathcal{P}(\Omega), \quad \mathcal{H}(P|R) = \mathcal{H}(P_{0T}|R_{0T}) + \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathcal{H}(P^{xy}|R^{xy}) P_{0T}(dxdy),$$

where for any $P \in \mathcal{P}(\Omega)$ we denote by P_{0T} the probability measure $(X_0, X_T)_\#P$ and P^{xy} stands for the xy bridge of P , that is the probability measure obtained by conditioning P to the event $\{X_0 = x, X_T = y\}$. Of course, the xy bridge is only P_{0T} -almost surely well defined. From this identity it is immediately seen that any optimizer in (1.1.2) is such that $P^{xy} = R^{xy} P_{0T}$ almost surely and therefore (1.1.2) is seen to be equivalent [126] to

$$\inf\{\mathcal{H}(\pi|R_{0T}) : \pi \in \Pi(\mu, \nu)\}, \quad (1.1.3)$$

where $\mathcal{H}(\cdot|R_{0T})$ denotes relative entropy w.r.t. to R_{0T} and $\Pi(\mu, \nu) \subseteq \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of couplings of μ and ν , that is to say the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ whose first marginal is μ and whose second marginal is ν . Problem (1.1.3) is known as the *static formulation* of the Schrödinger problem.

Stochastic control formulation This formulation is a consequence of the well known fact that for $P \in \mathcal{P}(\Omega)$, the condition $\mathcal{H}(P|R) < +\infty$ is equivalent [126, 162] to the existence a stochastic process $(\alpha_t^P)_{t \in [0, T]}$ that is adapted to the canonical filtration of Ω , with

$$\mathbb{E} \left[\int_0^T |\alpha_t^P|^2 dt \right] < +\infty$$

and such that under P

$$X_t - \int_0^t \alpha_s^P ds \quad \text{is a Brownian motion.} \quad (1.1.4)$$

Moreover, we have the following representation formula

$$\mathcal{H}(P|R) = \frac{1}{2} \mathbb{E} \left[\int_0^T |\alpha_t^P|^2 dt \right]$$

These considerations naturally lead to the following equivalent formulation of (1.1.2), that we shall call control formulation because of the fact that $(\alpha_t^P)_{t \in [0, T]}$ is naturally interpreted as a control variable.

$$\inf \left\{ \frac{1}{2} \mathbb{E} \left[\int_0^T |\alpha_t^P|^2 dt \right] : P \in \mathcal{P}(\Omega), (X_0)_\#P = \mu, (X_T)_\#P = \nu, (1.1.4) \text{ holds.} \right\}$$

Fluid dynamic formulation The fluid dynamic formulation is the one that one gets replacing empirical path measures with the *empirical* flow $(\mu_t^N)_{t \in [0, T]} \subset \mathcal{P}(\mathbb{R}^d)$ in Schrödinger's thought experiment. Optimizing the corresponding large deviations rate function, computed for example in [104], amounts to the following problem

$$\inf \left\{ \frac{1}{2} \int_0^T \int_{\mathbb{R}^d} |\alpha_t|^2 d\rho_t dt : (\rho, \alpha) \in \text{Adm}(\mu, \nu) \right\}. \quad (1.1.5)$$

In the above, a pair (ρ, α) belongs to the admissible set $\text{Adm}(\mu, \nu)$ if $(\alpha_t)_{t \in [0, T]} \in H_{-1}((\rho_t)_{t \in [0, T]})$, where $H_{-1}((\rho_t)_{t \in [0, T]})$ is the closure in $L^2((\rho_t)_{t \in [0, T]})$ of the set

$$\{\alpha : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \text{ s.t } \alpha = \nabla \psi, \psi \in C_c^\infty([0, T] \times \mathbb{R}^d)\},$$

and it provides with a weak solution of the Fokker-Planck equation

$$\begin{cases} \partial_t \rho_t - \frac{1}{2} \Delta \rho_t + \nabla \cdot (\alpha_t \rho_t) = 0, \\ \rho_0 = \mu, \rho_T = \nu. \end{cases} \quad (1.1.6)$$

The fluid-dynamic formulation of the Schrödinger problem has been recently established and revisited by various authors, see [139] for very general results. (1.1.5) admits an interpretation as the problem of controlling the Fokker Planck equation through its drift in such a way that minimum effort is made and the target configuration is reached at terminal time. Note that if we remove the term $-\frac{1}{2} \Delta$ from (1.1.6) then (1.1.5) is closely related to the Benamou-Brenier formulation of optimal transport [24].

1.1.1 Towards an abstract formulation

When translating Schrödinger's thought experiment into a rigorous mathematical problem, we made the choice of modeling the *système de particules en diffusion* through a collection of independent Brownian motions. This choice is convenient as it makes the calculation of the rate function simpler, but of course one is left to wonder with what happens when particles truly interact with each other. Moreover, from a physical standpoint, having independent particles is not so satisfactory. Following the heuristic reasons that led to (1.1.2), one quickly realizes that as soon as a large deviation principle for either the empirical flow or the empirical path measure is available, then the corresponding Schrödinger problem can be cast as the problem of minimizing the rate function among all flows or path measures having prescribed marginals at the initial and final time. Since the large deviation rate function is known explicitly for a rather large class of interacting particle systems, there is plenty of Schrödinger worth studying. A good part of my research activity has been devoted to the understanding of some of these problems, and in particular of the ones I am about to introduce.

1.2 A gallery of examples

In this section we illustrate some of the Schrödinger problems that will be discussed at a later stage in the thesis. Although a static formulation is to be expected only for the case of independent particles, in most cases, besides the original large deviations formulation, both a stochastic control and a fluid-dynamic formulation are available. For each of the following examples we shall only present one of these two alternative formulations. As a general convention, regardless of the type of Schrödinger problem under consideration, optimizers are called *Schrödinger bridges* when the problem is in stochastic control formulation. For problems in fluid dynamic formulation, optimizers are called entropic interpolations, although we may occasionally use a different terminology. The optimal value of a Schrödinger problem will generally be called *entropic cost*, or simply cost.

The classical Schrödinger problem (CSP) This problem correspond a general version of (1.1.2) in which particles are still independent but the single particle dynamics can incorporate a gradient drift term and the ambient space is a smooth, connected Riemannian manifold M without boundary. That is to say, we are looking at

$$\begin{cases} dX_t^i = -\nabla U(X_t^i) dt + dB_t^i, \\ X_0^i \sim \mathbf{m}, \quad i = 1, \dots, N. \end{cases} \quad (1.2.1)$$

where $\mathbf{m}(dx) = Z^{-1} \exp(-2U(x)) dx$ is the invariant measure¹. Arguing on the basis of Sanov's Theorem we obtain the large deviations formulation of the classical Schrödinger problem as

¹ \mathbf{m} may have infinite mass and therefore so does R . However, relative entropy w.r.t. R can still be defined [163].

$$\inf \{ \mathcal{H}(\mathbb{P}|\mathbb{R}) : \mathbb{P} \in \mathcal{P}(\mathcal{C}([0, T]; M)), (X_0)_\# \mathbb{P} = \mu, (X_T)_\# \mathbb{P} = \nu \}, \quad (\text{CSP})$$

where \mathbb{R} is the law of

$$\begin{cases} dX_t = -\nabla U(X_t)dt + dB_t, \\ X_0 \sim \mathbf{m}. \end{cases}$$

The case when the drift is not of gradient type is of clear interest, but won't be further discussed in the rest of the manuscript. The consequences of lower bounds on the Ricci curvature of M on the Schrödinger problem will be extensively investigated in the upcoming chapters.

The kinetic Schrödinger problem (KSP) For this problem, particles are still independent, but instead of following (1.2.1), they evolve according to Langevin dynamics

$$\begin{cases} dX_t^i = V_t^i dt, \\ dV_t^i = -\nabla U(X_t^i)dt - \gamma V_t^i dt + \sqrt{2\gamma} dB_t^i, \\ (X_0^i, V_0^i) \sim \mathbf{m}, \end{cases} \quad (1.2.2)$$

where $\gamma > 0$ is the friction parameter and $\mathbf{m}(dx dv) = Z^{-1} \exp(-2U(x) - |v|^2) dx dv$ the invariant measure. The PDE describing the evolution of the probability density of $(X_t^i, V_t^i)_{t \in [0, T]}$ is well known in the literature as the *kinetic Fokker-Planck equation*. Going back to Schrödinger's thought experiment and given that the V_t^i are velocity variables, it is natural in this context to condition only on the observation of the configuration of the spatial variables X_t^i at the initial and final times. There is no difficulty in adapting the heuristic justification of (1.1.2) through Sanov's theorem to this slightly different setting to obtain the corresponding Schrödinger problem, whose stochastic control formulation is as follows.

$$\inf \left\{ \mathcal{H}((X_0, V_0)_\# \mathbb{P} | \mathbf{m}) + \frac{1}{4\gamma} \mathbb{E} \left[\int_0^T |\alpha_t^{\mathbb{P}}|^2 dt \right] : \mathbb{P} \text{ admissible} \right\}, \quad (\text{KSP})$$

where a path probability measure $\mathbb{P} \in \mathcal{P}(\mathcal{C}([0, T]; \mathbb{R}^d \times \mathbb{R}^d))$ is admissible if and only if under \mathbb{P} , there exist a Brownian motion $(B_t)_{t \in [0, T]}$ and an adapted process $(\alpha_t^{\mathbb{P}})_{t \in [0, T]}$ such that $\mathbb{E}[\int_0^T |\alpha_t^{\mathbb{P}}|^2 dt] < +\infty$ and

$$\begin{cases} dX_t = V_t dt, \\ dV_t = -\nabla U(X_t)dt - \gamma V_t dt + \alpha_t^{\mathbb{P}} dt + \sqrt{2\gamma} dB_t, \\ X_0 \sim \mu, X_T \sim \nu \end{cases} \quad (1.2.3)$$

holds. (KSP) is called the kinetic Schrödinger problem: in its fluid dynamic formulation, it is the problem of steering the kinetic Fokker Planck equation towards the spatial configuration ν with minimum effort. Note that the presence of the extra term $\mathcal{H}((X_0, V_0)_\# \mathbb{P} | \mathbf{m})$ in the cost function is due to the fact that in (KSP) only the spatial variables are constrained and not the velocity variables. The study of the long time behavior of its solutions is, as expected, quite delicate: this is a manifestation of the hypocoercive nature of the kinetic Fokker Planck equation [66].

The mean field Schrödinger problem (MFSP) Here, the particle system is given by the N -particles approximation of the McKean-Vlasov dynamics. More precisely, for any N we consider $(X_t^i)_{1 \leq i \leq N, t \in [0, T]}$ defined by

$$\begin{cases} dX_t^i = b(\mu_t^N, X_t^i)dt + \sigma(\mu_t^N, X_t^i) \cdot dB_t^i, \\ X_0^i \sim \mu, \end{cases} \quad (1.2.4)$$

where the drift b and the diffusion coefficient σ satisfy suitable regularity assumptions and I recall that μ_t^N denotes the empirical distribution at time t , namely

$$\mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}.$$

A large deviations principle for the empirical flow $(\mu_t^N)_{t \in [0, T]}$ was established in [104], whereas results for the empirical path measure have been obtained in [42] and lead to

$$\inf \{ \mathcal{H}(P | \Gamma(P)) : P \in \mathcal{P}(\mathcal{C}([0, T]; \mathbb{R}^d)), (X_0)_{\#} P = \mu, (X_T)_{\#} P = \nu \}, \quad (\mathbf{MFSP})$$

In the above, for a given P , $\Gamma(P)$ is the law on $\mathcal{C}([0, T]; \mathbb{R}^d)$ of the controlled diffusion process

$$\begin{cases} dZ_t = b(P_t, Z_t)dt + \sigma(P_t, Z_t) \cdot dB_t, \\ Z_0 \sim \mu, \end{cases} \quad (1.2.5)$$

where I denote by $(P_t)_{t \in [0, T]}$ is the marginal flow of P . Note that the fixed point equation $\Gamma(P) = P$ defines the McKean-Vlasov dynamics for (1.2.4). The stochastic control formulation is in terms of a McKean-Vlasov control problem.

Schrödinger problem for the simple exclusion process Given a discretization $\mathbb{T}_N = \{i/N, i = 1, \dots, N\}$ of the unit circle, the simple symmetric exclusion process is a continuous time Markov chain on the configuration space $\{0, 1\}^{\mathbb{T}_N}$. We interpret $\eta(i) = 0$ as the fact that for the configuration η there is no particle at site i , whereas $\eta(i) = 1$ means that site i is occupied by a particle. The dynamics is governed by the exclusion rule, which I briefly describe. Attached to each site i there is an exponential clock of rate $1/2N^2$: when the clock rings, if only one of the sites $i, i+1$ is currently occupied, then the particle sitting in the occupied site jumps to the empty site. Therefore, the generator \mathcal{L}_N of the Markov chain is given for $f : \{0, 1\}^{\mathbb{T}_N} \rightarrow \mathbb{N}$ by

$$\mathcal{L}_N f(\eta) = \frac{1}{2N^2} \sum_{i=1}^N [f(\eta^{i, i+1}) - f(\eta)]$$

with

$$\eta^{i, i+1}(j) = \begin{cases} \eta(i+1), & \text{if } j = i \\ \eta(i), & \text{if } j = i+1 \\ \eta(j) & \text{otherwise.} \end{cases}$$

Given any path $(\eta_t)_{t \in [0, T]}$ the empirical density at time t is defined by

$$\mu_t^N : \mathbb{T} \rightarrow [0, 1], \quad \mu_t^N(\theta) = \sum_{i=1}^N \eta_t(i) \chi_{(i/N, i+1/N)}(\theta).$$

Note that, in contrast with the other examples considered so far, μ_t^N is a random function and not a random probability measure. A large deviations principle for the empirical density flow has been established in [154] and yields the corresponding Schrödinger problem in fluid dynamic form

$$\inf \left\{ \frac{1}{2} \int_0^T \int_{\mathbb{T}} |\alpha_t|^2 \rho_t (1 - \rho_t) d\theta dt : (\rho, \alpha) \in \text{Adm}(\mu, \nu) \right\}, \quad (\mathbf{SESP})$$

where given smooth density profiles μ, ν we say that $(\rho, \alpha) \in \text{Adm}(\mu, \nu)$ if and only if $\rho_0 = \mu, \rho_T = \nu, \alpha$ belongs to $H_{-1}((\rho_t(1 - \rho_t))_{t \in [0, T]})$, defined as the closure in $L^2((\rho_t(1 - \rho_t))_{t \in [0, T]})$ of

$$\{\nabla H : H \in C_c^\infty([0, T] \times \mathbb{T})\},$$

and if it provides with a solution of the non linear continuity equation

$$\partial_t \rho_t - \frac{1}{2} \Delta \rho_t + \nabla \cdot (\alpha_t \rho_t (1 - \rho_t)) = 0.$$

In all the above expressions, ∇ stands for the gradient taken w.r.t. the spatial (angle) variable θ . (**SESP**) is only a particular instance of a much more general problem, that we construct replacing the simple exclusion process with a

lattice gas, i.e. a collection of indistinguishable particles $(X_t^1, \dots, X_t^N)_{t \geq 0}$ performing random walks on a subset Λ of a lattice (typically \mathbb{Z}^d) and interacting through their empirical distribution: notable examples of lattice gases are the zero-range dynamics and the Kawasaki-Glauber dynamics. In this general framework the large deviations principle is provided by the so called *fundamental formula* of Macroscopic Fluctuations Theory [31] and is expressed in terms of two functions, called *mobility* and *diffusivity*. Although the Schrödinger problem for lattice gases is probably the one that best describes Schrödinger's thought experiment, at the moment of writing and to the best of my knowledge, its understanding is quite limited. In this manuscript, I shall give at various occasions some perspectives on conjectured results, their geometrical interpretation as well as put forward some heuristics that may at a later stage yield rigorous proofs of the most relevant claims.

Other Schrödinger-like problems The short list I made is far from being complete, and many interesting descendants of the original Schrödinger problem exist and are being currently investigated: to name one, the multiplicative Schrödinger problem studied by S.Pal and coauthors [186].

1.2.1 A first encounter with the toy model

All the model examples I have just introduced share a common interpretation: they may be seen as the problem of correcting a given reference dynamics, in such a way that a target configuration at time T is reached and minimum correction effort is made. The interaction mechanism of the underlying particle system dictates the equation satisfied by the reference dynamics and the way in which the control effort is measured. This geometric structure is captured in the following simple toy model on \mathbb{R}^d , that I will call toy Schrödinger problem in the sequel:

$$\inf \left\{ 2E(x) + \frac{1}{2} \int_0^T |u_t|^2 dt : \dot{x}_t^u = -\nabla E(x_t^u) + u_t, x_0^u = x, x_T^u = y \right\} := C_T(x, y). \quad (\text{TSP})$$

Adding the constant $2E(x)$ in the objective function, thought irrelevant for the minimization problem, has the advantage of making $C_T(x, y)$ symmetric in x, y . This property is particularly useful in the study of the time-reversal properties of Schrödinger bridges, as I will explain in the next chapters. In the toy model the reference dynamics is chosen to be $\dot{x}_t = -\nabla E(x_t)$, as this equation is the simplest possible *gradient flow* one can think of. In most of the examples above, the reference dynamics admits a variational interpretation as a gradient flow once the space of probability measures is equipped with the correct metric. For example, in a fundamental contribution Jordan, Kinderlehrer and Otto [151] showed that the Fokker-Planck equation is the gradient flow of the relative entropy in the space of probability measures endowed with the Wasserstein distance. Understanding the toy model is very useful in the analysis of the various Schrödinger problems we shall encounter throughout this manuscript. In some cases, by considering an adequate Riemannian structure on the space of probability measures, this correspondence becomes in some sense *exact*. Having a *geometric* structure instead of a *metric* structure provides with a richer toolbox: in particular, we can compute the acceleration of a curve, that is the covariant derivative of the velocity along the curve itself. I will show in Chapter 2 that computing the acceleration of a Schrödinger bridge is key to gain a deep geometric intuition on its dynamics that eventually leads to a fine control of its ergodic behavior. Leaving all precise definitions to the upcoming chapter, let me just mention that in the case of (**CSP**), the right (formal) Riemannian metric to consider is the so called Otto metric [182], whose associated geodesic distance is the Wasserstein distance of order two.

1.2.2 Two questions

I conclude this introduction to the Schrödinger problem posing two questions about Schrödinger bridges. A good part of this manuscript is devoted to illustrate what partial answers I could find, together with my collaborators.

- (i) What equations govern the dynamics of Schrödinger bridges? What is their interpretation? How to rigorously establish existence and uniqueness?

I shall give a more precise formulation of each of these questions at a later stage. For the moment, let me just say that answers to question (i) rely on mathematical objects that are quite different, depending on the type of formulation we

work with. In the sequel I shall report on existence results in the form of Forward Backward Stochastic Differential Equations (FBSDEs), coupled Hamilton Jacobi Bellman and Fokker Planck (HJB-FP) PDE systems and Hamilton Jacobi equations on metric spaces.

(ii) Where do Schrödinger bridges spend most of their time? What is their long time and ergodic behavior?

Finding good answers for (ii) is obviously very important for developing the theory of Schrödinger bridges, but not only. Indeed, in the effort of finding quantitative entropy dissipation estimates for entropic functionals along Schrödinger bridges, one discovers a novel class of functional inequalities involving the entropic cost, that includes both generalizations of well known inequalities such as Talagrand's transport-entropy inequality, and genuinely novel inequalities that do not have a classical counterpart. Moreover, answering (1.2.2) may be viewed as a first step towards a general quantitative understanding of the *turnpike property* in stochastic control. This property stipulates that optimal solutions of dynamic control problems are made of three pieces: a rapid transition from the initial state to the steady state, the *turnpike*, then a long stationary phase localized around the turnpike, and finally another rapid transition to reach the final state. Although this phenomenon has been widely studied in deterministic control [206] and despite the immense body of quantitative results on the speed of convergence to equilibrium for Markov processes, much less appears to be known concerning the ergodicity of *controlled* Markov processes and stochastic control in general, and even less so about mean field stochastic control. Many of the upcoming ergodic results are indeed quantitative versions of the turnpike property for Schrödinger bridges.

Constructing Schrödinger bridges

Contents

2.1	The fg-decomposition for the static problem	11
2.1.1	Optimal transport and cyclical monotonicity	12
2.1.2	The dual problems	12
2.1.3	A non linear monotonicity principle for Schrödinger type problems	13
2.1.4	Shape theorems	15
2.2	Newton's law for the entropic interpolation	17
2.2.1	A second encounter with the toy model	17
2.2.2	Otto calculus and the geometry of optimal transport	17
2.2.2.1	Heuristics	17
2.2.2.2	Some rigorous constructions	19
2.2.3	Newton's law	23
2.3	McKean Vlasov FBSDE for the mean field Schrödinger problem	25
2.3.1	Mean Field Schrödinger bridges	25
2.3.2	On the relation between Newton's laws and FBSDEs	27
2.4	Perspectives	28
2.4.1	Mean Field Schrödinger problem	28
2.4.2	Lattice gases	28

This chapter focuses on the general problem of finding an equation for Schrödinger bridges. There are several possible approaches:

- A constrained optimization approach, that I discuss at section 2.1, in which one tries to establish the so called fg -decomposition of optimal solutions to the static formulation.
- A Riemannian geometry viewpoint, where one interprets optimality conditions for the fluid dynamic formulation as Newton's laws in the Riemannian manifold of optimal transport. This is the content of section 2.2.
- A stochastic control viewpoint, where one works with forward-backward stochastic differential equations (FBSDEs), that are of mean field (or McKean-Vlasov) type as soon as the underlying particle system is not made of independent particles. I discuss this approach at section 2.3.

Finally, I conclude the chapter with an overview of possible research directions.

2.1 The fg -decomposition for the static problem

In this section, I present an approach based on the so called *monotonicity principle* to show existence of dual optimizers for static Schrödinger-type problem that I developed in [8] together with J.Backhoff and M.Beiglböck. As I will explain more clearly below, one of the interesting aspects of this approach is that is completely independent of convex duality and can successfully be used to obtain shape theorems for the minimization of non-convex energy functionals, thus going beyond the classical Schrödinger problem. In order to better introduce the monotonicity principle and the dual Schrödinger problem, I begin with a short recapitulation of some fundamental facts on the Monge-Kantorovich optimal transport problem.

2.1.1 Optimal transport and cyclical monotonicity

Given marginals μ and ν on \mathbb{R}^d with finite second moment and a continuous cost function $c(\cdot, \cdot)$ the celebrated Monge-Kantorovich problem is the problem of minimizing an average cost $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ among all couplings of μ and ν

$$\inf \left\{ \int c \, d\pi : \pi \in \Pi(\mu, \nu) \right\}. \quad (2.1.1)$$

The notion of *c-cyclical monotonicity* leads to a geometric and variational characterization of optimal couplings. Its relevance for (2.1.1) has been highlighted by Gangbo and McCann [129], following earlier work of Knott and Smith [155] and Rüschendorf [194] among others.

Definition 2.1. A set $\Gamma \subseteq \mathbb{R}^d \times \mathbb{R}^d$ is *c-cyclically monotone* if any measure α that is finite and supported on finitely many points in Γ , is a cost-minimizing transport between its marginals. I.e., if α' has the same marginals as α , then

$$\int c \, d\alpha \leq \int c \, d\alpha'. \quad (2.1.2)$$

A transport plan π is called *c-cyclically monotone* if it is concentrated on such a set Γ , i.e. if there is such a π with $\pi[\Gamma] = 1$.

The definition I gave is a slightly non-standard one that is not inherently tied to the transport problem and serves the exposition more directly. I refer to [209, Exercise 2.21] for equivalence to the more familiar way of stating *c-cyclical monotonicity* of a set Γ : usually one requires that for any $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$, $y_{n+1} = y_1$ it holds

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

The equivalence of optimality and *c-cyclical monotonicity* has been established under progressively weaker regularity assumptions. Based on [6, 191, 197, 20, 33] the following "*monotonicity principle*" holds true:

Theorem 2.1.1. Let $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, \infty)$ be measurable and assume that $\pi \in \Pi(\mu, \nu)$ is a transport plan with finite cost $\int c \, d\pi \in \mathbb{R}_+$. Then π is optimal if and only if π is *c-cyclically monotone*.

The importance of this result stems from the observation that it is often an elementary and feasible task to see whether a transport behaves optimally on a finite number of points. But this would be a priori of no help for a problem where single points do not carry positive mass. Theorem 2.1.1 provides the required remedy to this obstacle as it establishes the connection to optimality on a "pointwise" level.

2.1.2 The dual problems

Dual Monge Kantorovich problem The Monge-Kantorovich problem may be viewed as an infinite dimensional linear program. As in the finite dimensional case, a dual problem can be constructed invoking convex duality and reads as

$$\sup \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \varphi, \psi \in \mathcal{C}_b(\mathbb{R}^d), \varphi + \psi \leq c \right\}. \quad (2.1.3)$$

Existence of optimal solutions for the primal problem is easily obtained leveraging weak lowersemicontinuity of the cost function and weak compactness of $\Pi(\mu, \nu)$. On the contrary, existence of optimal dual solutions is more delicate [4, Thm. 1.17]. The relation between primal and dual optimal solution is given in the following *shape theorem*, which can be deduced from [4, Thm 1.13].

Theorem 2.1.2. Assume that μ, ν have a second moment and let $c(x, y) = |x - y|^2$. Then $\pi \in \Pi(\mu, \nu)$ is optimal for (2.1.1) if and only if there exist a concave function φ such that $\max\{\varphi, 0\} \in L^1(\mu)$ and the support of π is contained in the superdifferential of φ . In this case, (φ, ψ) is an optimal solution of (2.1.3), where

$$\psi(y) = \inf_{x \in \mathbb{R}^d} c(x, y) - \varphi(x).$$

Dual Schrödinger problem Let me now come back to the static Schrödinger problem that, for a general probability measure ρ on $\mathbb{R}^d \times \mathbb{R}^d$ is given by

$$\inf\{\mathcal{H}(\pi|\rho) : \pi \in \Pi(\mu, \nu)\}.$$

Relying again on convex duality arguments, one obtains the dual problem

$$\sup\left\{\int \varphi d\mu + \int \psi d\nu - \log \int \exp(\psi + \varphi) d\rho : \varphi, \psi \in \mathcal{C}_b(\mathbb{R}^d)\right\}.$$

Remark that so far I have not made any specific assumption on ρ but even if I had done so, there is in general no reason why dual attainment should take place in a space as regular as $C_b(\mathbb{R}^d)$. At the formal level one expects the unique primal and dual optimizers to be related through the functional equation

$$\frac{d\pi}{d\rho}(x, y) = \exp(\varphi(x) + \psi(y)) \quad (2.1.4)$$

Equivalently, defining $f = \exp(\varphi)$, $g = \exp(\psi)$ we can rewrite the above as

$$\frac{d\pi}{d\rho}(x, y) = f(x)g(y) \quad (2.1.5)$$

In the sequel, I will refer to (2.1.5) as to the fg -decomposition of optimal solutions. Moreover, imposing that a plan of the form (2.1.5) is admissible, we find the so called *Schrödinger system*

$$\begin{cases} f(x) \int g(y) \rho^x(dy) = \frac{d\mu}{d\rho_0}(x), & \rho_0 - \text{a.s.} \\ g(y) \int f(x) \rho^y(dx) = \frac{d\nu}{d\rho_1}(y), & \rho_1 - \text{a.s.} \end{cases}$$

where $\rho^x(dy)$ is the conditional distribution of ρ defined formally by

$$\rho^x[A] = \frac{\rho[\{x\} \times A]}{\rho[\{x\} \times \mathbb{R}^d]}$$

and $\rho^y(dx)$ is defined analogously. Likewise, ρ_0 (resp. ρ_1) stands for the first (resp. second) marginal of ρ . Turning the above formal statements into a rigorous shape theorem is a non trivial mathematical problem, that has a quite long history beginning with the early contributions of Fortet and Beurling [128, 32], followed by the seminal work of Csiszar [101]. In more recent works, Borwein and Lewis [39] and Borwein, Lewis and Nussbaum [40] proposed an approach to entropy minimization that combines fixed point-arguments and convex optimization techniques. Convex duality is also at the heart of the proof strategy of Pennanen and Perkkiö [189]. A different viewpoint is adopted by Rüschendorf and Thomsen [195]: therein the shape of the optimal measure is found as a consequence of the closedness property of sum spaces of integrable functions. A recent article of Carlier and Laborde [55] obtains the fg -decomposition for multimarginal generalizations of the Schrödinger problem. A large part of the above mentioned results is surveyed by Léonard in [163]. This author has also proven shape theorems for the Schrödinger problem in [160, 161].

2.1.3 A non linear monotonicity principle for Schrödinger type problems

Together with Julio and Mathias, we proved a monotonicity principle for a general version of the Schrödinger problem, see (hSP) below, and applied it to derive a rigorous version of (2.1.4). As I said above, our approach has the nice feature of making no use of convex duality. As a consequence, it also applies to the minimization of energy functionals that, unlike the relative entropy, are not convex. To illustrate the potential of our approach, we applied our results to obtain a shape theorem for the optimal solutions of a non-convex Schrödinger problem with congestion. Moreover, all our results do not need to assume the base space to be \mathbb{R}^d , a Polish space is enough.

A ‘formal’ non-linear monotonicity principle In order to build intuition, I begin by stating a non-linear monotonicity principle for an abstract infinite dimensional optimization problem which is "formal" in the sense that I do not give a rigorous proof or precise conditions under which it is expected to hold. I consider a Polish space \mathcal{Z} with \mathcal{B} its Borel sigma-algebra and a family \mathcal{F} of real-valued functions on \mathcal{Z} . Furthermore, I assume that \mathcal{F} is a subset of $\mathcal{C}_b(\mathcal{Z})$, the space of continuous bounded functions. Given a functional

$$G : \mathcal{P}(\mathcal{Z}) \rightarrow [0, +\infty],$$

we are interested in the following problem

$$\inf \left\{ G(\pi) : \pi \in \mathcal{P}(\mathcal{Z}), \int_{\mathcal{Z}} f d\pi = 0, \forall f \in \mathcal{F} \right\}. \quad (2.1.6)$$

The standing assumption on G is that there exist directional derivatives with representation via functions, i.e. for any π in the domain there exists $\delta G_{\pi} : \mathcal{Z} \rightarrow (-\infty, \infty]$ measurable such that

$$\forall \bar{\pi}, \quad \lim_{\varepsilon \rightarrow 0} \frac{G(\pi + \varepsilon(\bar{\pi} - \pi)) - G(\pi)}{\varepsilon} = \int_{\mathcal{Z}} \delta G_{\pi} d(\bar{\pi} - \pi).$$

Positive finite measures α, α' with equal mass and finite support shall be called *competitors* if

$$\int_{\mathcal{Z}} f d(\alpha - \alpha') = 0, \forall f \in \mathcal{F}.$$

We then expect the following:

Lemma 2.1.1 (Non-Linear Monotonicity Principle, formal version). *Suppose π^* is an optimizer for problem (2.1.6). Then*

- π^* attains the linearized problem

$$\inf \left\{ \int_{\mathcal{Z}} c d\pi : \pi \in \mathcal{P}(\mathcal{Z}), \int f d\pi = 0, \forall f \in \mathcal{F} \right\}, \text{ where } c := \delta G_{\pi^*},$$

- There exist a Borel set $\Gamma_{\pi^*} \subset \mathcal{Z}$ such that $\pi^*[\Gamma_{\pi^*}] = 1$ having the following property: given competitors $\alpha, \bar{\alpha}$, with $\text{supp}(\alpha) \subseteq \Gamma_{\pi^*}$ we have

$$\int \delta G_{\pi^*} d\alpha \leq \int \delta G_{\pi^*} d\bar{\alpha}.$$

A Rigorous Non-Linear Monotonicity Principle In the rest of this section, fix a continuous function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying at least:

$$h \text{ is differentiable on } (0, \infty) \text{ and the limit } h'(0) := \lim_{x \rightarrow 0} h'(x) \text{ exists.} \quad (2.1.7)$$

Let me now specify the setting of the former paragraph to formulate the class of static Schrödinger type problems that are the object of this section. I do so by choosing

$$\mathcal{Z} := \mathcal{X} \times \mathcal{Y},$$

where \mathcal{X}, \mathcal{Y} are Polish spaces. As for the constraints set \mathcal{F} , we are interested in

$$\mathcal{F}_{\mu, \nu} := \mathcal{F}_{\mu} \cup \mathcal{F}_{\nu}, \quad \mathcal{F}_{\mu} := \left\{ f - \int_{\mathcal{X}} f d\mu : f \in C_b(\mathcal{X}) \right\} \quad \mathcal{F}_{\nu} := \left\{ g - \int_{\mathcal{Y}} g d\nu : g \in C_b(\mathcal{Y}) \right\},$$

for given probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ such that $\mu \otimes \nu \sim \rho$. With these specifications, our minimization problem clearly becomes:

$$\inf \left\{ \int_{\mathcal{X} \times \mathcal{Y}} h \left(\frac{d\pi}{d\rho}(x, y) \right) \rho(dx, dy) : \pi \in \Pi(\mu, \nu) \right\}. \quad (\mathbf{hSP})$$

Notice that for the choice $h(x) = x \log(x)$, problem **(hSP)** is indeed the classical Schrödinger problem. We established the following

Lemma 2.1.2 (Non-Linear Monotonicity Principle). *Suppose that h is twice differentiable on \mathbb{R}_+ with $h'' \geq C$ everywhere for some $C \in \mathbb{R}$ and that $\lim_{x \rightarrow +\infty} h'(x) = +\infty$. Furthermore, assume that either h' is lower bounded or $\lim_{x \downarrow 0} h'(x) = -\infty$ and let π^* be an optimizer of problem **(hSP)**. Then there exist sets $\Gamma_{\pi^*}, \Gamma_\rho$ such that $\rho[\Gamma_\rho] = \pi^*[\Gamma_{\pi^*}] = 1$ and for all competitors α, α' with $\text{supp}(\alpha) \subseteq \Gamma_{\pi^*}, \text{supp}(\alpha') \subseteq \Gamma_\rho$ we have*

$$\int h' \left(\frac{d\pi^*}{d\rho} \right) d\alpha \leq \int h' \left(\frac{d\pi^*}{d\rho} \right) d\alpha'. \quad (2.1.8)$$

A more explicit version of Lemma 2.1.2 in the case of the classical Schrödinger Problem has been obtained in parallel by Bernton, Ghosal and Nutz in [29], where it is furthermore leveraged to obtain large deviations estimates. Over the past few years, variants of the classical monotonicity principle of Definition 2.1 have been applied in transport problems for finitely or infinitely many marginals [187, 78, 18], the martingale version of the optimal transport problem [21, 180, 23], stochastic portfolio theory [185], the Skorokhod embedding problem [147], the distribution constrained optimal stopping problem [19, 22] and the weak transport problem [140, 9, 10]. What all these articles have in common is that the original idea is applied to other infinite dimensional *linear* optimization problems. Our contribution promoted the idea that this optimality principle can be useful beyond linear problems and in fact to problems that are not susceptible to a convex duality approach. Given the versatile applicability of the idea in various linear optimization problems, the extension to other non-linear problems appears highly promising. To the best of our knowledge, even in the restricted framework of **(hSP)**, the case when h is not convex had not been treated before.

2.1.4 Shape theorems

We now rigorously derive necessary optimality conditions for Problem **(hSP)**. In the case of the classical Schrödinger problem $h(x) = x \log(x)$, the functions φ, ψ appearing in the next theorems are known as *Schrödinger potentials*, see [163, Sec 2.]. When h is not convex, φ and ψ are not dual optimizers in general. However, they still admit the interpretation of Lagrange multipliers.

Theorem 2.1.3. *Assume that Problem **(hSP)** is finite and π^* is an optimizer thereof. Importantly, assume that $\rho \sim \mu \otimes \nu$. Let $h : [0, \infty) \rightarrow (-\infty, \infty)$ be twice continuously differentiable, $\lim_{x \rightarrow 0} h'(x) = -\infty$, $\lim_{x \rightarrow +\infty} h'(x) = +\infty$ and $\inf_{\mathbb{R}_+} h'' > -\infty$. Then $\pi^* \sim \rho$ and there exist measurable functions $\varphi : \mathcal{X} \rightarrow [-\infty, +\infty)$ and $\psi : \mathcal{Y} \rightarrow [-\infty, +\infty)$ such that*

$$h' \left(\frac{d\pi^*}{d\rho}(x, y) \right) = \varphi(x) + \psi(y), \quad \rho - a.s. \quad (2.1.9)$$

As I have said before Theorem 2.1.3, is well known for a when h is convex. However, no result for non convex h was known earlier, to the best of our knowledge, and none of the existing proofs in the convex case relies on the monotonicity principle. The above theorem applies to $h(x) = x \log(x)$ (where $h'(x) = 1 + \log(x)$) but not to $h(x) = x^2$ (where $h'(x) = 2x$). This latter case and similar ones are covered by the following complementary theorem.

Theorem 2.1.4. *Assume that Problem **(hSP)** is finite and π^* is an optimizer thereof. Assume that $\rho \sim \mu \otimes \nu$. Let $h : [0, \infty) \rightarrow (-\infty, \infty)$ be strictly increasing, continuously differentiable, $\lim_{x \rightarrow 0} h'(x) = 0$, $\lim_{x \rightarrow +\infty} h'(x) = +\infty$, $\inf_{\mathbb{R}_+} h'' > -\infty$. Then there exist measurable functions $\varphi : \mathcal{X} \rightarrow [-\infty, +\infty)$ and $\psi : \mathcal{Y} \rightarrow [-\infty, +\infty)$ such that*

$$h' \left(\frac{d\pi^*}{d\rho}(x, y) \right) = (\varphi(x) + \psi(y))_+, \quad \rho - a.s. \quad (2.1.10)$$

Remark that uniqueness of an optimizer to Problem **(hSP)** is guaranteed if h is strictly convex. On the other hand, Conditions (2.1.9)-(2.1.10) do not characterize optimizers even when these are unique (e.g. when h' is not one-to-one).

We now study the converse direction: how structure of a measure implies optimality.

Theorem 2.1.5. *Let $h : [0, \infty) \rightarrow (-\infty, \infty)$ be strictly convex, lower-bounded, and continuously differentiable, $\lim_{x \rightarrow 0} h'(x) = 0$, $\lim_{x \rightarrow +\infty} h'(x) = +\infty$, and $h(2x) \leq ah(x) + bx + c$ for constants a, b, c . Suppose that $\pi^* \in \Pi(\mu, \nu)$ is absolutely continuous with respect to ρ , with*

$$h' \left(\frac{d\pi^*}{d\rho}(x, y) \right) = (\varphi(x) + \psi(y))_+, \quad \rho - a.s.$$

for measurable $\varphi : \mathcal{X} \rightarrow [-\infty, +\infty)$ and $\psi : \mathcal{Y} \rightarrow [-\infty, +\infty)$. Then π^* is optimal for **(hSP)**.

With the same techniques used to prove Theorem 2.1.5, variants of this result can be established if $h'(0) \in [-\infty, \infty)$, covering in particular the Schrödinger problem. As for Theorem 2.1.4, it can be adapted to cover the case $h'(0) \in (-\infty, \infty)$.

Example: Schrödinger Problem with Congestion Imagine that $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ denote respectively origins and destinations for car users in a city. Hence an origin-destination pair (x, y) can stand for the route that a car has to travel from x to y . Experts have determined that $\rho \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is the optimal use of the road network (here $\rho(dx, dy)$ is the infinitesimal proportion of cars taking route (x, y)) in the stationary case. However, the actual proportion of car trips origins and car trip destinations are described by $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ respectively, rather than the marginals of ρ . In the vanilla version of the Schrödinger Problem we aim to determine a minimizer π^* of the relative entropy $\int \frac{d\pi}{d\rho} \log \left(\frac{d\pi}{d\rho} \right) d\rho$ over $\pi \in \Pi(\mu, \nu)$, $\pi \ll \rho$, amounting to the distribution of car trips compatible with the experts' guess ρ and the marginal information μ and ν . However, we may also want to consider congestion effects, codified by an added term $f \left(\frac{d\pi}{d\rho} \right)$ with $f(\cdot)$ increasing, the idea being that adding traffic above the experts' recommendation should be more costly than the opposite. This way we arrive at the non-convex Schrödinger-type problem of minimizing $\int \left[\frac{d\pi}{d\rho} \log \left(\frac{d\pi}{d\rho} \right) + f \left(\frac{d\pi}{d\rho} \right) \right] d\rho$ under the same constraints. The optimality condition in Theorem 2.1.3 now reads:

$$(\log + f') \left(\frac{d\pi^*}{d\rho} \right) = \varphi(x) + \psi(y),$$

from which π^* can even be determined depending on the choice of f .

Some perspectives on the mean field Schrödinger problem A simplified discrete-time version of (MFSP) consists in finding the most likely evolution conditionally to observations at initial and terminal times of the particle system $(X_t^i)_{i=1, \dots, N; t=0,1,2}$ where (X_0^1, \dots, X_0^N) are i.i.d. samples from a probability measure μ on \mathbb{R}^d and

$$X_{t+1}^i - X_t^i = -\frac{1}{N} \sum_{j \leq N} \nabla W(X_t^i - X_t^j) + \xi_t^i, \quad i = 1, \dots, N, \quad t = 0, 1. \quad (2.1.11)$$

Here the random variables $(\xi_t^i)_{i=1, \dots, N; t=0,1}$ are i.i.d. standard Gaussians. The large deviations rate function for the empirical distribution of the particle system (2.1.11) in the regime $N \rightarrow +\infty$ is known explicitly (see [124] for the analysis of the toy model (2.1.11)) and leads to the following problem formulation

$$\inf \left\{ \int h \left(\frac{d\pi}{dR(\pi)}(x_0, x_1, x_2) \right) R(\pi)(dx_0 dx_1 dx_2) : \pi \in \Pi(\mu, \nu) \right\}. \quad (2.1.12)$$

In the above $h(x) = x \log x$ and, adapting the usual notation, I denoted by $\Pi(\mu, \nu)$ the subset of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ whose first marginal ($t = 0$) is μ and whose last ($t = 2$) marginal is ν . Finally, for a given π , $R(\pi) \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ is defined as the law of the controlled discrete stochastic differential equation

$$\begin{cases} Z_{t+1} = Z_t - \int \nabla W(Z_t - x_t) \pi(dx_0, dx_1, dx_2) + \xi_t, & t = 0, 1, \\ Z_0 \sim \mu, \end{cases} \quad (2.1.13)$$

where (ξ_0, ξ_1) are i.i.d. standard Gaussians. Despite several analogies with (hSP), including the fact that the function $R(\cdot)$ naturally introduces non-convexity into the problem, the analysis of (2.1.12) is outside the reach of our work, essentially because the “reference” measure $R(\pi)$ depends on π . However, the heuristics put forward in the introduction based on the linearization procedure still apply and leads to natural conjectures on the kind of monotonicity principle and shape theorem for optimizers to be expected in this situation. For this reason, our article may be seen as a first step in the direction of developing and exploiting ever more powerful monotonicity principles. One of the main motivations for validating such conjectures for Problem (2.1.12) lies in the fact that a shape theorem for the mean field Schrödinger problems yields existence of solutions for the coupled Fokker Planck-Hamilton Jacobi Bellman system describing the dynamics of mean field Schrödinger bridges, see (2.4.2) below.

2.2 Newton's law for the entropic interpolation

In order to develop a good intuition on why Schrödinger bridges are expected to be solutions to a Newton's law, i.e. an equation of the form $\ddot{X}_t = \nabla I(X_t)$, it is instructive to revisit the toy model.

2.2.1 A second encounter with the toy model

Operating the change of variables $v_t = -\nabla E(x_t^u) + u_t$ it is not hard to see that **(TSP)** is equivalent to

$$\inf \left\{ \int_0^T \frac{1}{2} |v_t^2| + \frac{1}{2} |\nabla E(x_t^v)|^2 dt : \dot{x}_t^v = v_t, x_0^v = x, x_T^v = y \right\}. \quad (2.2.1)$$

A standard variational analysis reveals that optimal curves satisfy Newton's law

$$\ddot{x}_t^v = \frac{1}{2} \nabla (|\nabla E|^2)(x_t^v), \quad x_0^v = x, x_T^v = y, \quad \text{with } I(\cdot) = |\nabla E|^2(\cdot). \quad (2.2.2)$$

Recalling the interpretation of the toy model as the problem of finding the best approximation of a gradient flow under marginal constraints, equation (2.2.2) gains a natural interpretation: it tells that optimizers, who would like but cannot be gradient flows, *accelerate* as gradient flows do. Indeed, if $\dot{y}_t = -\nabla E(y_t)$ is a gradient flow, then computing its acceleration we obtain

$$\begin{aligned} \ddot{y}_t &= -\nabla^2 E(y_t) \cdot \dot{y}_t \\ &= \nabla^2 E(y_t) \cdot \nabla E(y_t) \\ &= \frac{1}{2} \nabla (|\nabla E|^2)(y_t). \end{aligned}$$

The goal of the next sections is to answer the question: does an equation like (2.2.2) holds for the entropic interpolation? In order to provide the answer, let me take some time to introduce the Riemannian formalism on the space of probability measures known as *Otto calculus*.

2.2.2 Otto calculus and the geometry of optimal transport

Aim of this section is to describe the formal Riemannian metric on $\mathcal{P}_2(M)$ known as Otto metric. In preparation for the results of the next section, I will compute explicitly the Levi Civita connection as well as the gradient and Hessian of the relative entropy. Most of the following presentation is inspired from [4]. I stress that, although it is not possible to have a fully rigorous definition of the Otto metric, some rigorous constructions are possible [138], and constitute the framework I will use to prove that entropic interpolations solve Newton's law. But before doing so, let me introduce a notational convention that simplifies many of the expressions to come: whenever there is no ambiguity, I will not specify the time interval over which a curve or a vector field is defined. For example, I write (μ_t) instead of $(\mu_t)_{t \in [0, T]}$.

2.2.2.1 Heuristics

Here, in order to lighten notation and explain the main ideas in the simplest possible way, I shall work on \mathbb{R}^d and at the formal level. However, all construction carry over to Riemannian manifolds in a natural way. A fundamental result of Brenier asserts that if μ is *regular*, and in particular if μ admits a density against the Lebesgue measure, then for any $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a convex function θ such that the coupling

$$\pi = (\mathbf{id}, \nabla \theta)_{\#} \mu$$

is optimal for the Monge Kantorovich problem (2.1.1) with $c(x, y) = |x - y|^2$, where I denote \mathbf{id} the identity map. Moreover, the corresponding *displacement interpolation* $(\mu_t)_{t \in [0, 1]}$ defined via

$$\mu_t = (\mathbf{id} + t \nabla \varphi)_{\#} \mu, \quad \text{with } \varphi = \nabla \theta - \mathbf{id}, \quad (2.2.3)$$

then (μ_t) is a constant speed geodesic in the sense that

$$W_2(\mu_s, \mu_t) = |t - s|W_2(\mu, \nu) \quad \forall 0 \leq s, t \leq 1.$$

Of course, the above definition of geodesic is purely metric, and not geometric and we have to work more to properly identify the underlying Riemannian structure. If we were on a classical smooth finite dimensional Riemannian manifold, we could associate to any geodesic defined over the interval $[0, 1]$ a vector in the tangent space by inverting the exponential map. The above discussion highlights how, at least if the starting point is regular, we can associate to a geodesic a gradient vector field, that is $\nabla\varphi$ in (2.2.3). Moreover, $\nabla\varphi$ belongs to $L^2(\mu)$. Thus, it is tempting to consider the following definition for the tangent space

$$\mathcal{T}_\mu\mathcal{P}_2 = \overline{\{\nabla\varphi : \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}. \quad (2.2.4)$$

The next step is that of defining the Riemannian metric on the tangent space. To this aim, I recall that on a standard Riemannian manifold one can recover the squared norm of the difference of two vectors in the tangent space by expanding the distance between the corresponding geodesics around $h = 0$. Following this thread, for given tangent vectors $\nabla\varphi, \nabla\psi \in \mathcal{T}_\mu\mathcal{P}_2$, I "define" the Riemannian metric by considering

$$|\nabla\varphi - \nabla\psi|_{\mathcal{T}_\mu\mathcal{P}_2} := \lim_{h \downarrow 0} \frac{W_2(\mu_h, \nu_h)}{h}, \quad \mu_h = (\mathbf{id} + h\nabla\varphi)_\# \mu, \nu_h = (\mathbf{id} + h\nabla\psi)_\# \mu.$$

The limit can be computed explicitly and leads to consider the following inner product on $\mathcal{T}_\mu\mathcal{P}_2$.

$$\langle \nabla\varphi, \nabla\psi \rangle_{\mathcal{T}_\mu\mathcal{P}_2} = \int_{\mathbb{R}^d} \nabla\varphi \cdot \nabla\psi \, d\mu.$$

I have just defined the *Otto metric*. I now proceed on to show how to compute gradients, velocities and accelerations.

First order calculus Here I deal with the problem of defining the velocity of a curve that is not necessarily a geodesic. I accomplish this following again the intuition we have on finite dimensional Riemannian manifolds, where the speed of a curve is found looking for the geodesic that best approximates over a short time interval. That is to say, if $(\mu_t)_{t \in [0, T]}$ is a "smooth" curve, its velocity at t is identified with the tangent vector $v_t \in \mathcal{T}_{\mu_t}\mathcal{P}_2$ such that

$$\lim_{h \rightarrow 0} \frac{d(\mu_{t+h}, \gamma_h)}{h} = 0, \quad \gamma_h = (\mathbf{id} + hv_t)_\# \mu_t \quad (2.2.5)$$

This definition, although it conveys the right geometrical intuition, is not of practical use. In order to obtain a more efficient definition, we invoke the powerful characterization of absolutely continuous curves [5] in $\mathcal{P}_2(\mathbb{R}^d)$ that gives existence of a unique family of vector fields (v_t) solving the continuity equation for (μ_t)

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

in the weak sense and such that $v_t \in \mathcal{T}_{\mu_t}\mathcal{P}_2$ t -almost everywhere. The vector field v_t is indeed the velocity of the curve (μ_t) at t in the sense of (2.2.5). To get convinced of this, consider the family of *flow maps* $X(t, s, x)$ generated by (v_t) defined by

$$\begin{cases} \frac{d}{ds} X(t, s, x) = v_s(X(t, s, x)), \quad \forall t \in [0, T], x \in \text{supp}(\mu_t), \text{ a.e. } s \in [0, T], \\ X(t, t, x) = x, \quad \forall t \in [0, T], x \in \text{supp}(\mu_t). \end{cases} \quad (2.2.6)$$

The flow maps have the following important properties

$$\begin{aligned} X(r, s, X(t, r, \cdot)) &= X(t, s, \cdot), \quad \forall t, r, s \in [0, T], \\ X(t, s, \cdot)_\# \mu_t &= \mu_s, \quad \forall t, s \in [0, T]. \end{aligned} \quad (2.2.7)$$

But then, for a fixed t we have

$$\mu_{t+h} = X(t, t+h, \cdot)_\# \mu_t \stackrel{(2.2.6)}{=} (\mathbf{id} + hv_t + o(h))_\# \mu_t$$

from which (2.2.5) follows. We have now all the tools at hand to ¹ [24] interpret the Benamou-Brenier formula as the fact that the geodesic distance generated by Otto's metric is the Wasserstein distance. Indeed, the formula reads as

$$W_2^2(\mu, \nu) = \inf_{\mu_t, v_t} \int_0^1 |v_t|_{\mathcal{T}_{\mu_t} \mathcal{P}_2}^2 dt,$$

where (μ_t) varies in the set of absolutely continuous curves joining μ and ν and (v_t) is the velocity of (μ_t) .

Second order calculus I now turn to the definition of the Levi-Civita connection. Given a curve $(\mu_t)_{t \in [0, T]}$ and a tangent vector $u_t \in \mathcal{T}_{\mu_t} \mathcal{P}_2$ for all $t \in [0, T]$, the basic idea behind the construction of the covariant derivative is to say that u_t is a parallel vector field in $\mathcal{P}_2(\mathbb{R}^d)$ along (μ_t) if and only if for any x , the vector field $s \mapsto u_s(X(0, s, x))$ is a parallel vector field along the curve $s \mapsto X(0, s, x)$ in \mathbb{R}^d . In the case when the base manifold is the euclidean space \mathbb{R}^d , this condition takes the particularly simple form

$$u_{t+h}(X(t, t+h, x)) = u_t(x) \quad \forall x \in \mathbb{R}^d, t, h \in [0, T].$$

Moreover, if we denote by $\text{proj}_{\mu_t} : L^2(\mu_t) \rightarrow \mathcal{T}_{\mu_t} \mathcal{P}_2$ the projection operator, and recalling that we assumed $u_t \in \mathcal{T}_{\mu_t} \mathcal{P}_2$, the above equation is in fact equivalent to

$$\text{proj}_{\mu_t}(u_{t+h}(X(t, t+h, \cdot))) = u_t(\cdot) \quad \forall t, h \in [0, T].$$

Therefore, using that $X(t, t+h, x) = x + hv_t(x) + o(h)$, one expects that the vector field (u_t) to be parallel along the curve (μ_t) if and only if

$$\text{proj}_{\mu_t}(\partial_t u_t + Du_t \cdot v_t) = 0,$$

and the covariant derivative of (u_t) along (μ_t) should be given by

$$\frac{\mathbf{D}}{dt} u_t := \text{proj}_{\mu_t}(\partial_t u_t + Du_t \cdot v_t). \quad (2.2.8)$$

2.2.2.2 Some rigorous constructions

I provide here the fundamental definitions that allow to turn some of the above heuristic arguments into rigorous mathematical statements. From now on, I no longer work on \mathbb{R}^d , but the base space is a smooth Riemannian manifold (M, g) . I assume (M, g) to be complete, connected, closed, without boundary, and of finite dimension and denote its Levi-Civita connection by $\bar{\nabla}$. The material of this section is essentially taken from [138].

Wasserstein distance on M The Wasserstein distance $W_2(\mu, \nu)$ between $\mu, \nu \in \mathcal{P}_2(M)$ and I recall that the Wasserstein distance of order two is defined as

$$W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{M^2} d_M(x, y) \pi(dx dy),$$

where d_M is the Riemannian distance on M . In the rest of this section, curves are always defined on the time interval $[0, T]$, unless otherwise stated.

Geodesics, Velocity fields I recall here the notion of absolutely continuous curve and geodesic in the context of general metric spaces

Definition 2.2. *Let (E, d) be a metric space. A curve (x_t) is absolutely continuous over $[\varepsilon, T - \varepsilon]$ provided that for some integrable function f*

$$\forall \varepsilon \leq r < s \leq T - \varepsilon \quad d(x_r, x_s) \leq \int_r^s f(t) dt.$$

¹In the original result of Benamou and Brenier v_t is not the velocity field of (μ_t) , but just an arbitrary weak solution to the continuity equation. However, it is easy to see that the representation formula for the Wasserstein distance remains true if we restrict the minimization to the couples (μ_t, v_t) such that (v_t) is the velocity field of (μ_t) .

In all what follows, I will write "absolutely continuous curve" and mean "absolutely continuous over $[\varepsilon, T - \varepsilon]$ for all $\varepsilon \in (0, T)$ ". The fundamental characterization of absolutely continuous curves on $(\mathcal{P}_2(M), W_2(\cdot, \cdot))$, see [4, Th. 2.29], states that if (μ_t) is an absolutely continuous curve then there exists a Borel family of vector fields (v_t) such that the *continuity equation*

$$\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \quad (2.2.9)$$

holds in the sense of distributions and such that

$$v_t \in \mathcal{T}_{\mu_t} \mathcal{P}_2 \quad t - \text{a.e.},$$

where the definition of $\mathcal{T}_{\mu} \mathcal{P}_2$ was given at (2.2.4). We call $(v_t)_{t \in [0, T]}$ the *velocity field* of (μ_t) . Conversely, if (v_t) is a Borel family of vector fields satisfying (2.2.9) such that

$$\forall \varepsilon \in (0, T), \int_{\varepsilon}^{T-\varepsilon} |v_t|_{L^2(\mu_t)} dt < +\infty, \quad v_t \in \mathcal{T}_{\mu_t} \mathcal{P}_2 \quad t - \text{a.e.},$$

then (μ_t) is absolutely continuous and (v_t) is its velocity field.

Definition 2.3. *Let (E, d) be a metric space. A curve $(x_t)_{t \in [0, 1]}$ is a constant speed geodesic if and only if*

$$\forall s, t \in [0, 1] \quad d(x_s, x_t) = |t - s|d(x_0, x_1).$$

(E, d) is said to be a geodesic space provided that for any pair of points there exist a constant speed geodesic connecting them.

It turns out that ([4, Th 2.10]) $(\mathcal{P}_2(M), W_2(\cdot, \cdot))$ is a geodesic space.

Regular curves and flow maps I give the definition of regular curve following closely [138, Def. 2.8], the only difference being that I define regularity over $[\varepsilon, T - \varepsilon]$ for $\varepsilon \in (0, T)$, instead of looking at $[0, 1]$.

Definition 2.2.1. *For $\varepsilon \in (0, T)$, an absolutely continuous curve (μ_t) is regular over $[\varepsilon, T - \varepsilon]$ provided*

$$\int_{\varepsilon}^{T-\varepsilon} |v_t|_{\mathcal{T}_{\mu_t} \mathcal{P}_2}^2 dt < +\infty \quad (2.2.10)$$

and

$$\int_{\varepsilon}^{T-\varepsilon} L(v_t) dt < +\infty, \quad (2.2.11)$$

where for a smooth vector field ξ , $L(\xi)$ is defined as

$$L(\xi) = \sup_{\substack{x \in M \\ w: |w|=1}} |\overline{\nabla}_w \xi(x)|.$$

For non smooth vector fields, the general definition of L can be found at [138, Def 2.1]; in this manuscript we will only be concerned with the smooth case. In all what follows, by regular curve I mean "regular over $[\varepsilon, T - \varepsilon]$ for all $\varepsilon \in (0, T)$ ". If (μ_t) is a regular curve and (v_t) its velocity field, by means of standard Cauchy-Lispchitz theory one can show that there exists a unique family of maps, called *flow maps*, $X(t, s, \cdot) : \text{supp}(\mu_t) \rightarrow \text{supp}(\mu_s)$ such that for any $\varepsilon \in (0, T)$, $t \in (0, T)$, $x \in \text{supp}(\mu_t)$ the curve $s \mapsto X(t, s, x)$ is absolutely continuous over $(\varepsilon, T - \varepsilon)$ and satisfies (2.2.6). Moreover, the flow maps enjoy the properties (2.2.7).

The maps $(\tau_x)_s^t$ and $\tau_s^t(u)$ These maps are needed to define the covariant derivative: the first family acts between tangent spaces of the base manifold M , whereas the second family acts on vector fields: the following are Definition 2.9 and 2.12 in [138].

Definition 2.2.2. *Let (μ_t) be a regular curve and $(X(t, s, \cdot))_{s, t \in [0, T]}$ its flow maps. Given $s, t \in [0, T]$ and $x \in \text{supp}(\mu_t)$, we let $(\tau_x)_s^t : T_{X(t, s, x)} M \rightarrow T_x M$ be the map which associate to $v \in T_{X(t, s, x)} M$ its parallel transport along the absolutely continuous curve $r \mapsto X(t, r, x)$ from $r = s$ to $r = t$.*

The map $(\tau_x)_s^t$ is used to transport vectors along regular curves in M . In the next definition we shall see how the maps τ_s^t do the same for vector fields along regular curves in $\mathcal{P}_2(M)$. It should be stressed that the next definition is *not* the parallel transport. Indeed, in general, $\tau_s^t(u)$ may not be in $\mathcal{T}_{\mu_t}\mathcal{P}_2$.

Definition 2.2.3. *Let (μ_t) be a regular curve, $X(t, s, \cdot)$ its flow maps and u a vector field in $L^2(\mu_s)$. Then $\tau_s^t(u)$ is the vector field in $L^2(\mu_t)$ defined by*

$$\tau_s^t(u)(x) = (\tau_x)_s^t(u \circ X(t, s, x)).$$

The maps τ_s^t have similar properties to the flow maps: for example they enjoy the group property

$$\tau_s^t = \tau_r^t \circ \tau_s^r. \quad (2.2.12)$$

Moreover, τ_s^t is an isometry from $L^2(\mu_s)$ to $L^2(\mu_t)$, i.e.

$$\forall u \in L^2(\mu_s), \quad \int_M |u|^2 d\mu_s = \int_M |\tau_s^t(u)|^2 d\mu_t. \quad (2.2.13)$$

Vector fields along curves It is not possible in general to define a notion of covariant derivative for vector fields on $\mathcal{P}_2(M)$. However, for the purposes of this manuscript it is enough to define covariant derivative of absolutely continuous vector fields along regular curves, and this is what I am going to do in the next paragraphs.

Definition 2.2.4. *A vector field along a curve (μ_t) is a Borel map $(t, x) \mapsto u_t(x) \in T_x M$ such that $u_t \in L^2(\mu_t)$ for a.e. t . It will be denoted by (u_t) .*

Observe that also non tangent vector fields are considered in this definition, i.e. u_t may not be a gradient. Here is the definition of absolutely continuous vector field along a curve, see [138, Def. 3.2].

Definition 2.2.5. *Let (u_t) be a vector field along the regular curve (μ_t) and $\tau_s^t(u)$ be given by Definition 2.2.3. We say that (u_t) is absolutely continuous over $[\varepsilon, T - \varepsilon]$ provided the map*

$$t \mapsto \tau_t^{t_0}(u_t) \in L^2_{\mu_{t_0}}$$

is absolutely continuous in $L^2_{\mu_{t_0}}$ for all $t_0 \in [\varepsilon, T - \varepsilon]$.

It can be seen that the choice of t_0 is irrelevant and one could then set $t_0 = \varepsilon$ in the definition above. As before, by absolutely continuous vector field I mean "absolutely continuous over $[\varepsilon, T - \varepsilon]$ for all $\varepsilon \in (0, T)$ ".

Total derivative and covariant derivative We are ready to define the total derivative of an absolutely continuous vector field as in [138, Def 3.6]. Note that this is not yet the covariant derivative.

Definition 2.2.6. *Let (u_t) be an absolutely continuous vector field along the regular curve (μ_t) . Its total derivative is defined as*

$$\frac{\mathbf{d}}{dt} u_t := \lim_{h \rightarrow 0} \frac{\tau_{t+h}^t(u_{t+h}) - u_t}{h} \quad t - a.e., \quad (2.2.14)$$

where the limit is intended in $L^2(\mu_t)$.

To define the covariant derivative, we consider the orthogonal projection $\text{proj}_\mu : L^2(\mu) \rightarrow \mathcal{T}_\mu\mathcal{P}_2$ of the total derivative. The following is [4, Def. 6.8] for the flat case. For the general case, I refer to Definition 5.1 and discussion thereafter in [138].

Definition 2.2.7. *Let (u_t) be an absolutely continuous and tangent vector field along the regular curve (μ_t) . Its covariant derivative is defined as*

$$\frac{\mathbf{D}}{dt} u_t := \text{proj}_{\mu_t} \left(\frac{\mathbf{d}}{dt} u_t \right) \quad t - a.e..$$

It can be shown that if (μ_t) is a regular curve and its velocity field (v_t) is absolutely continuous, then the acceleration of (μ_t) can be computed by means of (2.2.8) and we have

$$\frac{\mathbf{D}}{dt} v_t = \partial_t v_t + \frac{1}{2} \nabla (|v_t|^2). \quad (2.2.15)$$

Levi-Civita connection The covariant derivative we have just defined is indeed the Levi-Civita connection, in the sense that it satisfies the *compatibility of the metric* and the *torsion free identity*. The compatibility of the metric says that if $(u_t^1), (u_t^2)$ are tangent absolutely continuous vector fields along the regular curve (μ_t) , then

$$\frac{d}{dt}\langle u_t^1, u_t^2 \rangle_{\mathcal{T}_{\mu_t}\mathcal{P}_2} = \left\langle \frac{D}{dt}u_t^1, u_t^2 \right\rangle_{\mathcal{T}_{\mu_t}\mathcal{P}_2} + \left\langle u_t^1, \frac{D}{dt}u_t^2 \right\rangle_{\mathcal{T}_{\mu_t}\mathcal{P}_2}. \quad (2.2.16)$$

Since we do not have a notion of covariant derivative for vector fields at hand, the torsion free identity cannot be expressed in the familiar way $[X, Y] = \bar{\nabla}_X Y - \bar{\nabla}_Y X$ and is therefore formulated as follows: if $(\mu_t^1), (\mu_t^2)$ are two regular curves and $(u_t^i), i = 1, 2$ are tangent vector fields along $(\mu_t^i), i = 1, 2$ with $u_0^1 = v_0^2$ and $u_0^2 = v_0^1$, then we have:

$$\left. \frac{d}{dt} \left(\frac{d}{dh} F_\varphi(\mu_t^{2,h}) \Big|_{h=0} \right) \right|_{t=0} - \left. \frac{d}{dt} \left(\frac{d}{dh} F_\varphi(\mu_t^{1,h}) \Big|_{h=0} \right) \right|_{t=0} = \left\langle \nabla \varphi, \frac{D}{dt}u_t^2 - \frac{D}{dt}u_t^1 \right\rangle_{\mathcal{T}_{\mu_t}\mathcal{P}_2} \Big|_{t=0},$$

where F_φ is the map $\mu \mapsto \int \varphi d\mu$ for some $\varphi \in C_c^\infty(M)$ and for any $i = 1, 2$ and $t \in [0, T]$, the curve $h \mapsto \mu_t^{i,h}$ is the constant speed geodesic $h \mapsto (\mathbf{id} + hu_t^i) \# \mu_t^i$, that is well defined on a neighbourhood of $h = 0$. Note that a simple calculation gives

$$\left. \frac{d}{dh} F_\varphi(\mu_t^{i,h}) \Big|_{h=0} = \int \langle \nabla \varphi, u_t^i \rangle d\mu_t^i,$$

but I prefer to keep the more implicit notation as it conveys better the underlying geometric concept. Proofs of the compatibility with the metric and torsion free identity can be found in [4, Sec 6.2] for $M = \mathbb{R}^d$ and in [138, Sec. 5.1] for the general case.

Gradient of the relative entropy and Fisher information For most functionals of interest on $\mathcal{P}_2(M)$ the existence of a gradient in the classical sense does not hold. Here, following [4, Eq 3.50], I define gradients through directional derivatives. Although a stronger definition through the notion of subdifferential can be given for the relative entropy (see [5, Ch 9,10]), this does not seem to be the case for the Fisher information, that will play a crucial role in the Newton's law. In this definition and in the rest of this section by $\mu \in \mathcal{C}_+^\infty(M)$ we mean that $\mu \ll \mathbf{vol}$ and $\frac{d\mu}{d\mathbf{vol}} \in \mathcal{C}_+^\infty(M)$, where \mathbf{vol} is the volume measure of (M, g) .

Definition 2.2.8. Let $\mu \in \mathcal{C}_+^\infty(M)$. We say that $\mathcal{F} : \mathcal{P}_2(M) \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is differentiable at μ if there exists a vector field $w \in \mathcal{T}_\mu\mathcal{P}_2$ such that for all $\varphi \in \mathcal{C}_c^\infty(M)$

$$\lim_{h \rightarrow 0} \frac{\mathcal{F}((\mathbf{id} + h\nabla\varphi) \# \mu) - \mathcal{F}(\mu)}{h} = \langle w, \nabla\varphi \rangle_{\mathcal{T}_\mu\mathcal{P}_2}. \quad (2.2.17)$$

It follows that if \mathcal{F} is differentiable at μ , then there exists a unique w fulfilling (2.2.17). We then call w the *gradient* of \mathcal{F} at μ , and denote it $\nabla^{W_2}\mathcal{F}(\mu)$. If we denote by $\mathcal{H}(\cdot|\mathbf{vol})$ the relative entropy with respect to the volume measure \mathbf{vol} , it is known since the seminal paper [151] that

$$\nabla^{W_2}\mathcal{H}(\cdot|\mathbf{vol})(\mu) = \nabla \log \frac{d\mu}{d\mathbf{vol}}, \quad (2.2.18)$$

provided $\nabla \log \mu \in L^2(\mu)$. Recalling the definition of the Fisher information functional $\mathcal{I}(\cdot)$

$$\mathcal{I}(\mu); = \begin{cases} \int_M |\nabla \log \frac{d\mu}{d\mathbf{vol}}|^2 d\mu, & \text{if } \mu \ll \mathbf{vol}, \nabla \log \frac{d\mu}{d\mathbf{vol}} \in L^2(\mu), \\ +\infty, & \text{otherwise,} \end{cases} \quad (2.2.19)$$

we immediately derive from (2.2.18) its geometric interpretation as the squared norm of the gradient of the relative entropy:

$$\mathcal{I}(\mu) = |\nabla^{W_2}\mathcal{H}(\cdot|\mathbf{vol})|_{\mathcal{T}_\mu\mathcal{P}_2}^2. \quad (2.2.20)$$

From now on, since there is no ambiguity, I will often use the shorthand notation μ for $\frac{d\mu}{d\mathbf{vol}}$. The gradient of the Fisher information at μ is the vector field

$$\nabla^{W_2}\mathcal{I}(\mu) = \nabla \left(-|\nabla \log \mu|^2 - 2\Delta \log \mu, \right), \quad (2.2.21)$$

provided the right hand side is well defined and belongs to $L^2(\mu)$ and μ is regular enough. A detailed computation of the gradient of the Fisher information can be found in [210].

Hessian of the entropy and convexity Recall that on a smooth finite dimensional Riemannian manifold whose Levi-Civita connection is $\bar{\nabla}$, the Hessian of f at x applied to $v \in \mathcal{T}_x M$ is defined through (see e.g. [105, Ex. 11, pg. 141])

$$\nabla_x^2 f(v) = \bar{\nabla}_v \nabla f(x) = \frac{\mathbf{D}}{dt} \nabla f(x_t) \Big|_{t=0},$$

where (x_t) is any curve such that $x_0 = x$, $\dot{x}_0 = v$. Here, we are only interested in defining a kind of Hessian for $\mathcal{H}(\cdot|\mathbf{vol})$. Therefore, as we did before, to simplify the definitions, we restrict to a very special setting.

Definition 2.2.9. *Let M be compact. Consider a measure $\mu \in \mathcal{C}_+^\infty(M)$ and a vector field $v \in \mathcal{C}_b^\infty(M) \cap \mathcal{T}_\mu \mathcal{P}_2$. We define*

$$\mathbf{Hess}_\mu^{W_2} \mathcal{H}(\cdot|\mathbf{vol})(v) = \frac{\mathbf{D}}{dt} \left(\nabla^{W_2} \mathcal{H}(\mu_t|\mathbf{vol}) \right) \Big|_{t=0},$$

where (μ_t) is any regular curve such that $\mu_0 = \mu$, $\mu_t \in \mathcal{C}_+^\infty([0, \varepsilon] \times M)$ for some $\varepsilon > 0$ and whose velocity field (v_t) is such that $v_0 = v$.

It is easy to see that this is a good definition in the sense that there is always one curve fulfilling the requirements and the value of the Hessian does not depend on the specific choice of the curve. If the *Ricci curvature* of M is bounded below or, more generally if the Bakry-Émery condition

$$\forall x \in M, \quad \mathbf{Ric}_x + 2\mathbf{Hess}_x U \geq \kappa \mathbf{id}$$

holds, where \mathbf{Ric}_x is the Ricci curvature tensor at x , then $\mathcal{H}(\cdot|\mathbf{vol})$ is displacement κ -convex, i.e. κ -convex along geodesics [171, 211]. This means that if the Bakry-Émery condition holds then

$$\forall \mu \in \mathcal{C}_+^\infty(M), v \in \mathcal{C}_b^\infty \cap \mathcal{T}_\mu \mathcal{P}_2, \quad \langle \mathbf{Hess}_\mu^{W_2} \mathcal{H}(\cdot|\mathbf{vol})(v), v \rangle_{\mathcal{T}_\mu \mathcal{P}_2} \geq \kappa |v|_{\mathcal{T}_\mu \mathcal{P}_2}^2. \quad (2.2.22)$$

Displacement convexity of the entropy and other functionals is at the heart of most quantitative results obtained by optimal transport methods, and plays a fundamental role in the quantitative analysis of Schrödinger bridges, as I will illustrate in Chapter 4.

2.2.3 Newton's law

Now that the geometric formalism has been introduced, I can state Newton's law for the entropic interpolation that, I recall, is the optimal flow in (1.1.5). Here, I make the following hypothesis on the manifold and the marginal measures

Assumptions 2.2.1. *M, μ, ν satisfy*

- M is compact,
- $\mathcal{H}(\mu|\mathbf{vol}), \mathcal{H}(\nu|\mathbf{vol}) < +\infty$,

and consider the Schrödinger problem (**CSP**) where R is the law of a Brownian motion on M over $[0, T]$ with initial distribution \mathbf{vol} : under the standing assumptions it is well known that the problem admits a unique solution. The following result is drawn from [81].

Theorem 2.2.2. *Let M, μ, ν be such that Assumption 2.2.1 holds. Then the entropic interpolation $(\mu_t)_{t \in [0, T]}$ is a regular curve and its velocity field $(v_t)_{t \in [0, T]}$ is absolutely continuous. Moreover, $(\mu_t)_{t \in [0, T]}$ satisfies the equation*

$$\forall 0 < t < T, \quad \frac{\mathbf{D}}{dt} v_t = \frac{1}{8} \nabla^{W_2} \mathcal{I}(\mu_t), \quad (2.2.23)$$

where $\frac{\mathbf{D}}{dt} v_t$ is the covariant derivative of (v_t) along (μ_t) and \mathcal{I} is the Fisher information functional (2.2.19).

The interested reader will find in the original article [95] various generalizations of this result under weaker assumptions on M, μ, ν and for more general reference measures. In particular, the case when R_{0T} is the joint law of the stationary solution of

$$dX_t = -\nabla U(X_t) dt + dB_t$$

is treated there as well as the case when R_{0T} is a *Feynman-Kac* penalisation of the Brownian motion [193].

Theorem 2.2.2 gives an answer to the problem of determining what second order equation captures the dynamics of the bridge of a diffusion process. Several authors (see e.g. [156, 157, 100, 205, 214, 213]) have proven results in this direction. In this respect, equation (2.2.23) has some nice features. The first one is that the acceleration we consider here is a true acceleration, in the sense that it corresponds to the covariant derivative associated with a Riemannian structure. One of the chief advantages of having an equation expressed within a Riemannian manifold is that it allows for explicit computations; we shall see how useful this is in view of obtaining quantitative estimates and functional inequalities in Chapter 4. Another interesting observation is that Theorem 2.2.23 may be viewed as the second order counterpart of the interpretation of the marginal flow of the Brownian motion as a gradient flow in the Wasserstein space. This result, proved in [151] is one of the reasons for the explosion of interest around optimal transport over the last two decades.

Related works After some manipulations, it is possible to reinterpret the fluid dynamic formulations of (CSP) obtained in [135] and [67] as variational problems in the Riemannian manifold of optimal transport: equation (2.2.23) is then the associated Euler-Lagrange equation. Thus, in principle we could link (2.2.23) with the theory of Hamiltonian systems in $\mathcal{P}_2(M)$ developed in [3]. However, our proof is based on probabilistic arguments, and the Schrödinger problem is not covered by the results therein. By changing the sign in the right hand side of (2.2.23), one gets a nice connection with the Schrödinger equation, see [210, 75]. Using the hot gas experiment, we can give an heuristic for equation (2.2.23) to hold. Finally, it is worth mentioning that (2.2.23) justifies a posteriori the validity of the toy model in capturing the essential geometrical structure of the Schrödinger problem. This has led some authors to introduce general abstract versions of the Schrödinger problem that, unlike the problems discussed in this manuscript, do not necessarily admit a large deviation interpretation, see [176, 136, 158, 167, 74] for a sample of the recent developments.

Sketch of proof For simplicity, I take $M = \mathbb{R}^d$ whence \mathbf{vol} is the Lebesgue measure. Now that the intuition is all there, the formal proof is "just" a single computation. Indeed, taking conditional expectation in (2.1.5) yields an analogous of the fg -decomposition (2.1.5) valid for any $t \in [0, T]$, i.e. we have

$$d\mu_t = f_t g_t d\mathbf{vol}$$

where (f_t, g_t) solve the forward-backward system

$$\begin{cases} \partial_t f_t = \frac{1}{2} \Delta f_t, \\ \partial_t g_t = -\frac{1}{2} \Delta g_t, \\ f_0 g_0 = \frac{d\mu}{d\mathbf{vol}}, f_T g_T = \frac{d\nu}{d\mathbf{vol}}. \end{cases}$$

From this decomposition, we obtain that the velocity field of (μ_t) is

$$v_t = \frac{1}{2} \nabla (\log g_t - \log f_t) \tag{2.2.24}$$

Let us now compute the acceleration $\frac{D}{dt} v_t$ using the formula (2.2.15). We find

$$\begin{aligned} \partial_t v_t &= -\frac{1}{4} \nabla \left(\frac{\Delta f_t}{f_t} + \frac{\Delta g_t}{g_t} \right) \\ &= -\frac{1}{4} \nabla \left(\Delta \log \mu_t + |\nabla \log f_t|^2 + |\nabla \log g_t|^2 \right) \\ &= -\frac{1}{8} \nabla \left(2\Delta \log \mu_t + |\nabla \log \mu_t|^2 + |\nabla \log g_t - \nabla \log f_t|^2 \right) \\ &\stackrel{(2.2.21)}{=} \frac{1}{8} \nabla^{W_2} \mathcal{I}(\mu_t) - \frac{1}{8} \nabla (|\nabla \log g_t - \nabla \log f_t|^2) \end{aligned}$$

On the other hand we have directly from (2.2.24)

$$\frac{1}{2} \nabla |v_t|^2 = \frac{1}{8} \nabla (|\nabla \log g_t - \nabla \log f_t|^2),$$

which gives the desired result, recalling the definition of $\frac{D}{dt} v_t$ given at (2.2.15).

2.3 McKean Vlasov FBSDE for the mean field Schrödinger problem

I now proceed to the analysis of optimality conditions for Schrödinger problems in stochastic control formulation focusing on (MFSP). At the conceptual level, I abandon any considerations about Riemannian geometry and rather work with stochastic calculus. The mean field Schrödinger problem we studied in [7] corresponds to consider in the N -particle system, the following simplified version of (1.2.4), where interaction between particles is only in the drifts and through a pair potentials W :

$$\begin{cases} dX_t^i = -\nabla W * \mu_t^N(X_t^i)dt + dB_t, \\ X_0^i \sim \mu, \end{cases} \quad (2.3.1)$$

Correspondingly, in this setting the controlled diffusion process $\Gamma(P)$ defined at (1.2.5) is given by

$$\begin{cases} dZ_t = -\nabla W * P_t(Z_t)dt + dB_t, \\ Z_0 \sim \mu, \end{cases}$$

and (MFSP) is the task of minimizing $\mathcal{H}(P|\Gamma(P))$ among all path probability measures satisfying given marginal constrains. The Mc-Kean Vlasov limit of (2.3.1) is sometimes called granular media equation and is widely studied in the literature. It is beyond the scope of this thesis to make a comprehensive overview of the important contributions to the study of the granular media equation. Instead, I limit myself to mention [61, 64, 37, 202] as these works are among those closely related to the results we derived. We wrote our optimality conditions in terms of a McKean-Vlasov FBSDE. Classical FBSDEs have a long history and are at the heart of the probabilistic approach to stochastic control [169] as they provide with a pathwise representation of solutions to Hamilton Jacobi Bellman PDEs. McKean-Vlasov FBSDEs are much more recent objects, introduced first by Carmona and Delarue with the aim of developing a probabilistic analysis of mean field games [57] and mean field control [58], see also the monograph [59]. The mean field Schrödinger problem falls indeed into the realm of mean field stochastic control problems.

2.3.1 Mean Field Schrödinger bridges

We make the following assumption on the potential W and on the measure inputs of the problem

Assumptions 2.3.1. *The interaction potential W satisfies*

$$W \text{ is of class } \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}) \text{ and symmetric, i.e. } W(\cdot) = W(-\cdot). \quad (2.3.2)$$

Moreover, the Hessian of W is bounded

$$\sup_{z, v \in \mathbb{R}^d, |v|=1} v \cdot \nabla^2 W(z) \cdot v < +\infty. \quad (2.3.3)$$

The initial and final distributions μ, ν satisfy

$$\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d) \text{ and } \tilde{\mathcal{F}}(\mu), \tilde{\mathcal{F}}(\nu) < +\infty, \quad (2.3.4)$$

where the free energy $\tilde{\mathcal{F}}$ is defined for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ by

$$\tilde{\mathcal{F}}(\mu) = \begin{cases} \int_{\mathbb{R}^d} \log \frac{d\mu(x)}{d\mathbf{vol}} \mu(dx) + \int_{\mathbb{R}^d} W * \mu(x) \mu(dx), & \text{if } \mu \ll \mathbf{vol} \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.3.5)$$

Under these assumptions, MFSP becomes the following McKean-Vlasov control problem

$$\inf \left\{ \frac{1}{2} \mathbb{E} \left[\int_0^T |\alpha_t^P|^2 dt \right] : P \in \mathcal{P}(\Omega), (X_0)_\# P = \mu, (X_T)_\# P = \nu, P \in \text{Adm}(\mu, \nu) \right\},$$

where a path probability measure \mathbb{P} belongs to the admissible set $\text{Adm}(\mu, \nu)$ if and only if there exists a predictable process $(\alpha_t^{\mathbb{P}})_{t \in [0, T]}$ s.t.

$$\mathbb{E}_{\mathbb{P}} \left[\int_0^T |\alpha_t^{\mathbb{P}}|^2 dt \right] < +\infty \quad (2.3.6)$$

and such that

$$X_t - \int_0^t (-\nabla W * \mathbb{P}_s(X_s) + \alpha_s^{\mathbb{P}}) ds \quad (2.3.7)$$

is a Brownian motion under \mathbb{P} . I recall that $(X_t)_{t \in [0, T]}$ always denote the canonical process.

Theorem 2.3.2. *Let Assumption (2.3.1) hold. Then (MFSP) admits at least an optimal solution. Moreover, if \mathbb{P} is optimal, the following holds*

(i) *There exists $\Psi \in H_{-1}((\mathbb{P}_t)_{t \in [0, T]})$ such that*

$$(dt \times d\mathbb{P}\text{-a.s.}) \quad \alpha_t^{\mathbb{P}} = \Psi_t(X_t). \quad (2.3.8)$$

(ii) *The process $t \mapsto \Psi_t(X_t)$ is continuous² and the process $(M_t)_{t \in [0, T]}$ defined by*

$$M_t := \Psi_t(X_t) - \int_0^t \tilde{\mathbb{E}}_{\tilde{\mathbb{P}}} \left[\nabla^2 W(X_s - \tilde{X}_s) \cdot (\Psi_s(X_s) - \Psi_s(\tilde{X}_s)) \right] ds \quad (2.3.9)$$

is a continuous martingale under \mathbb{P} on $[0, T[$, where $(\tilde{X}_t)_{t \in [0, T]}$ is an independent copy of $(X_t)_{t \in [0, T]}$ defined on some probability space $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}})$ and $\tilde{\mathbb{E}}_{\tilde{\mathbb{P}}}$ denotes the expectation on $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}})$.

Item (i) tells that optimal controls are Markov, square integrable and almost of gradient type. The key message of Theorem 2.3.2 is delivered at item (ii) where we give a rigorous meaning to the McKean-Vlasov FBSDE in the unknowns (X, Y, Z) :

$$\begin{cases} dX_t = -\tilde{\mathbb{E}}[\nabla W(X_t - \tilde{X}_t)]dt + Y_t dt + dB_t \\ dY_t = \tilde{\mathbb{E}}[\nabla^2 W(X_t - \tilde{X}_t) \cdot (Y_t - \tilde{Y}_t)]dt + Z_t \cdot dB_t \\ X_0 \sim \mu, X_T \sim \nu. \end{cases} \quad (2.3.10)$$

To see this, all we need to do is take $Y_t := \Psi_t(X_t)$ and reinterpret (2.3.7) for the dynamics of the canonical process X and (2.3.9) for the dynamics of Y , in the latter case using the martingale representation theorem. Moreover, at least formally one could see that $Z_t = D\Psi_t(X_t)$. The terminal condition $X_T \sim \nu$ is unconventional for FBSDEs and we can view the forward-backward system above as an example *planning McKean-Vlasov FBSDE*. To the best of my knowledge, such class of equations had not been studied prior to our work. On the other hand, coupled PDE systems of planning type arise naturally in the theory of mean field games and have recently received quite some attention, see e.g. [144].

Proof strategy At the moment we considered it, the system (2.3.10) was beyond the scope of existing FBSDE theory, such as Carmona and Delarue's [58, 57, 56, 59]. This is due to the above mentioned terminal constraint but also to the nature of the coefficients driving the dynamics of Y and the assumptions we made on W . Thus, we developed an ad hoc proof strategy that deviates from the standard *modus operandi* and relies essentially on an adaptation of stochastic calculus of variations [100] to the mean field setting, on a variational characterization of martingales due to Émery [112] and on results on the invertibility of shifts in Wiener space due to Üstünel [207]. In a nutshell, by controlling the behavior of the objective function under the action of shifts τ_h^ε of the form

$$\tau_h^\varepsilon : \Omega \longrightarrow \Omega, \quad (\tau_h^\varepsilon(\omega))_t = \omega_t + \varepsilon \int_0^t h_s(\omega) ds.$$

²More precisely, it has a continuous version adapted to the \mathbb{P} -augmented canonical filtration.

where $(h_t)_{t \in [0, T]}$ is a nice adapted process satisfying the loop condition

$$\int_0^T h_t dt = 0, \quad \text{P - a.s.},$$

we arrive at

$$\mathbb{E}\left[\int_0^T M_t h_t dt\right] = 0,$$

where M_t is the process defined at (2.3.9). The desired conclusion is then a direct consequence of Émery's result.

2.3.2 On the relation between Newton's laws and FBSDEs

I feel that there are two interesting observations to be made. The first one is that, at least at the formal level, in the setting of Theorem 2.3.2 entropic interpolations are still described by a Newton's law in the Riemannian metric of optimal transport. That is to say, if P is any optimizer and $(\mu_t)_{t \in [0, T]}$ its marginal flow, then

$$\frac{\mathbf{D}}{dt} v_t = \nabla^{W_2} \frac{1}{8} \mathcal{I}_{\tilde{\mathcal{F}}}(\mu_t), \quad \mu_0 = \mu, \mu_T = \nu$$

where (v_t) is the velocity field and $\frac{\mathbf{D}}{dt}$ is the covariant derivative defined in the former section. The functional $\mathcal{I}_{\tilde{\mathcal{F}}}$ is the Fisher information functional corresponding to $\tilde{\mathcal{F}}$:

$$\mathcal{I}_{\tilde{\mathcal{F}}}(\mu) = \int_{\mathbb{R}^d} \left| \nabla \log \left(\frac{d\mu}{d\text{vol}} \right) + 2\nabla W * \mu \right|^2(x) d\mu.$$

In fact, the validity of Newton's law extends to a much larger setting than the two examples discussed so far. I refer to [136] for an inspiring heuristic discussion on Newton's laws and generalized versions of the Schrödinger problem. The second interesting observation to be made is that, setting $W = 0$ for simplicity in Theorem 2.3.2, we recover the well known result that classical Schrödinger bridges are solutions to the FBSDE

$$\begin{cases} dX_t = Y_t dt + dB_t \\ dY_t = Z_t \cdot dB_t \\ X_0 \sim \mu, X_T \sim \nu, \end{cases} \quad (2.3.11)$$

enjoying the additional property that Y_t is (almost) of the form $Y_t = \nabla \psi_t(X_t)$. But then, since entropic interpolation are Newton's laws we have obtained the following

Informal statement *We have:*

- (i) If $(X_t, Y_t, Z_t)_{t \in [0, T]}$ is a solution for the FBSDE (2.3.11) such that $Y_t = \nabla \psi_t(X_t)$ for some time-varying potential ψ , then the marginal flow $(\mu_t)_{t \in [0, T]}$ of X_t is a solution for the Newton's law (2.2.23).
- (ii) If P is the (classical) Schrödinger bridge between μ and ν , then under P the canonical process $(X_t)_{t \in [0, T]}$ is such that there exist processes $(Y_t)_{t \in [0, T]}, (Z_t)_{t \in [0, T]}$ with the property that $(X_t, Y_t, Z_t)_{t \in [0, T]}$ is a solution for (2.3.11) and Y_t is as in (i)

We have therefore extended the stimulating parallelism between Itô and Otto calculus from first to second order calculus. More precisely we have

- (i) Classical SDEs provide with a pathwise representation of Wasserstein gradient flows. This was known since [151] and recently revisited and expanded in [153].
- (ii) FBSDEs provide with a pathwise representation of Newton's laws in the Riemannian manifold of optimal transport.

Note that there is no conceptual difficulty in adapting the above statements to include the interaction potential W . To sum up what I have presented in the last two sections, one could say that the dynamics of Schrödinger bridges can be equivalently be described in two seemingly very different ways by saying either that "the acceleration is the gradient of the Fisher information" or that "the drift is a martingale".

2.4 Perspectives

2.4.1 Mean Field Schrödinger problem

The coupled PDE system The FBSDE (2.3.10) provides with a probabilistic representation of the PDE system

$$\begin{cases} \partial_t \mu_t(x) - \frac{1}{2} \Delta \mu_t(x) + \nabla \cdot ((-\nabla W * \mu_t(x) + \nabla \psi_t(x)) \mu_t(x)) = 0 \\ \partial_t \nabla \psi_t(x) + \frac{1}{2} \nabla \Delta \psi_t(x) + \nabla^2 \psi_t(x) \cdot (-\nabla W * \mu_t(x) + \nabla \psi_t(x)) \\ = \int_{\mathbb{R}^d} \nabla^2 W(x - \tilde{x}) \cdot (\nabla \psi_t(x) - \nabla \psi_t(\tilde{x})) \mu_t(d\tilde{x}), \\ \mu_0(x) = \mu(x), \mu_T(x) = \nu(x). \end{cases} \quad (2.4.1)$$

Note that the second equation is indeed an equation in the unknown $\nabla \psi$, that has to be understood componentwise. Theorem 2.3.2 yields solutions in $H_{-1}((\mu_t)_{t \in [0, T]})$. However, observing that

$$\int_{\mathbb{R}^d} \nabla^2 W(x - \tilde{x}) \cdot (\nabla \psi_t(x) - \nabla \psi_t(\tilde{x})) \mu_t(d\tilde{x}) + \nabla^2 \psi_t(x) \cdot \nabla W * \mu_t(x) = \nabla \left(\int_{\mathbb{R}^d} \nabla W(\cdot - \tilde{x}) \cdot (\nabla \psi_t(\cdot) - \nabla \psi_t(\tilde{x})) \mu_t(d\tilde{x}) \right)(x),$$

we deduce that if we had a smooth solution for (2.4.1), then $(\mu_t, \psi_t)_{t \in [0, T]}$ would constitute a solution of

$$\begin{cases} \partial_t \psi_t(x) + \frac{1}{2} \Delta \psi_t(x) + \frac{1}{2} |\nabla \psi_t(x)|^2 = \int_{\mathbb{R}^d} \nabla W(x - \tilde{x}) \cdot (\nabla \psi_t(x) - \nabla \psi_t(\tilde{x})) \mu_t(d\tilde{x}), \\ \partial_t \mu_t(x) - \frac{1}{2} \Delta \mu_t(x) + \nabla \cdot ((-\nabla W * \mu_t(x) + \nabla \psi_t(x)) \mu_t(x)) = 0, \\ \mu_0(x) = \mu(x), \mu_T(x) = \nu(x). \end{cases} \quad (2.4.2)$$

This last PDE system has the typical structure encountered also in mean field games: a (non linear) Fokker Planck equation coupled with an HJB equation. However, due to the form of the interaction coefficients, it appears that, at least to the best of my knowledge, the existing theory [144] does not guarantee existence of solutions. Moreover, note that at the conceptual level, (MFSP) is *not* a mean field game. Thus, the question of how to profit from and refine the powerful machinery developed for coupled FP-HJB systems in order attack (2.4.2) is a natural one.

On the interaction coefficients The study of instances of (MFSP) where interaction happens also at the level of the diffusion matrix is of clear interest. For the particle system (1.2.4), this means choosing $\sigma(x, \mu)$ to have a non trivial dependence on μ . However, the proof methods we used in [7] do not permit to analyse this situation, essentially because of the well known fact that the law of two SDEs with different diffusion coefficients are not mutually absolutely continuous. But then, how do we obtain an FBSDE characterization of mean field Schrödinger bridges in this case? One possibility we are currently exploring is that of lifting the monotonicity principle from the static to the dynamic formulation of (CSP), and eventually move on to consider (MFSP). Recall that one of the strengths of building on the monotonicity principle is that we can avoid any convex duality argument. This is exactly what is needed, as the functional $\mathcal{H}(P|\Gamma(P))$ is not convex in general. Another research line (in collaboration with Zhenjie Ren) I am trying to pursue to tackle instances of MFSP with interaction in the diffusion coefficients is that of defining variations in a different way than what we did in [7], working on the controls instead of the canonical process.

2.4.2 Lattice gases

In Chapter 1 I introduced the Schrödinger problem for the simple exclusion process. As it was hinted there, this problem is a special representative of a larger class, that I shall call the Schrödinger problem for lattice gases, henceforth (SPLG). The content of this short section is organized in two paragraphs. In the first one, I briefly present the problem formulation using the terminology and formalism of Macroscopic Fluctuation Theory (MFT). It is however beyond the scope of this thesis to provide with a precise account of MFT and to make any physically relevant claim: I redirect the reader interested in MFT to [31]. In a second paragraph I present some results of the recent note [72], that I have written in collaboration with A. Chiarini and L. Tamanini. In this work, we formally obtained optimality conditions and convexity estimates for entropic functionals along Schrödinger bridges. Currently, we are working towards transforming the heuristic calculations we made into rigorous results.

Schrödinger problem for lattice gases Stochastic lattice gases may be described as a large collection of interacting particles performing random walks on the macroscopic blow-up $\Lambda_N = N\Lambda \cap \mathbb{Z}^d$ of a set $\Lambda \subseteq \mathbb{R}^d$. Particles are indistinguishable and the effect of the interaction is that the jump rates depend on the empirical particle distribution. Thermal reservoirs are modeled by adding creation/annihilation of particles at the boundary $\partial\Lambda_N$ and the influence of an external field is modeled by perturbing the rates giving a net drift toward a specified direction. Notable examples include the simple exclusion process, zero-range processes and the Glauber-Kawasaki dynamics. In the diffusive limit $N \rightarrow +\infty$, the empirical distribution of these systems typically converges to a hydrodynamic limit equation whose form is as follows

$$\partial_t \mu_t(x) + \nabla \cdot (J(t, \mu_t(x))) = 0, \quad \text{where } J(t, \mu_t(x)) = -D(\mu_t(x))\nabla \mu_t(x) + \chi(\mu_t(x))E(t).$$

where, as I already did sometimes before, I make no distinction between a probability measure and its density against the Lebesgue measure. The vector field $J(t, \mu_t(x))$ is called the hydrodynamic current, $\chi(\mu_t(x)) \geq 0$ is the mobility, $D(\mu_t(x)) \geq 0$ the diffusion coefficient and $E(t)$ the external field. I make here the assumption that D and χ are connected through the Einstein relation

$$D(\cdot) = f''(\cdot)\chi(\cdot), \tag{2.4.3}$$

where f is a convex function. The above equations are supplemented by the boundary conditions

$$f'(\mu_t(x)) = \lambda(t, x), \quad x \in \partial\Lambda.$$

All coefficients χ, D, E, λ are macroscopic quantities reflecting the microscopic description of the lattice gas: for example, the simple exclusion process corresponds to $\chi(\mu) = \mu(1 - \mu)$, $J(t, \mu) = -\nabla \mu$ and its hydrodynamic limit was studied in [154]; the zero-range dynamics corresponds to $\chi(\mu) = \varphi(\mu)$, $J(t, \mu) = -\varphi'(\mu)\nabla \mu$, see [31]. The rate function quantifying the large deviations from the hydrodynamic limit is given by the following "fundamental formula" in the context of macroscopic fluctuation theory [31]

$$\frac{1}{4} \int_0^T \int_{\Lambda} (j_t(x) - J(t, \mu_t(x))) \cdot \chi(\mu_t(x))^{-1} \cdot (j_t(x) - J(t, \mu_t(x))) \, dx dt, \tag{2.4.4}$$

where μ is the local density of particles and j_t is connected to μ_t by the equation (2.4.5) below. It is worth mentioning that the rate function (2.4.4) captures the large deviations behavior of other relevant interacting systems beyond lattice gases such as Ginzburg-Landau models [107]. The Schrödinger problem for lattice gases is therefore given by

$$\inf \left\{ \frac{1}{4} \int_0^T \int_{\Lambda} (j_t(x) - J(t, \mu_t(x))) \cdot \chi(\mu_t(x))^{-1} \cdot (j_t(x) - J(t, \mu_t(x))) \, dx dt : (\mu, j) \in \text{Adm}(\mu, \nu) \right\}, \tag{SPLG}$$

where a pair (μ, j) is admissible if and only if it is a (weak) solution of

$$\partial_t \mu_t(x) + \nabla \cdot j_t(x) = 0, \quad \mu_0 = \mu, \mu_T = \nu \tag{2.4.5}$$

Optimality conditions In order to simplify the exposition, I limit the discussion to the simpler situation of periodic boundary conditions, where $\Lambda = \mathbb{T}^d$ is the d -dimensional torus and there is no external field, that is $E = 0$. As a further simplification, I also assume that the mobility χ is a scalar quantity. Following a classical argument, we expect the optimal current and density for (SPLG) need to satisfy $\chi(\mu)^{-1}(j - J(t, \mu)) = \nabla H$ for some scalar field H .

But then, we can operate a change of variable and rewrite (SPLG) as the problem of minimizing

$$\frac{1}{4} \int_0^T \int_{\Lambda} |\nabla H_t(x)|^2 \chi(\mu_t(x)) \, dx dt \tag{2.4.6}$$

among all curves (μ, H) satisfying the equation

$$\partial_t \mu_t + \nabla \cdot (\chi(\mu_t)(\nabla H_t - f''(\mu)\nabla \mu)) = 0$$

and such that $\mu_0 = \mu$, $\mu_T = \nu$. A heuristic variational analysis leads to conjecture that the optimal pair solves the following HJB-FP system

$$\begin{cases} \partial_t \mu_t(x) - \nabla \cdot (D(\mu_t(x)) \nabla \mu_t(x)) + \nabla \cdot (\chi(\mu_t(x)) \nabla H_t(x)) = 0, \\ \partial_t H_t(x) + D(\mu_t(x)) \Delta H_t(x) + \frac{1}{2} |\nabla H_t(x)|^2 \chi'(\mu_t(x)) = 0, \\ \mu_0 = \mu, \mu_T = \nu. \end{cases} \quad (2.4.7)$$

The question of how to give a rigorous meaning to (2.4.7) is very intriguing. Note that under further restrictions on the coefficients and getting rid of the diffusion coefficients (i.e. setting $D \equiv 0$), this system has been solved in [48]. Let us now discuss how (2.4.7) can be brought into a form analogous to the celebrated *Schrödinger system*

$$\begin{cases} \partial_t H_t(x) + \frac{1}{2} |\nabla H_t(x)|^2 + \frac{1}{2} \Delta H_t(x) = 0, \\ -\partial_t \hat{H}_t(x) + \frac{1}{2} |\nabla \hat{H}_t(x)|^2 + \frac{1}{2} \Delta \hat{H}_t(x) = 0. \end{cases}$$

that express optimality conditions for the classical Schrödinger problem [163]. The Schrödinger system has a peculiar structure in the sense that H and \hat{H} evolve according to two HJB equations which are *identical* except for the term containing the time derivative, that changes sign from one equation to the other. The relation between the Schrödinger system and the HJB-FP system is given in the classical SP through the Hopf-Cole transform. It is natural to wonder whether such a structure is conserved in (SPLG). The answer is affirmative. Indeed introducing a suitable change of variables inspired by considerations on time reversal for diffusion processes, we can rephrase (2.4.7) as

$$\forall t, x \in [0, T] \times \Lambda, \quad \hat{H}_t(x) + H_t(x) = 2f'(\mu_t(x)). \quad (2.4.8)$$

Then we can rewrite optimality conditions in the equivalent form

$$\begin{cases} \partial_t H_t(x) + D(\mu_t(x)) \Delta H_t(x) + \frac{1}{2} |\nabla H_t(x)|^2 \chi'(\mu_t(x)) = 0, \\ -\partial_t \hat{H}_t(x) + D(\mu_t(x)) \Delta \hat{H}_t(x) + \frac{1}{2} |\nabla \hat{H}_t(x)|^2 \chi'(\mu_t(x)) = 0. \end{cases}$$

Beyond Otto calculus We now shift the attention towards Newton's law. What we want to know is: is there a Newton's law associated with (2.4.7)? Of course, the answer depends on the choice of an appropriate Riemannian metric. It is not hard to see that unless $D(\mu) = \mu$, Otto's metric is *not* the right choice and that the correct metric has to incorporate some knowledge of the mobility coefficient. In [106] the authors construct a class of distances associated with a concave mobility. This strongly suggests that the good Riemannian metric to consider is constructed defining the tangent space at μ as

$$\overline{\{\nabla \varphi : \varphi \in \mathcal{C}_c(\mathbb{R}^d)\}}^{L^2(\chi(\mu))}.$$

and the inner product on the tangent space by

$$\int_{\Lambda} \nabla \varphi \cdot \nabla \psi(x) \chi(\mu) dx$$

Accordingly, the velocity of a curve can be defined via the non linear continuity equation

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

What remains to be done is to understand and correctly define the Levi-Civita connection. Although the geodesic equations rigorously studied in [48] give some intuition on how to define parallel transport, the picture is not clear yet and a notion that verifies the fundamental axioms, i.e. compatibility with the metric and torsion is still missing. I plan to address this problem in the near future.

Hamilton Jacobi equations for controlled gradient flows

Contents

3.1	Introduction	31
3.1.1	A third encounter with the toy model	31
3.2	Gradient flows in EVI formulation and the Tataru distance	33
3.2.1	Evolutional Variational Inequality	34
3.2.2	The Tataru distance	34
3.2.3	Development of a proper lower and upper bound	34
3.3	Two comparison principles	36
3.3.1	More regular Hamiltonians	37
3.3.2	Some examples	39
3.4	Perspectives	40

In this chapter I report on the results of a joint work with D. Tonon and R. Kraaij [87] that is concerned with uniqueness of viscosity solutions for a class of Hamilton Jacobi equations on metric spaces that is intrinsically connected with an abstract version of the Schrödinger problem. The chapter is organized as follows.

- In the introductory section, I explain why this equation arises naturally working on the toy model and giving some background on Hamilton Jacobi equations on infinite dimensional spaces.
- In section 3.2, I introduce gradient flows in EVI formulation and the Tataru distance, that play a major role in establishing the main results, that are stated and commented section 3.3.

As usual, I conclude the chapter with some perspectives on future research.

3.1 Introduction

3.1.1 A third encounter with the toy model

If we relax the final constraint $x_T^u = y$ in (TSP) and replace it by a terminal cost φ we obtain

$$\inf \left\{ \frac{1}{2} \int_0^T |u_t|^2 dt + \varphi(x_T^u) : \dot{x}_t^u = -\nabla E(x_t^u) + u_t, x_0^u = x \right\}. \quad (3.1.1)$$

To solve this problem, a classical strategy in optimal control is to solve the associated Hamilton Jacobi equation

$$\begin{cases} \partial_t f_t^\varphi(x) - \frac{1}{2} |\nabla f_t^\varphi(x)|^2 - \nabla E(x) \cdot \nabla f_t^\varphi = 0, \\ f_T^\varphi \equiv \varphi \end{cases} \quad (3.1.2)$$

and then show by a verification argument that the optimally controlled state process is given by

$$\dot{x}_t^u = -\nabla E(x_t^u) - \nabla f_t^\varphi(x_t^u), \quad x_0^u = x.$$

Equation (3.1.2) is in duality with (TSP) in the sense that

$$C_T(x, y) - 2E(x) = \sup_{\varphi \in \mathcal{C}_b(\mathbb{R}^d)} f_0^\varphi(x) - \varphi(y), \quad (3.1.3)$$

where I recall that $C_T(x, y)$ is the optimal value in (TSP). Note that, assuming $E = 0$ and integrating in μ and ν respectively the two summands in (3.1.3) we indeed obtain the dual of the Monge-Kantorovich problem: adding the term $-\frac{1}{2}\Delta f^\varphi$ in (3.1.2) yields one of the many formulation of the dual Schrödinger problem, see [137]. It appears that, as soon as the underlying particle system is not made of independent particles, in order to obtain a duality result for the corresponding entropic cost, working with finite dimensional Hamilton-Jacobi equations is not enough. The "right" HJ equation to consider has to be set on the space of probability measures and, once reinterpreted through Otto calculus, it has the geometric structure of (3.1.2). Of course, in the infinite dimensional framework existence and uniqueness of viscosity solutions becomes a much more delicate matter. In this chapter, I present the main ideas and results of [87] where, together with Daniela Tonon and Richard Kraaij, we have developed a method to prove the comparison principle and show uniqueness of solutions. In fact, we have proven two comparison principles. The first one is for a Hamiltonian whose domain is very small, but contains some irregular objects, namely the *Tataru distance* see section 3.2.2. The second comparison principle is for an Hamiltonian whose domain is constituted by smooth test functions and is therefore a much more tractable object for building an existence theory. The toy version of the equation we considered reads as

$$f + \lambda Hf = h, \quad Hf := -\langle \nabla f, \nabla E \rangle + \frac{1}{2}|\nabla f|^2 \quad (3.1.4)$$

and is meant to characterize the value function of the ergodic control problem

$$\sup \left\{ \int_0^{+\infty} e^{-\lambda t} \left(h(x^u(t)) - \frac{1}{2}|u(t)|^2 \right) dt : \dot{x}_t^u = -\nabla E(x_t^u) + u_t, x_0^u = x \right\}$$

Besides a harmless sign change, the main difference between the above control problem and the toy model is that instead of having a terminal cost, we have a running cost $e^{-\lambda t}h(\cdot)$, which we integrate over $[0, +\infty]$. When looking at the corresponding HJ equations, this means dropping the term containing the time derivative as well as the boundary condition and gaining another term that corresponds to the running cost. There are two reasons why we studied this equation: the first one is, as said above, that (3.1.4) is a proxy for (3.1.2), that is slightly simpler to analyse. The second is that well posedness for this equation is key for the succesful application of the techniques developed by Feng and coauthors to establish large deviations principles, see the monograph [120]. As I will illustrate below, our results largely expand the range of applicability of this approach. Essentially, our comparison principle works as soon as the energy functional E admits a gradient flow and is geodesically κ -convex, with $\kappa \in \mathbb{R}$. In fact, these two assumptions are both valid if we assume that \mathcal{E} admits a gradient flow in *EVI* formulation. This powerful notion introduced in [5], (also Definition 3.2.1 below) has far reaching consequences [177] that play a decisive role in the proofs of our comparison principles. As a guiding example to better understand what follows, the reader can keep in mind the case when equation (3.1.4) is set on the metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ and the energy functional, which we shall always denote by \mathcal{E} from now on, is the free energy functional $\tilde{\mathcal{F}}$ defined at (2.3.5) in the context of the mean field Schrödinger problem. I will provide more interesting examples of applications of our results in section 3.3.2.

Hamilton Jacobi equations in infinite dimensional spaces The theory of viscosity solutions for Hamilton Jacobi equations on infinite dimensional spaces was initiated by Crandall and Lions in a series of papers [99] in the setting of Hilbert spaces or Banach spaces possessing the Radon-Nykodim property. Recent applications in large deviations [122], functional inequalities [141], statistical mechanics [30, 31], and McKean-Vlasov control [58] have motivated the development of a theory of viscosity solutions for Hamilton Jacobi equations on metric spaces that are not necessarily Hilbert, and in particular over the space of probability measures endowed with a transport-like distance. A first approach to HJ equations on metric spaces exploits the possibility of lifting the space of probability distributions to the space of square integrable random variables in order to take advantage of the Hilbertian structure of the latter: we refer to [14, 190, 28] for some results recently obtained following this method. A second approach is

more intrinsic and consists of working directly on the space of probability measures and develop all the relevant notions therein. One can do so using either the linear derivative, as shown in [43] in the context of McKean-Vlasov control for jump processes, or relying on the notion of subdifferential provided by optimal transport [5]. The connections between the intrinsic approach and the extrinsic notion of derivative obtained through the above mentioned lifting procedure have been clarified in [134]. In our work followed the intrinsic approach and in particular we build on the achievements of the research program carried out by Feng and his coauthors [122, 121, 120, 123], who developed *ad hoc* technique to deal with equations whose geometric structure is the same as (3.1.4): I shall discuss in more details the relation between our contribution and these works in the next section. Several other important contributions [2, 132, 133, 134, 212] adopt the intrinsic approach to show well posedness of Hamilton Jacobi equations on metric spaces. In all these works it is assumed that the variations of the Hamiltonian with respect to the measure argument can be locally controlled by the metric d in some way. Since we require very little from the energy functional \mathcal{E} beyond the existence of an EVI gradient flow, this assumption is systematically violated in most instances of (3.1.4) that we consider. This happens already in the basic example when \mathcal{E} is the relative entropy. It is worth mentioning that operating the formal change of variable $\tilde{f} = f - \mathcal{E}$ and setting $\lambda = 1$ allows to rewrite formally (3.1.4) in the form

$$f - \frac{1}{2}|\text{grad}f|^2 + \mathcal{F} = 0, \quad (3.1.5)$$

$\mathcal{F}(\cdot) = \frac{1}{2}|\text{grad}\mathcal{E}|^2(\cdot) + \mathcal{E}(\cdot) - h(\cdot)$. This equation has been often studied in the literature on infinite dimensional HJ equations. However, our main geometrical assumption, that is formally equivalent to the semiconvexity of \mathcal{E} , does not give the control on the growth of $\frac{1}{2}|\text{grad}\mathcal{E}|^2(\cdot)$ needed to successfully apply the techniques developed in the above mentioned references. Our proof of the comparison principle combines some rather classical ingredients such as Ekeland's perturbed optimization principle and the duplication of variables technique, which becomes here a quadruplication to get around the fact that \mathcal{E} may well be infinite, together with some more original ones. A first one worth mentioning is the use of the Tataru distance ([203, 204]) as a penalization function in Ekeland's principle, that allows us remove compactness assumptions both for the level sets of the energy functional \mathcal{E} and for metric balls: applications to the infinite dimensional setting seem to be limited to Hilbert spaces [98, 120]. A second original element of our proof strategy is the systematic use of the properties of EVI gradient flows, in particular of their regularizing properties that include energy dissipation and distance contraction estimates. Indeed, gradient flows play an crucial role in *a)* Defining suitable upper and lower bounds for the formal Hamiltonian that depend on \mathcal{E} and d only *b)* the construction of the Tataru distance and *c)* developing all the necessary estimates for the proof of the main result, in particular to bound the difference of the Hamiltonians in the proof of the comparison principle and for carrying out the approximation steps through which we transfer the comparison principle from the less regular Hamiltonians to the more regular Hamiltonians, see Definitions 3.2.2 and 3.3.2 below.

Master equation and mean field games The recent years have witnessed fundamental advances on the understanding of the master equation arising in the theory of mean field games, see [50] and the works [212, 130, 131, 49] for a sample of the recent progresses. Such equation aims at characterizing the limiting behavior of Nash equilibria in the many players regime. It has been noticed in [27] that the master equation shares some properties with infinite dimensional Hamilton-Jacobi equations, and in particular with those characterizing the value function of McKean-Vlasov control problems. However, these two equations remain conceptually different as explained [60]. For example, despite some analogies between the "monotonicity" assumption that is typically imposed on the coefficients of the master equation and the geodesic convexity of the energy functional \mathcal{E} that underlies all our computations, these two geometrical assumptions are not directly related and enter the respective equations in a different way. In the recent article [130], the authors get past the classical monotonicity assumption and indeed obtain well posedness for the master by means of displacement convexity. Still, the equation considered there and (3.1.4) are of different nature.

3.2 Gradient flows in EVI formulation and the Tataru distance

I now fix a complete metric space (E, d) and a lowersemicontinuous functional \mathcal{E} and discuss EVI gradient flows and the Tataru distance.

3.2.1 Evolutional Variational Inequality

The theory of gradient flows on metric spaces has known an impressive development in the last two decades. In order to get around the lack of differentiability of energy functionals, various formulations of the gradient flow equation have been proposed. Most of them are reviewed in the monograph [5]. Let me define the EVI formulation of gradient flows.

Definition 3.2.1. *Given $\kappa \in \mathbb{R}$, we say that a solution of the (\mathbf{EVI}_κ) inequality is a continuous curve $\gamma : [0, +\infty) \rightarrow \mathcal{D}(\mathcal{E})$ such that*

$$\frac{1}{2} \frac{d^+}{dt} (d^2(\gamma(t), \rho)) \leq \mathcal{E}(\rho) - \mathcal{E}(\gamma(t)) - \frac{\kappa}{2} d^2(\gamma(t), \rho), \quad \forall \rho \in \mathcal{D}(\mathcal{E}), t \in [0, +\infty). \quad (\mathbf{EVI}_\kappa)$$

Here $\frac{d^+}{dt}$ denotes the upper right time derivative.

An (\mathbf{EVI}_κ) gradient flow of \mathcal{E} defined in $D \subset \overline{\mathcal{D}(\mathcal{E})}$ is a family of continuous maps $S(t) : D \rightarrow D, t \geq 0$ such that for every $\pi \in D$:

- it holds

$$S[\pi](0) = \pi, \quad S[\pi](t+h) = S[S[\pi](t)](h) \quad \forall t, h \geq 0, \quad (3.2.1)$$

- The curve $t \mapsto S[\pi](t)$ is a solution to (\mathbf{EVI}_κ) .

The curve $(S[\pi](t))_{t \geq 0}$ is the gradient flow of \mathcal{E} started at π . To lighten the notation, from now on, we will denote with $(\pi(t))_{t \geq 0}$ the gradient flow $(S[\pi](t))_{t \geq 0}$.

For our purposes, we found it convenient to work with (\mathbf{EVI}_κ) gradient flows. This is not surprising: the (\mathbf{EVI}_κ) formulation is the most demanding one and therefore has deeper implications in terms of dissipation estimates and regularizing effects [177].

3.2.2 The Tataru distance

The standard procedure to prove the comparison principle consists in using a doubling variables method. However, when doing this in our context, we run into the known issue that optimizers are not attained, essentially because we are working on an infinite dimensional space. This issue is usually solved via Ekeland's variational principle. Nevertheless, for our setting, in which the Hamiltonian contains an unbounded term, choosing the metric $d(\cdot, \cdot)$ as penalization function in Ekeland's principle does not allow to replicate the classical proofs of the comparison principle, essentially because we do not gain enough control on the difference of the Hamiltonians. Following [98, 203, 204, 120], we need to apply Ekeland variational principle with the *Tataru distance* $d_T(\cdot, \cdot)$ as a penalization function, and the reason why we do so is that, in contrast with the original distance, $d_T(\cdot, \cdot)$ has the property of being Lipschitz along the gradient flow. More specifically, the definition of $d_T(\pi, \rho)$ we used, that is slightly different from the original one [203] is the following:

$$d_T(\pi, \rho) = \inf_{t \geq 0} \{t + d(\pi, \exp(\hat{\kappa}t)\rho(t))\}, \quad \forall \pi, \rho \in E,$$

where $\rho(\cdot)$ is the gradient flow and $\hat{\kappa} = \min\{\kappa, 0\}$. Note that d_T is not symmetric in its arguments and therefore is not a metric. The announced Lipschitzianity property e is its Lipschitzianity along the gradient flow reads as follows:

$$\forall \gamma, \rho \in E, \quad |d_T(\gamma(t), \rho) - d_T(\gamma(0), \rho)| \leq t. \quad (3.2.2)$$

It seems that, prior to our work, the use of the Tataru distance in the context of infinite dimensional Hamilton Jacobi equations was limited to Hilbert spaces [98, 120].

3.2.3 Development of a proper lower and upper bound

Let me rewrite (3.1.4) in more abstract terms

$$f + \lambda Hf = h, \quad Hf := -\langle \text{grad} f, \text{grad} \mathcal{E} \rangle + \frac{1}{2} |\text{grad} f|^2, \quad (3.2.3)$$

where I used the notation grad instead of ∇ to emphasize that we are working on a (so far unspecified) abstract Riemannian manifold. The first problem to overcome is that of finding a way to make sense of all the terms appearing in (3.2.3). To treat the term $|\text{grad}f|^2$ we can use the metric notion of *local slope* [5]. However, it still remains to understand how to treat the term $-\langle \text{grad}\mathcal{E}, \text{grad}f \rangle$. Note that if we were working on a smooth finite dimensional manifold and if \mathcal{E} was smooth, we would write

$$-\langle \text{grad}\mathcal{E}, \text{grad}f \rangle(\rho) = \left. \frac{d}{dt} f(\rho(t)) \right|_{t=0}. \quad (3.2.4)$$

Unfortunately, the derivative of f along the gradient flow may not be always well defined. This is why we chose work with upper and lower bounds for H as in [2] for example. If we consider a test function of the form

$$f^\dagger(\pi) = \frac{1}{2}ad^2(\pi, \rho) + bd_T(\pi, \mu) + c \quad (3.2.5)$$

where $a, b > 0$, $c \in \mathbb{R}$ and $\rho, \mu \in E$ with $\mathcal{E}(\rho) < +\infty$ then, arguing on the basis of (3.2.4), (EVI $_\kappa$) and using the properties of the Tataru distance such as (3.2.2) we obtain, the following formal upper bound

$$Hf^\dagger(\pi) \leq a[\mathcal{E}(\rho) - \mathcal{E}(\pi)] - a\frac{\kappa}{2}d^2(\pi, \rho) + b + \frac{1}{2}a^2d^2(\pi, \rho) + abd(\pi, \rho) + \frac{1}{2}b^2.$$

The main advantage of working with the upper bound is that, unlike the formal Hamiltonian H it is a well defined functional and can be used as a starting point to give a rigorous meaning to our HJ equation. An analogous calculation can be used to produce a lower bound, leading to the following definition. Of course, the delicate part in working with upper and lower bounds is that they have to be tight enough for showing uniqueness of viscosity solutions.

Definition 3.2.2. 1. For each $a > 0, b > 0, c \in \mathbb{R}$, and $\mu, \rho \in E : \mathcal{E}(\rho) < \infty$ let f^\dagger and g^\dagger be given for any $\pi \in E$ by

$$\begin{aligned} f^\dagger(\pi) &:= \frac{1}{2}ad^2(\pi, \rho) + bd_T(\pi, \mu) + c, \\ g^\dagger(\pi) &:= a[\mathcal{E}(\rho) - \mathcal{E}(\pi)] - a\frac{\kappa}{2}d^2(\pi, \rho) + b + \frac{1}{2}a^2d^2(\pi, \rho) + abd(\pi, \rho) + \frac{1}{2}b^2. \end{aligned}$$

Then the operator $\tilde{H}_\dagger \subseteq C_b(E) \times USC(E)$ is defined by

$$\tilde{H}_\dagger := \{(f^\dagger, g^\dagger) \mid a > 0, b > 0, c \in \mathbb{R}, \mu, \rho \in E : \mathcal{E}(\rho) < \infty\}.$$

2. For each $a > 0, b > 0, c \in \mathbb{R}$, and $\pi, \gamma \in E : \mathcal{E}(\gamma) < \infty$ let f^\ddagger and g^\ddagger be given for any $\mu \in E$ by

$$\begin{aligned} f^\ddagger(\mu) &:= -\frac{1}{2}ad^2(\gamma, \mu) - bd_T(\mu, \pi) + c \\ g^\ddagger(\mu) &:= a[\mathcal{E}(\mu) - \mathcal{E}(\gamma)] + a\frac{\kappa}{2}d^2(\gamma, \mu) - b + \frac{1}{2}a^2d^2(\gamma, \mu) - abd(\gamma, \mu). \end{aligned}$$

Then the operator $\tilde{H}_\ddagger \subseteq C_b(E) \times LSC(E)$ is defined by

$$\tilde{H}_\ddagger := \{(f^\ddagger, g^\ddagger) \mid a > 0, b > 0, c \in \mathbb{R}, \pi, \gamma \in E : \mathcal{E}(\gamma) < \infty\}.$$

The definition of viscosity solution we used is the following

Definition 3.2.3. Fix $\lambda > 0$ and $h^\dagger, h^\ddagger \in C_b(E)$. Consider the equations

$$f - \lambda A_\dagger f = h^\dagger, \quad (3.2.6)$$

$$f - \lambda A_\ddagger f = h^\ddagger. \quad (3.2.7)$$

We say that u is a (viscosity) subsolution of equation (3.2.6) if u is bounded, upper semi-continuous and if for all $(f, g) \in A_{\dagger}$ there exists a sequence $(\pi_n)_{n \in \mathbb{N}} \in E$ such that¹

$$\lim_{n \uparrow \infty} u(\pi_n) - f(\pi_n) = \sup_{\pi} u(\pi) - f(\pi), \quad (3.2.8)$$

$$\lim_{n \uparrow \infty} u(\pi_n) - \lambda g(\pi_n) - h^{\dagger}(\pi_n) \leq 0. \quad (3.2.9)$$

We say that v is a (viscosity) supersolution of equation (3.2.7) if v is bounded, lower semi-continuous and if for all $(f, g) \in A_{\ddagger}$ there exists a sequence $(\pi_n)_{n \in \mathbb{N}} \in E$ such that

$$\lim_{n \uparrow \infty} v(\pi_n) - f(\pi_n) = \inf_{\pi} v(\pi) - f(\pi),$$

$$\lim_{n \uparrow \infty} v(\pi_n) - \lambda g(\pi_n) - h^{\ddagger}(\pi_n) \geq 0.$$

If $h^{\dagger} = h^{\ddagger}$, we say that u is a (viscosity) solution of equations (3.2.6) and (3.2.7) if it is both a subsolution of (3.2.6) and a supersolution of (3.2.7).

We say that (3.2.6) and (3.2.7) satisfy the comparison principle if for every subsolution u to (3.2.6) and supersolution v to (3.2.7), we have $\sup_E u - v \leq \sup_E h^{\dagger} - h^{\ddagger}$.

Let me conclude this section what follows I denote by I the squared metric slope of \mathcal{E} , i.e.

$$\sqrt{I}(\pi) = \lim_{\rho \rightarrow \pi} \frac{(\mathcal{E}(\pi) - \mathcal{E}(\rho))_+}{d(\rho, \pi)}.$$

To fix ideas, when \mathcal{E} is $\mathcal{H}(\cdot|\text{vol})$, the functional I is the Fisher information (2.2.19).

3.3 Two comparison principles

We obtained two comparison principles. The first is in terms of the upper and lower bounds $\tilde{H}_{\dagger}, \tilde{H}_{\ddagger}$ and holds under the following set of assumptions.

Assumption 3.3.1. We make the following assumptions of the metric space (E, d) and the energy functional \mathcal{E} .

1. (E, d) is a geodesic space, i.e. for any $\rho, \pi \in E$ there exists a curve $(\gamma_{\rho}^{\pi}(t))_{t \in [0,1]}$ such that $\gamma_{\rho}^{\pi}(0) = \rho, \gamma_{\rho}^{\pi}(1) = \pi$ and for all $s, t \in [0, 1]$

$$d(\gamma_{\rho}^{\pi}(s), \gamma_{\rho}^{\pi}(t)) = |t - s|d(\rho, \pi). \quad (3.3.1)$$

Such a curve will be called geodesic.

2. We assume that the energy functional $\mathcal{E} : E \rightarrow (-\infty, +\infty]$ is an extended functional such that:
 - It has a proper effective domain, i.e. $\mathcal{D}(\mathcal{E}) := \{\pi \in E : \mathcal{E}(\pi) < +\infty\} \neq \emptyset$.
 - It is lower semi-continuous

3. There exists $\kappa \in \mathbb{R}$ such that there exists an **(EVI) $_{\kappa}$** gradient flow of \mathcal{E} defined on E .

4. For any $\rho, \pi \in E$ satisfying $I(\rho) + \mathcal{E}(\pi) < +\infty$, there exist a geodesic γ_{ρ}^{π} such that, for any $\theta > 0$, there exists a curve $(\rho^{\theta}(t))_{t \in [0, \tau]}$ satisfying

$$\limsup_{t \downarrow 0} \frac{d(\rho^{\theta}(t), \gamma_{\rho}^{\pi}(t))}{t} \leq \theta, \quad (3.3.2)$$

and

$$\limsup_{t \downarrow 0} \frac{\mathcal{E}(\rho^{\theta}(t)) - \mathcal{E}(\rho)}{t} \leq \sqrt{I}(\rho)(d(\rho, \pi) + \theta). \quad (3.3.3)$$

We refer to (3.3.2) as to the angle condition.

¹In classical works on viscosity solutions, instead of working with the statement "there exists a sequence such that", one has "for all optimizers one has". Even though the classical stronger definition has advantages when proving the comparison principle, the weaker definition allows for easier approximation arguments.

Let me clarify a bit the content of item 4. In there, we ask that any geodesic can be approximated as well as needed with a smoother curve, typically another geodesic, along which the variations of \mathcal{E} can be controlled with the slope. This last requirement is coherent with the interpretation of the metric slope as the norm of the gradient of \mathcal{E} . Note that in most examples of interest (3.3.3) fails to be true if we replace $\rho^\theta(t)$ with an arbitrary geodesic and that item 4 does not imply the differentiability of \mathcal{E} . Here is our first comparison principle.

Theorem 3.3.2. *Let Assumption 3.3.1. Moreover, let $\lambda > 0$ and $h^\dagger, h^\ddagger : E \rightarrow \mathbb{R}$ be bounded and uniformly continuous. Let $u : E \rightarrow \mathbb{R}$ be a viscosity subsolution to $f - \lambda \tilde{H}_\dagger f = h^\dagger$ and let $v : E \rightarrow \mathbb{R}$ be a viscosity supersolution to $f - \lambda \tilde{H}_\ddagger f = h^\ddagger$. Then we have*

$$\sup_{\pi} u(\pi) - v(\pi) \leq \sup_{\pi} h^\dagger(\pi) - h^\ddagger(\pi).$$

Note that we formally have

$$Hf \leq \tilde{H}_\dagger f \quad \text{and} \quad \tilde{H}_\ddagger f \leq Hf.$$

Thanks to these inequalities the above result will give a formal comparison principle for equation (3.2.3).

3.3.1 More regular Hamiltonians

As already said, the Hamiltonians $\tilde{H}_\dagger, \tilde{H}_\ddagger$ are not the most convenient to use to prove existence of viscosity solutions as one may want to work with Hamiltonians whose domain contains smooth functions of distance-like objects. This motivated us to establish a viscosity push-over result (Theorem 3.3.4) that enables to transfer viscosity subsolutions (resp. supersolutions) from the regular Hamiltonian H_\dagger (resp. H_\ddagger), that I am going to define at Definition 3.3.1, to \tilde{H}_\dagger (resp. \tilde{H}_\ddagger). As a corollary, the comparison principle holds for the regular Hamiltonians. To state the transfer result, we need a further topological assumption: we ask that there exist a topology, weaker than the topology generated by the metric d , for which the intersection of metric balls with the level sets of \mathcal{E} are compact sets. Of course, this does not mean that such a set is compact in the stronger topology.

Assumption 3.3.3 (Topological assumptions). *We assume that there exists a topology on E that is weaker than the topology generated by d and such that*

- *The metric $d(\cdot, \cdot)$ is weakly lower semi-continuous. The energy functional \mathcal{E} is weakly lower semi-continuous on metric balls.*
- *For all $\rho \in E$ and $c, d \in \mathbb{R}$, the set*

$$K_{c,d}^\rho := \{\pi \in E : d(\rho, \pi) \leq c, \mathcal{E}(\pi) \leq d\} \quad (3.3.4)$$

is weakly compact.

Note that this assumption is satisfied in the reference example $(\mathcal{P}_2(\mathbb{R}^d), W_2(\cdot, \cdot))$ equipped with the free energy functional $\tilde{\mathcal{F}}$ of Definition (2.3.5) and this is essentially because balls in the W_2 -topology are compact in the W_p -topology for any $0 < p < 2$. Let me now proceed to define the regular Hamiltonians, acting on smooth test functions of squared distance objects. In order to understand the logic behind the following definition one can argue on the basis of (EVI $_\kappa$) and (3.2.4) to check that the proposed Hamiltonian is an upper bound (resp. lower bound) of the formal Hamiltonian H defined at (3.1.4). In the next definition, I use the set

$$\mathcal{T}_b = \{\varphi \in \mathcal{C}_\infty^b((0, \infty)^{k+1}; \mathbb{R}) : k \in \mathbb{N}, \partial_i \varphi \geq 0 \forall i\}.$$

Definition 3.3.1. *Let $\varphi \in \mathcal{T}_b$. Moreover, let $\boldsymbol{\rho} = (\rho_0, \dots, \rho_k) \in E^{k+1}$ be such that $I(\rho_i) < \infty$ for all i . We define*

$$f^\dagger(\pi) := \varphi\left(\frac{1}{2}d^2(\pi, \boldsymbol{\rho})\right) := \varphi\left(\frac{1}{2}d^2(\pi, \rho_0), \dots, \frac{1}{2}d^2(\pi, \rho_k)\right), \quad (3.3.5a)$$

$$g^\dagger(\pi) := \sum_{i=0}^k \partial_i \varphi\left(\frac{1}{2}d^2(\pi, \boldsymbol{\rho})\right) \left[\mathcal{E}(\rho_i) - \mathcal{E}(\pi) - \frac{\kappa}{2}d^2(\pi, \rho_i) \right] \quad (3.3.5b)$$

$$+ \frac{1}{2} \left(\sum_{i=0}^k \partial_i \varphi\left(\frac{1}{2}d^2(\pi, \boldsymbol{\rho})\right) d(\pi, \rho_i) \right)^2.$$

and define H_{\dagger} by

$$H_{\dagger} := \{(f^{\dagger}, g^{\dagger}) \mid \varphi \in \mathcal{T}_b, I(\rho_i) < +\infty \forall i\}.$$

Arguing similarly, if $\gamma = (\gamma_0, \dots, \gamma_k) \in E^{k+1}$ is such that $I(\gamma_i) < \infty$ for all i , we consider

$$f^{\ddagger}(\mu) := -\varphi\left(\frac{1}{2}d^2(\gamma, \mu)\right), \quad (3.3.6a)$$

$$\begin{aligned} g^{\ddagger}(\mu) := & \sum_{i=0}^k \partial_i \varphi\left(\frac{1}{2}d^2(\gamma, \mu)\right) \left[\mathcal{E}(\mu) - \mathcal{E}(\gamma_i) + \frac{\kappa}{2}d^2(\gamma_i, \mu) \right] \\ & + \frac{1}{2} \left(\sum_{i=0}^k \partial_i \varphi\left(\frac{1}{2}d^2(\gamma, \mu)\right) d(\gamma_i, \mu) \right)^2. \end{aligned} \quad (3.3.6b)$$

and define H_{\ddagger} by

$$H_{\ddagger} := \{(f^{\ddagger}, g^{\ddagger}) \mid \varphi \in \mathcal{T}_b, I(\gamma_i) < +\infty \forall i\}.$$

Here is the announced transfer result.

Theorem 3.3.4. *Let Assumption 3.3.1 and Assumption 3.3.3 hold. Moreover, let $\lambda > 0$ and h be weakly continuous.*

Let u be a viscosity subsolution to $f - \lambda H_{\dagger} f = h$. Then the weak upper semi-continuous regularization \hat{u} of u is a viscosity subsolution to $f - \lambda \tilde{H}_{\dagger} f = h$.

Similarly, let v be a viscosity supersolution to $f - \lambda H_{\dagger} f = h$. Then the weak lower semi-continuous regularization \hat{v} of v is a viscosity supersolution to $f - \lambda \tilde{H}_{\dagger} f = h$.

As a corollary: the comparison principle holds for H_{\dagger} and H_{\ddagger} .

The proof of Theorem 3.3.4 is composed of seven steps. At each but the last step we defined a new Hamiltonian $H_{i,\dagger}$, $i = 1, \dots, 6$ and prove a statement that, loosely speaking looks like

$$u \text{ subsolution of } u - \lambda H_{i,\dagger} u = h \quad \Rightarrow \quad u \text{ subsolution of } u - \lambda H_{i+1,\dagger} u = h$$

with the understanding that $H_{0,\dagger} = H_{\dagger}$ and $H_{7,\dagger} = \tilde{H}_{\dagger}$. As i increases, the domain of $H_{i,\dagger}$ includes objects that are more and more closer to the Tataru distance. Once again, all our proofs are built around the dissipation estimates and regularization effects that come with gradient flows in (\mathbf{EVI}_{κ}) form. Since the proof of certain steps is quite technical and delicate, giving a sketch of the proof of Theorem 3.3.4 is beyond the scope of this presentation. However, for the curious reader I report here a table that summarizes in what consists each of the seven steps.

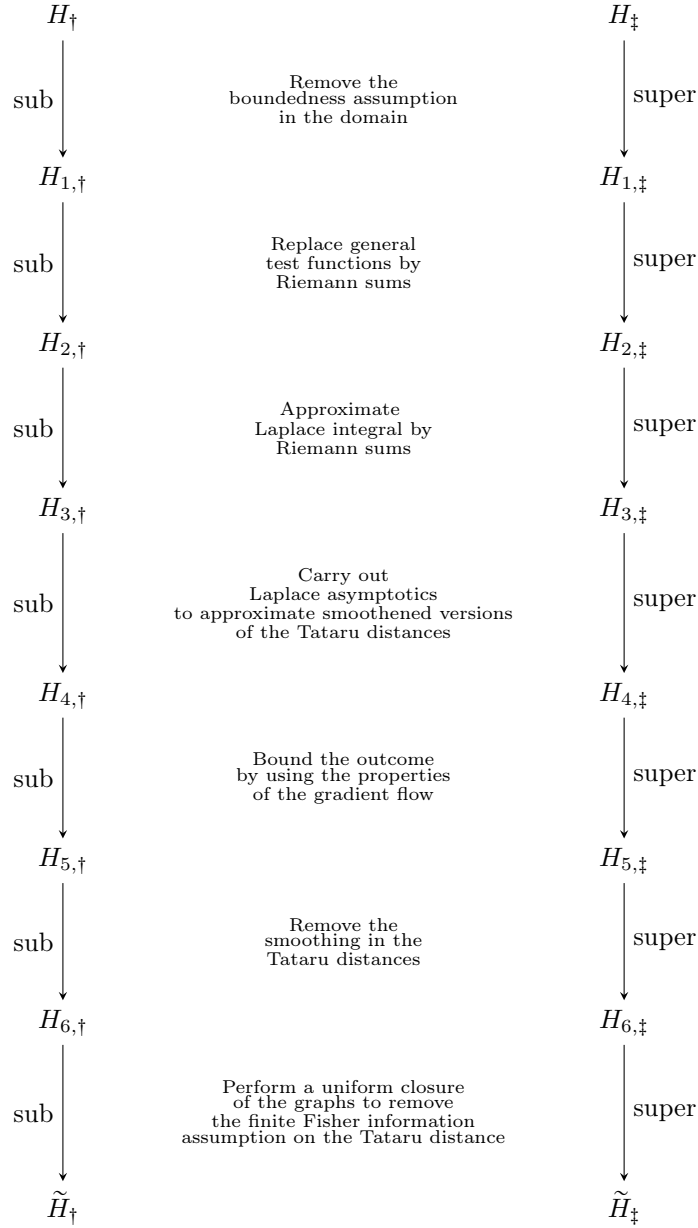


Figure 3.1 – Carrying over viscosity sub- and supersolutions.

In this diagram, an arrow connecting an operator A with operator B with subscript 'sub' means that viscosity subsolutions of $f - \lambda A f = h$ are also viscosity subsolutions of $f - \lambda B f = h$. Similarly for arrows with a subscript 'super'.

3.3.2 Some examples

Let me present the macro-example that motivated our study. The metric space under consideration is $E = \mathcal{P}_2(\mathbb{R}^d)$ that I equip with the Kantorovich-Wasserstein distance $W_2(\cdot, \cdot)$ of order two. For given functions $F : \mathbb{R}_+ \rightarrow \mathbb{R}$, $V : \mathbb{R}^d \rightarrow \mathbb{R}$, $W : \mathbb{R}^d \rightarrow \mathbb{R}$, I consider the energy functional \mathcal{E} defined by

$$\mathcal{E}(\rho) := \int F\left(\frac{d\rho}{d\text{vol}}(x)\right) dx + \int V(x)\rho(dx) + \frac{1}{2} \int W(x-y)\rho(dx) \otimes \rho(dy), \tag{3.3.7}$$

setting $\mathcal{E}(\rho) = +\infty$ as soon as ρ is not absolutely continuous w.r.t the Lebesgue measure **vol**. The celebrated *McCann condition*, introduced in [171] is a set of assumptions of F, W, V ensuring that \mathcal{E} is κ -convex along geodesics. In particular, this condition holds for the model example (2.3.5) as soon as the Hessian of the interaction potential W is uniformly bounded below from a constant, that could possibly be negative. Other interesting examples can be obtained replacing Boltzmann's entropy, that corresponds to $F(r) = r \log r$, with another internal energy functional such as those typically encountered in the study of porous media equations, i.e. of the form $F(r) = \frac{r^\alpha - r}{\alpha - 1}$ with $\alpha \geq 1 - 1/d$.

Theorem 3.3.5. *Let $(E, d) = (\mathcal{P}_2(\mathbb{R}^d), W_2(\cdot, \cdot))$, and \mathcal{E} be defined by (3.3.7) with F, V, W satisfying the McCann condition. Then the conclusion of Theorem 3.3.2 holds. Moreover, if there exist $0 < p < 2$ such that*

$$\min\{V, W\}(x) \geq -a - b|x|^p \text{ for some } a, b > 0 \text{ and all } x \in \mathbb{R}^d, \quad (3.3.8)$$

then the conclusion of Theorem 3.3.4 also holds.

It is worth noting that even the basic case when \mathcal{E} is the classical Boltzmann entropy on $\mathcal{P}_2(\mathbb{R}^d)$ is not covered by the findings of [121] since neither metric balls nor the sublevels of \mathcal{E} are compact. At the level of the Schrödinger problem, this corresponds to taking a system of independent Brownian particles. Other examples we discuss in the article include controlled Cahn-Hilliard and Allen-Cahn equations. In these case, the underlying space is an Hilbert space and the methods of [120] also apply.

3.4 Perspectives

Lattice gases It is likely that our results cover some instances of (**SPLG**), and in particular (**SESP**). In that case the metric to consider is the one introduced in [106] with mobility $\rho(1 - \rho)$ and the energy functional \mathcal{E} is given by

$$\mathcal{E}(\rho) = \int_{\mathbb{T}^d} \rho \log \rho + (1 - \rho) \log(1 - \rho) \, d\text{vol}$$

In fact, gradient flows in EVI form with $\kappa = 0$ have been constructed in [62] for a general functional satisfying what the authors call the *generalized McCann's condition*. This is a promising sign telling that our main results could apply this interesting case, although we still need to check all details.

Existence of solutions and dynamic HJ equations Although showing existence of solutions for the smoother Hamiltonians is arguably a simpler task than the comparison principle, at least for some of the main examples, it still remains to be done. In particular, what we would like to do is to prove an existence result that is valid in an abstract setting, hopefully under the exact same assumptions needed for the proof of the comparison principle. Finally, in order to obtain duality results for Schrödinger problems analogous to the "toy" result (3.1.3), we would like to reintroduce a $\partial_t u$ term in the equation, thus going back to solving an abstract version of (3.1.2).

McKean–Vlasov control In fact, the equation we look at includes as a special case a large class McKean–Vlasov control problems. Given the interest surrounding this emerging research field it would very interesting to see whether the techniques we developed here can be applied to tackle general instances of the McKean–Vlasov control problem. From that perspective, the boundedness assumption we made on the running cost h can be seen as a limitation. On the other hand, in that context one typically works with a much less general energy functional \mathcal{E} .

Turnpike property and functional inequalities

Contents

4.1	Fourth encounter with the toy model	42
4.2	Convexity of the entropy along Schrödinger bridges	43
4.3	The total energy and long time behavior of the entropic cost	46
4.3.1	The conserved total energy and the energy-transport inequality	46
4.3.2	Long time behavior of the entropic cost	47
4.3.3	Short-time behavior of the entropic cost	47
4.4	Entropic turnpike estimates	48
4.4.1	Classical Schrödinger problem	48
4.4.1.1	Curvature dimension condition	48
4.4.1.2	Variational interpretation of local logarithmic Sobolev inequalities	50
4.4.2	Mean Field Schrödinger problem	51
4.4.2.1	Exponential convergence to equilibrium and the turnpike property	51
4.4.3	Kinetic Schrödinger problem	52
4.5	Turnpike estimates in Wasserstein distance for non uniformly convex potentials	53
4.6	Perspectives	56

Let us revisit once more Schrödinger's thought experiment. To keep things simple we consider independent particles, and we assume each particle to follow the law of a stationary Ornstein-Uhlenbeck process. Thus, we are looking at

$$\begin{cases} dX_t^i = -X_t^i dt + dB_t^i, \\ X_0^i \sim \mathbf{m}, \quad i = 1, \dots, N, \end{cases}$$

where $\mathbf{m}(dx) \propto \exp(-|x|^2)dx$ is the invariant measure. The general problem we are interested in is to understand the most likely evolution of the particle system conditionally on the observations $\{\mu_0^N \approx \mu, \mu_T^N \approx \nu\}$ of the empirical configuration at times 0 and T . Our intuition suggests that if T and N are large, an efficient way for the empirical flow $(\mu_t^N)_{t \in [0, T]}$ to travel from μ to ν is to remain for the most time in close proximity of the invariant measure \mathbf{m} because this is the natural tendency of the particle system. To make this just a bit more precise we could say that, because of the ergodic properties of the Ornstein-Uhlenbeck process, in absence of external observations we have the following

(a) $\mu_0^N, \mu_{T/2}^N, \mu_T^N$ are three almost independent random probability measures

(b) $\mu_{T/2}^N$ is found with high probability in an (exponentially) small neighborhood of \mathbf{m}

But then, because of item (a), property (b) should hold even *conditionally* on $\{\mu_0^N = \mu, \mu_T^N = \nu\}$. Letting $N \rightarrow +\infty$ and transferring these observations to one of the possible control formulations of the Schrödinger problem means that we conjecture that Schrödinger bridges satisfy the *turnpike property*, as I am going to explain in the next paragraph. Over the past three years I have been very much interested in establishing the turnpike property for Schrödinger bridges in a quantitative form.

Turnpike property The turnpike property stipulates that optimal solutions of dynamic control problems are made of three pieces: a rapid transition from the initial state to the steady state, the *turnpike*, then a long stationary phase localized around the turnpike, and finally another rapid transition to reach the final state. The first turnpike theorems have been established in the 60's for problems arising in econometry [172]. The study of the turnpike property in deterministic control is an active field of research: results for deterministic finite dimensional problems are by now available, see [206] for finite non-linear control, [118] for applications to machine learning. I further refer to the monographs [215, 216] for an extended treatment of the subject. The understanding of this phenomenon in stochastic control seems to be much more limited with some notable exceptions. In particular, there are some works in the literature dealing with the long time behavior of Hamilton Jacobi Bellman equations, see e.g. [97]. Finally, the McKean-Vlasov control formulation of (MFSP) connects our findings with the study of the long time behavior of mean field games [51, 52, 46, 53]. However, the equations we are looking at are quite different. A fundamental difference is that the coupling terms in the version of the mean field Schrödinger problem we considered are not monotone in the sense of [47, Eq.(7) pg. 8].

Functional inequalities Functional inequalities such as Poincaré inequality, Talagrand's transport entropy inequality and the logarithmic Sobolev inequality are powerful tools to analyze the long time behavior of Markov processes and find plenty of applications in other fields such as spectral theory, the concentration of measure phenomenon, and hypercontractivity of Markov semigroups. Many monographs are devoted to functional inequalities and their applications, I refer to [13] as it contains the background material needed for this chapter, and much more. Why do I discuss functional inequalities here? Because it happens that in the attempt of showing exponential turnpike estimates for Schrödinger bridges, sometimes one discovers generalizations of classical functional inequalities where the entropic transportation cost takes the role of the transport distance, and sometimes it also happens that one unveils new functional inequalities that do not really have a "classical" counterpart. This is the case of the so called energy-entropy inequality, see (4.3.4) below.

4.1 Fourth encounter with the toy model

In this section, I obtain the turnpike property on the toy model. The transfer of results from the toy model to the various Schrödinger problems encountered in this manuscript is more or less difficult depending on the instance under consideration and it is realized by means of either Otto calculus (when possible), Itô calculus or the so called Γ calculus [11]. For the toy model, ergodicity of the underlying particle system is modeled through strong convexity of E , i.e. we assume that

$$\exists \kappa > 0 \text{ s.t. } \quad \nabla^2 E(x) \geq \kappa, \quad \forall x \in \mathbb{R}^d.$$

If we call x_∞ the unique minimizer of E , then our convexity hypothesis gives

$$E(x) \geq \frac{\kappa}{2} |x - x_\infty|^2,$$

and therefore we can establish the turnpike property by bounding either $E(x_t^u)$ or $|x_t^u - \bar{x}|^2$ along optimal curves. Of course bounding E yields stronger results and in order to do so, a key quantity to study is the *forward corrector*, that is the function

$$t \mapsto c(t) = \frac{1}{2} \int_0^t |u_s|^2 ds,$$

where u is an optimal control in (TSP). Note that $c(0) = 0$ and $c(T) = C_T(x, y) - 2E(x)$ and that using the state equation $\dot{x}^u = -\nabla E(x_t^u) + u_t$ we can rewrite

$$c_f(t) = \frac{1}{2} \int_0^t |\dot{x}_s^u + \nabla E(x_s^u)|^2 ds$$

But then, thanks to our Newton's law (2.2.2) we find

$$\begin{aligned}
\frac{d^2}{dt^2}c_f(t) &= \frac{d}{dt} \frac{1}{2} |\dot{x}_t^u + \nabla E(x_t^u)|^2 \\
&= \langle \dot{x}_t^u + \nabla E(x_t^u), \ddot{x}_t^u + \nabla^2 E(x_t^u) \cdot \dot{x}_t^u \rangle \\
&\stackrel{(2.2.2)}{=} \langle \dot{x}_t^u + \nabla E(x_t^u), \nabla^2 E(x_t^u) \cdot [\nabla E(x_t^u) + \dot{x}_t^u] \rangle \\
&\geq \kappa |\nabla E(x_t^u) + \dot{x}_t^u|^2 \\
&= 2\kappa \frac{d}{dt} c_f(t).
\end{aligned}$$

We have thus closed the differential inequality $\frac{d^2}{dt^2}c_f(t) \geq 2\kappa \frac{d}{dt}c_f(t)$, that immediately gives the *corrector estimate*

$$c_f(t) \leq \frac{\exp(2\kappa t) - 1}{\exp(2\kappa T) - 1} [C_T(x, y) - 2E(x)] \quad (4.1.1)$$

This estimate, together with a twin estimate for the backward corrector, that we define leveraging the time symmetry of the toy model as

$$c_b(t) = \frac{1}{2} \int_0^t |\nabla E(x_{T-s}^u) - \dot{x}_{T-s}^u|^2 ds$$

are the two fundamental estimates on which the proof of the turnpike property rests. The other main estimates are indeed obtained combining these two. Indeed, along any optimal curve we have,

$$\forall t \in [0, T], \quad E(x_t^u) = c_f(t) + c_b(T-t) - (C_T(x, y) - 2E(x) - 2E(y)).$$

Setting $t = T/2$ in the above display and supposing that

$$\sup_{T \geq 0} C_T(x, y) < +\infty \quad (4.1.2)$$

then the corrector estimates imply that $c_f(t) + c_b(T-t) \lesssim \exp(-\kappa T)$. Therefore, if we could show that

$$|C_T(x, y) - 2E(x) - 2E(y)| \lesssim \exp(-\kappa T) \quad (4.1.3)$$

then we would immediately obtain

$$E(x_{T/2}^u) \lesssim \exp(-\kappa T), \quad (4.1.4)$$

that is the prototype of a turnpike estimate, as it shows exponential convergence of $x_{T/2}^u$ towards x_∞ as $T \rightarrow +\infty$. I will present in the upcoming sections some precise results that prove that (4.1.3) together with the boundedness of $C_T(x, y)$ hold in many examples of interest. In fact, the bounds I will obtain later on are much finer than the loose statements I give here and relate to a novel class of functional inequalities. In the above calculations, I chose to work with $t = T/2$ but this is only for simplicity, and one could work with arbitrary values of $t \in [0, T]$.

4.2 Convexity of the entropy along Schrödinger bridges

Here, I treat the classical Schrödinger problem (**CSP**). In the article [81] I obtained a convexity estimate for the relative entropy along entropic interpolations, that is equivalent to the Bakry-Émery condition.

Theorem 4.2.1. *Let M be compact and $\mathcal{L} = \frac{1}{2}\Delta - \nabla U \cdot \nabla$. The following are equivalent:*

1. *The Bakry-Émery condition*

$$\forall x \in M, \quad \mathbf{Ric}_x + 2\mathbf{Hess}_x U \geq \kappa \mathbf{Id} \quad (\text{CD}(\kappa, +\infty))$$

2. For any μ, ν such that $\mathcal{H}(\mu|\mathbf{m}), \mathcal{H}(\nu|\mathbf{m}) < +\infty$ and any $T > 0$, the entropic interpolation satisfies

$$\forall t \in [0, T] \quad \mathcal{H}(\mu_t|\mathbf{m}) \leq \frac{1 - \exp(-\kappa(T-t))}{1 - \exp(-\kappa T)} \mathcal{H}(\mu|\mathbf{m}) + \frac{1 - \exp(-\kappa t)}{1 - \exp(-\kappa T)} \mathcal{H}(\nu|\mathbf{m}) - \frac{\cosh(\frac{\kappa T}{2}) - \cosh(-\kappa(t - \frac{T}{2}))}{\sinh(\frac{\kappa T}{2})} \mathcal{C}_T(\mu, \nu). \quad (4.2.1)$$

where $d\mathbf{m} = \exp(-2U)d\mathbf{vol}$ is the reversible measure for \mathcal{L} .

The interest of Theorem 4.2.1 is threefold. In first place and although this estimate is not *per se* a turnpike result, it is an important step towards establishing actual turnpike estimates as it immediately yields a quantitative form of (4.1.2) that will also be used to obtain a sharp form of the loose bound (4.1.3). Secondly, this estimate is a strict generalization of the fundamental displacement convexity [171, 211] of relative entropy along geodesics, that reads as

$$\mathcal{H}(\mu_t|\mathbf{m}) \leq \mathcal{H}(\mu|\mathbf{m})(1-t) + \mathcal{H}(\nu|\mathbf{m})t - \kappa \frac{t(1-t)}{2} W_2^2(\mu, \nu). \quad (4.2.2)$$

where $(\mu_t)_{t \in [0,1]}$ is the geodesic (displacement interpolation) connecting μ and ν and $d\mathbf{m} = \exp(-2U)d\mathbf{vol}$. To be more precise, operating the change of variable $t = Ts$, multiplying by T on both sides of (4.2.1) below and letting $T \rightarrow 0$ retrieves formally (4.2.2). The rigorous justification of this limit hinges on the Γ -convergence of (CSP) to the dynamic formulation of the Monge-Kantorovich problem in the short time limit [159, 174]. This convergence result and its consequences are not a central theme of my thesis. However, it is worth noticing that they are at the origin of the interest that the Schrödinger problem has generated among analysts and researchers in machine learning. To the best of my knowledge, convexity estimates for the relative entropy were previously obtained only along geodesics via optimal transport methods or along gradient flows using Γ -calculus techniques. The convexity estimate (4.2.1) holds along entropic interpolations, that are curves sharing some features with both gradient flows and geodesics. As I have just explained, convexity along geodesics is retrieved for $T \rightarrow 0$. Moreover, it is not difficult to see that the exponential decay of the entropy along the gradient flow is recovered in the $T \rightarrow +\infty$ limit.

The corrector estimate At the heart of the proof Theorem 4.2.1 is an estimate akin to (4.1.1). If we use Otto calculus, it is just the literal transposition of the estimate on the toy model and reads as

$$\int_0^t |v_s + \nabla^{W_2} \mathcal{H}(\mu_s|\mathbf{m})|_{\mathcal{T}_{\mu_s} \mathcal{P}_2}^2 ds \leq \frac{\exp(2\kappa t) - 1}{\exp(2\kappa T) - 1} [\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu|\mathbf{m})], \quad (4.2.3)$$

where $(v_t)_{t \in [0, T]}$ is the velocity field of the entropic interpolation $(\mu_t)_{t \in [0, T]}$. As realized in [164], the derivatives of the entropy along Schrödinger bridges can be expressed by means of the Γ and Γ_2 operators. To see this, let me recall their definition beginning from Γ , given by

$$\Gamma(f, g) := \frac{1}{2} \left(\mathcal{L}(fg) - f\mathcal{L}g - g\mathcal{L}f \right) = \frac{1}{2} |\nabla f|^2, \quad f, g \in \mathcal{C}_c^\infty(M).$$

The operator Γ_2 is an iterated version of the Γ operator defined as follows

$$\Gamma_2(f, g) = \mathcal{L}\Gamma(f, g) - \Gamma(f, \mathcal{L}g) - \Gamma(\mathcal{L}f, g).$$

In their fundamental contribution, Bakry and Émery showed that the second derivative of the entropy along the heat flow is expressed in terms of the Γ_2 operator. Moreover, exploiting Bochner's identity they established that under $\text{CD}(\kappa, +\infty)$ the pointwise estimate

$$\forall f \in \mathcal{C}_c^\infty(M), x \in M \quad \Gamma_2(f, f)(x) \geq 2\kappa \Gamma(f, f)(x).$$

that eventually leads to their celebrated proof of the logarithmic Sobolev inequality. The last missing step needed to connect these differential operators with the Schrödinger problem is to recall that the entropic interpolation μ_t rewrites

as a product $d\mu_t = f_t g_t d\mathbf{m}$, where f_t satisfies the forward Kolmogorov equation and g_t the backward Kolmogorov equation. Using this decomposition, the corrector becomes

$$c_f(t) = \frac{1}{2} \int_0^T \int_M |\nabla \log g_s|^2 f_t g_t d\mathbf{m} ds.$$

Moreover, the corrector estimate (4.2.3) can indeed be equivalently expressed in the Γ -calculus framework and reads as follows

$$\frac{d^2}{dt^2} c_f(t) = \int \Gamma_2(\log g_t, \log g_t) f_t g_t d\mathbf{m} \geq 2\kappa \int \Gamma(\log g_t, \log g_t) f_t g_t d\mathbf{m} = \frac{d}{dt} c_f(t) \quad (4.2.4)$$

Note that the convexity of the entropy along the heat flow is equivalent to

$$\frac{d}{dt} \int \Gamma(\log g_t, \log g_t) g_t d\mathbf{m} = \int \Gamma_2(\log g_t, \log g_t) g_t d\mathbf{m} \geq 2\kappa \int \Gamma(\log g_t, \log g_t) g_t d\mathbf{m},$$

and is thus what one obtains from (4.2.4) setting $f \equiv 1$.

Entropic Talagrand inequality Let me recall Talagrand's transportation inequality [199, 183], valid under the Bakry-Émery condition.

$$\forall \mu \in \mathcal{P}_2(M), \quad W_2^2(\mu, \mathbf{m}) \leq \frac{2}{\kappa} \mathcal{H}(\mu | \mathbf{m}). \quad (\mathbf{TI}(\kappa))$$

An immediate corollary of Theorem 4.2.1 are the following generalizations

Corollary 4.2.1. *Let M be compact and $CD(\kappa, +\infty)$ hold for some $\kappa > 0$. Then, for any $\mu, \nu \in \mathcal{P}(M)$ and $T > 0$ we have*

$$\mathcal{C}_T(\mu, \mathbf{m}) \leq \frac{1}{1 - \exp(-\kappa T)} \mathcal{H}(\mu | \mathbf{m}) \quad (\mathbf{ETI}(\kappa, T))$$

and

$$\mathcal{C}_T(\mu, \nu) \leq \frac{1}{1 - \exp(-\frac{T\kappa}{2})} (\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})). \quad (4.2.5)$$

Using the asymptotic convergence of the entropic cost towards the Wasserstein distance

$$\lim_{T \rightarrow 0} T \mathcal{C}_T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu)$$

we immediately find that $\mathbf{ETI}(\kappa, T)$ becomes $\mathbf{TI}(\kappa)$. I called such inequality *entropic* Talagrand's inequality, owing to the fact that is obtained replacing $W_2^2(\cdot, \mathbf{m})$ with $\mathcal{C}_T(\cdot, \mathbf{m})$ and readjusting the constants in the classical Talagrand's inequality. In the article [93], written in collaboration with Luigia Ripani, we studied the main properties and some equivalent characterizations of both $\mathbf{ETI}(\kappa, T)$ and (4.2.5). Among other results and beyond its usefulness in establishing the turnpike property, we found that (4.2.5) is equivalent to a weak form of reverse hypercontractivity for the associated semigroup and implies some new dimension-free concentration of measure properties for the invariant measure \mathbf{m} . Moreover, $\mathbf{ETI}(\kappa, T)$ is equivalent to various contractivity properties for Hamilton Jacobi Bellman semigroups. I do not report these results here, since doing so would require to introduce quite some notation that would make this manuscript less readable. Instead, I warmly refer the interested reader to the original article and I limit the present discussion to the relations between $\mathbf{ETI}(\kappa, T)$, the logarithmic Sobolev inequality and $\mathbf{TI}(\kappa)$. To best present these connections, let me recall that \mathbf{m} satisfies the logarithmic Sobolev inequality with constant κ if and only if for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ whose density against \mathbf{m} is in the domain of \mathcal{L} we have

$$\mathcal{H}(\mu | \mathbf{m}) \leq \frac{1}{\kappa} \int |\nabla \log \frac{d\mu}{d\mathbf{m}}|^2 d\mu, \quad (\mathbf{LSI}(\kappa))$$

A remarkable result of Otto and Villani [183] states that $\mathbf{LSI}(\kappa) \Rightarrow \mathbf{TI}(\kappa)$ in a general setting. Quite interestingly, $\mathbf{ETI}(\kappa, T)$ can be placed in between these two classical inequalities.

Theorem 4.2.2. *For any $\kappa, T > 0$ we have the following relations*

(i) If \mathbf{m} satisfies $\mathbf{LSI}(\kappa)$, then it satisfies $\mathbf{ETI}(\kappa, T)$ for any $T > 0$.

(ii) If \mathbf{m} satisfies $\mathbf{ETI}(\kappa, T)$, then it satisfies $\mathbf{TI}(\kappa')$ with

$$\kappa' = \frac{1}{T \left(\frac{2}{1 - \exp(-\kappa T)} - 1 \right)}.$$

Similar statements about (4.2.5) can also be found in [93]. In order to prove Theorem 4.2.2 we followed a technique developed in [35, 143] that consists in finding the dual form of $\mathbf{ETI}(\kappa, T)$ and comparing it with the dual forms of $\mathbf{LSI}(\kappa)$ and $\mathbf{TI}(\kappa)$.

4.3 The total energy and long time behavior of the entropic cost

In this section, I report on the results of the article [81], written in collaboration with Luca Tamanini. In this work, we have investigated in detail the properties of the entropic cost $\mathcal{C}_T(\cdot, \cdot)$ as a function of the time parameter T . In the $T \rightarrow +\infty$ regime, our findings connect with the turnpike property, as I am going to explain.

4.3.1 The conserved total energy and the energy-transport inequality

The work we did together got us convinced of the relevance that the *conserved total energy* has in the study of the Schrödinger problem. In order to introduce this quantity, let me go back once more to the toy model. There, thanks to Newton's law (2.2.2) it is immediately seen that the quantity

$$t \mapsto \frac{1}{2} |\dot{x}_t^u|^2 - \frac{1}{2} |\nabla E(x_t^u)|^2 \quad (4.3.1)$$

is constant in t along the unique optimal curve (but of course depends on μ, ν and T). In this context, this expression indeed admits the interpretation of the sum of the kinetic and potential energies of a physical system, and this analogy is what motivated our definition. In the context of (\mathbf{CSP}) , we define

$$\mathcal{E}_T(\mu, \nu) = \frac{1}{2} |v_t|_{\mathcal{T}_{\mu_t} \mathcal{P}_2}^2 - \frac{1}{8} \mathcal{I}(\mu_t), \quad (4.3.2)$$

where I recall that $\mathcal{I}(\cdot)$ stands for the Fisher information functional and, following the general convention of this manuscript, I denote (μ_t) the entropic interpolation and (v_t) its velocity field. It not hard to see that the above is a good definition, in the sense that the right hand side does not depend on t and that the entropic interpolation is unique. In the Otto calculus interpretation of the entropic interpolation, $\mathcal{E}_T(\mu, \nu)$ has the exact same meaning of (4.3.1). It turns out that $\mathcal{E}_T(\mu, \nu)$ and $\mathcal{C}_T(\mu, \nu)$ are tightly connected, as the next theorem clearly shows.

Theorem 4.3.1. *Assume that the Bakry-Émery condition holds for some $\kappa \in \mathbb{R}$ and that μ, ν have compact support and bounded density against \mathbf{m} . Then the map $T \mapsto \mathcal{C}_T(\mu, \nu)$ is $C^1((0, \infty); \mathbb{R})$, twice differentiable a.e. and the first derivative is given by*

$$\frac{d}{dT} \mathcal{C}_T(\mu, \nu) = -\mathcal{E}_T(\mu, \nu), \quad \forall T > 0;$$

In words: the energy is equal to the time derivative of the cost up to a sign. Although this relation might look a bit surprising, one could have guessed this result by applying formally the *envelope theorem* to the Benamou-Brenier formulation of the Schrödinger problem. Very recently, the relation (4.3.1) has been established in a far more general context in [176] and plays a crucial role in the study of the long time behavior $\mathcal{C}_T(\mu, \nu)$, as I will discuss in the next section.

4.3.2 Long time behavior of the entropic cost

Here, I study the long time behavior of the entropic cost $\mathcal{C}_T(\mu, \nu)$ beginning with the most basic task, that is to correctly identify the limit. The right answer is not too hard to guess. Indeed, when $T \rightarrow +\infty$, the joint law $R_{0,T}$ converges to the independent coupling $\mathbf{m} \otimes \mathbf{m}$ and therefore the optimal coupling in the static version of the Schrödinger problem is expected to converge to the product coupling, thus providing with a natural interpretation for the following theorem.

Theorem 4.3.2. *Assume that the Bakry-Émery condition holds for $\kappa \geq 0$ and that \mathbf{m} is a probability measure. If $\mathcal{H}(\mu|\mathbf{m}), \mathcal{H}(\nu|\mathbf{m})$ have finite entropy, then*

$$\lim_{T \rightarrow \infty} \mathcal{C}_T(\mu, \nu) = \mathcal{H}(\mu|\mathbf{m}) + \mathcal{H}(\nu|\mathbf{m}).$$

Moreover, if $\mu, \nu \in \mathcal{P}(M)$ have compact support and bounded density against \mathbf{m}

$$\lim_{T \rightarrow \infty} \mathcal{E}_T(\mu, \nu) = 0.$$

The proof of (4.3.2) relies on a Γ -convergence argument and heat kernel estimates. In fact, in the article we obtained a stronger result than the one displayed above that allowed us to determine the long-time behavior of the fg decomposition of the optimal coupling in (CSP). Let us now move on to show an exponential rate of convergence for (4.3.2). The key message is that, under a positive curvature condition, the approximation error is asymptotically smaller than $\exp(-\kappa T/2)$, up to constant factors depending on $\mathcal{H}(\mu|\mathbf{m})$ and $\mathcal{H}(\nu|\mathbf{m})$.

Theorem 4.3.3. *Let us assume that $CD(\kappa, +\infty)$ holds with $\kappa > 0$ and that $\mathcal{H}(\mu|\mathbf{m}) + \mathcal{H}(\nu|\mathbf{m}) < +\infty$. Then for all $T > 0$ it holds*

$$|\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu|\mathbf{m}) - \mathcal{H}(\nu|\mathbf{m})| \leq \frac{2}{\exp(\kappa T/2) - 1} \left(\mathcal{H}(\mu|\mathbf{m}) + \mathcal{H}(\nu|\mathbf{m}) \right). \quad (4.3.3)$$

If moreover, μ, ν have compact support and bounded density against \mathbf{m} , we also have

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{\kappa \exp(-\kappa T/2)}{(1 - \exp(-\kappa T/2))^2} \left(\mathcal{H}(\mu|\mathbf{m}) + \mathcal{H}(\nu|\mathbf{m}) \right)^2. \quad (4.3.4)$$

Furthermore, the exponential rate $-\kappa/2$ in (4.3.3) and (4.3.4) is sharp.

In order to establish the above theorem, we obtained a novel functional inequality that we called *energy transport inequality* for obvious reasons. This inequality states that if the Bakry-Émery condition holds for some $\kappa > 0$ and μ, ν are regular enough, then we for all $T > 0$ we have

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{\kappa}{\exp(\kappa T/2) - 1} \sqrt{(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu|\mathbf{m}))(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\nu|\mathbf{m}))}. \quad (4.3.5)$$

The energy transport inequality does not really admit a counterpart among classical functional inequalities. Indeed, as $T \rightarrow 0$ both cost and energy converge to the squared Wasserstein distance when appropriately scaled! With Theorem 4.3.3 we have thus obtained a precise and sharp version of (4.1.3). This was the last missing estimate we needed and deriving a quantitative version of the turnpike property for (CSP) is now only a matter of putting all the pieces together. But before doing so, let me discuss a bit the *short-time* behavior of the cost.

4.3.3 Short-time behavior of the entropic cost

In this section I deviate from the thread of this chapter by discussing the $T \rightarrow 0$ behavior of the cost, that clearly is not related to the turnpike property. More precisely, I present here some results taken from [81] that contribute to the understanding of the relation between (CSP) and the Monge-Kantorovich problem. As I recalled above, this story begins with Mikami's paper where a rigorous convergence result of (CSP) towards (2.1.1) was obtained for the first time. Mikami's result has been later largely generalized by Léonard [159] and eventually reproven and strengthened in various forms over the last few years, see [29]. In [1], a further fundamental step forward was taken that consists in

computing the first order term in the expansion of $TC_T(\mu, \nu)$ around $T = 0$, thus revealing a new relation between De Giorgi's minimizing movement scheme and the Schrödinger problem. In a nutshell we have the following expansion

$$TC_T(\mu, \nu) = \frac{1}{2}W_2^2(\mu, \nu) + \frac{T}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + o(T). \quad (4.3.6)$$

Estimates of the form (4.3.6) that improve on the earlier findings of [1] have been obtained in [108, 116, 184] among others. Together with Luca I showed a non-asymptotic bound for $TC_T(\mu, \nu) - W_2^2(\mu, \nu)/2$ which is sharp in the limit $T \rightarrow 0$ and we computed the second order term in (4.3.6). Here is what we got:

$$TC_T(\mu, \nu) = \frac{1}{2}W_2^2(\mu, \nu) + \frac{T}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + \frac{T^2}{8} \int_0^1 \int_M |\nabla \log \mu_t^0|^2 d\mu_t^0 dt + o(T^2), \quad (4.3.7)$$

where $(\mu_t^0)_{t \in [0,1]}$ denotes the (unique) Wasserstein geodesic between μ and ν . Let us remark that (4.3.7) tells that the rescaled cost is convex around $T = 0$ and using functional inequalities such as the HWI inequality [183] one can also estimate from below its second derivative under the Bakry-Émery condition.

Theorem 4.3.4. *Assume that μ, ν have compact support and bounded density against \mathbf{m} . Then for all $T > 0$ we have:*

$$0 \leq TC_T(\mu, \nu) - \frac{1}{2}W_2^2(\mu, \nu) \leq \frac{T}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + \frac{T^2}{8} \int_0^1 \int_M |\nabla \log \mu_t^0|^2 d\mu_t^0 dt, \quad (4.3.8)$$

where $(\mu_t^0)_{t \in [0,1]}$ denotes the Wasserstein geodesic between μ and ν . If in addition the Bakry-Émery condition $CD(\kappa, +\infty)$ holds for some $\kappa \in \mathbb{R}$ and $\int_0^1 \int_M |\nabla \log \mu_t^0|^2 dt d\mu_t^0 < \infty$, then $T \mapsto TC_T(\mu, \nu)$ is twice differentiable at $T = 0$ and (4.3.7) holds.

Quite curiously, the second order expansion of the cost finds applications in machine learning, see [73] where a proof of (4.3.7) in the Euclidean case has also been obtained.

4.4 Entropic turnpike estimates

4.4.1 Classical Schrödinger problem

Building on the results of the former section, one could show that turnpike estimates of the form (4.1.4) hold. Since these estimates take the same form for (CSP) and (MFSP), I am only going to present them in the latter case, specifying what adaptations need be to pass from one problem to the other. All this is done at section 4.4.2.1. In the upcoming sections, I shall only focus on results that have not yet found a counterpart in the mean field setting.

4.4.1.1 Curvature dimension condition

Together with Ivan Gentil and Gauthier Clerc we worked towards understanding what can be said about the long time behavior of entropic interpolations assuming, instead of the Bakry Émery condition $(CD(\kappa, +\infty))$, a *curvature-dimension* condition. In particular, we focused on the $CD(0, n)$ condition, that can be stated in terms of the generator \mathcal{L} and the Γ_2 operator as follows

$$\forall f \in C_c^\infty, \quad \Gamma_2(f, f) \geq \frac{1}{n} (\mathcal{L}f)^2. \quad (CD(0, n))$$

General curvature-dimension conditions encompassing both $CD(\kappa, +\infty)$ and $CD(0, n)$ have been put forward in the seminal work of Bakry and Émery [11]. In another remarkable work, Erbar Kuwada and Sturm obtained, among many other results, characterizations of general curvature-dimension conditions in terms of a kind of twisted convexity estimate for the relative entropy along displacement interpolations. One of the main reasons why $CD(0, n)$ is an interesting working hypothesis, is that it covers the fundamental example of the Brownian particles in \mathbb{R}^d , when the generator is simply $\frac{1}{2}\Delta$. In this case, the condition holds with $n = d$ and the measure \mathbf{m} is the Lebesgue measure, that is obviously not a probability measure. For this reason, it does not really make sense to speak of a turnpike theorem in this context, as the turnpike simply does not exist. However, something can still be said about the ergodic behavior of entropic interpolations, and this is what I am going to explain in the next two paragraphs.

Convergence to the gradient flow on a fixed window Most turnpike results discussed in this manuscript concern the behavior of entropic interpolations on a time scale that is of the same order $O(T)$ of the time-window over which the Schrödinger problem is set. As I have explained above, we cannot expect results of this type under the $CD(0, n)$ condition. However, we can still translate the intuition that Schrödinger's thought experiment provides us with by saying that, if we look at a fixed time t and if we make an observation at time $T \gg 1$, this observation has almost no effect on the dynamics of the particle systems at time t , that looks almost like the unconditional dynamics. For the case of the independent particles, this would mean that if T is very large, then the entropic interpolation μ_t is almost equal to the law at t of a Brownian motion started at μ . Here is a quantitative version of this statement.

Theorem 4.4.1. *Let $CD(0, n)$ hold, $\mu, \nu \in \mathcal{P}_2(M)$ be two absolutely continuous and compactly supported measures with smooth density w.r.t. \mathbf{m} and $(\mu_t)_{t \in [0, T]}$ the entropic interpolation from μ to ν . Moreover, let us denote by $(\rho_t)_{t \geq 0}$ the law at time t of the SDE*

$$\begin{cases} dX_t = -\nabla U(X_t)dt + dB_t, \\ X_0 \sim \mu \end{cases} \quad (4.4.1)$$

Then, for every $T > 1$ and $t \in (0, T)$ we have

$$W_2(\mu_t, \rho_t) \leq 2\sqrt{2(\mathcal{H}(\nu|\mathbf{m}) - \mathcal{H}(\mu|\mathbf{m})) + C_1(\mu, \nu) + 2n \log(T)} \left(\sqrt{T} - \sqrt{T-t} \right).$$

Furthermore for any $a > 1$, there exists a constant $C > 0$, such that for all $T \geq a$ and $t \in [0, a]$,

$$W_2(\mu_t, \rho_t) \leq C \sqrt{\frac{n \log(T)}{T}}.$$

Note that in the above theorem, although it is not written explicitly, the law μ_t is of course dependent on T . The proof exploits the information contained in $CD(0, n)$ in order to obtain an adapted version of the corrector estimate (4.2.3). The resulting bounds may be seen as a generalization of Costa's inequality (see [201] for recent developments) and improve on the earlier findings of [192]. I do not give more details at this point since I will revisit corrector estimates under $CD(0, n)$ in relation with local logarithmic Sobolev inequalities in the next section. The conclusion of Theorem 4.4.1 may be not be optimal. More precisely, it is not optimal for \mathbb{R}^d equipped with the usual Laplacian operator, where convergence happens at rate $1/T$. However, we do not know whether it is possible to improve on Theorem 4.4.1 assuming the $CD(0, n)$ condition only.

Long time behavior cost and energy In parallel with Theorem 4.3.3, we managed to obtain the following asymptotic results for the entropic cost and the conserved total energy.

Theorem 4.4.2. *Assume that $CD(0, n)$ holds and that $\mu, \nu \in \mathcal{P}_2(M)$ are two compactly supported absolutely continuous measures with smooth positive densities w.r.t. \mathbf{m} . Then for every $T > 0$ we have*

$$-\mathcal{E}_T(\mu, \nu) \leq \frac{2n}{T}, \quad \mathcal{C}_T(\mu, \nu) \leq C_1(\mu, \nu) + 2n \log(T).$$

The compactness assumption of the above theorem is there merely for technical reasons, and it is very likely to be unnecessary. The asymptotic rates can be seen to be sharp for Brownian particles on \mathbb{R}^d . In particular, note that, in stark contrast with Theorem 4.3.3 the cost may diverge as $T \rightarrow +\infty$.

A turnpike estimate Under the $CD(0, n)$ condition the following estimates for the Fisher information along the heat flow (4.4.1) is well-known

$$\mathcal{I}(\rho_t|\mathbf{m}) = \int |\nabla \log\left(\frac{d\rho_t}{d\mathbf{m}}\right)|^2 d\rho_t \leq \frac{n}{2t}, \quad \mu \in \mathcal{P}_2(M), t > 0. \quad (4.4.2)$$

We obtained an analogous estimate along the entropic interpolations.

Theorem 4.4.3. *Let $\mu, \nu \in \mathcal{P}_2(M)$ be two compactly supported absolutely continuous measures with smooth positive densities against \mathbf{m} . Then for every $T > 0$ and $t \in (0, T)$ we have*

$$\mathcal{I}(\mu_t | \mathbf{m}) \leq \frac{n}{2t} + \frac{n}{2(T-t)}.$$

This estimate is efficient at a timescale of order T , and not simply for a fixed t as the one I showed above. It is as close as we can get to a turnpike theorem. As it implies the (4.4.2), it is sharp. If \mathbf{m} is a probability measure, we can deduce from (4.4.2) an algebraic turnpike estimate.

4.4.1.2 Variational interpretation of local logarithmic Sobolev inequalities

Together with Ivan and Gauthier, we also noticed [77] that an appropriate form of the corrector estimates yields a variational interpretation of *local* logarithmic Sobolev inequalities, such as the celebrated Li-Yau inequality [166] valid under $\text{CD}(0, n)$ and the so called *heat kernel* logarithmic Sobolev inequalities valid under $\text{CD}(\kappa, +\infty)$, see [13, Sec 5.5]. Optimal transport and Otto calculus have successfully been applied to provide a geometrical interpretation and new stronger forms of several classical functional inequalities. For example, the celebrated HWI inequality [183] has indeed been discovered following this powerful heuristic guideline. However, to the best of our knowledge, an optimal transport interpretation of local inequalities was missing and the message of this section is that Otto calculus, when applied to the Schrödinger problem solves this problem. I will give the announced interpretation of the Li-Yau inequality, referring to the note we wrote for heat kernel logarithmic Sobolev inequalities.

Li-Yau from the toy model to (CSP) Although it is not immediately clear how to do so, a bit of thinking suggests that the reduction of $\text{CD}(0, n)$ to the toy model is the following

$$\forall x, v \in \mathbb{R}^d, \quad \langle \nabla^2 E(x) \cdot v, v \rangle \geq \frac{1}{n} (\langle \nabla E, v \rangle)^2.$$

Arguing on the basis of Newton's law and $\text{CD}(0, n)$ we can close a differential inequality for the *modified corrector* $\tilde{c}_f(t) = c_f(t) - t\mathcal{E}_T(x, y)$ which reads as follows

$$\frac{d^2}{dt^2} \tilde{c}_f(t) \geq \frac{1}{2n} \left(\frac{d}{dt} \tilde{c}_f(t) \right)^2.$$

In the above display, $\mathcal{E}_T(x, y)$ is the conserved total energy (4.3.1). This differential inequality is equivalent to concavity of the map $[0, T] \ni t \mapsto \exp\left(-\frac{\tilde{c}_f(t)}{2n}\right)$: recalling that u_t is the optimal control in (TSP) and using some basic convexity estimates, we deduce the following

$$\exp\left(-\frac{1}{2n} [C_T(x, y) + 2E(y) - 2E(x) - T\mathcal{E}_T(x, y)]\right) \leq 1 - \frac{T}{2n} (|u_0|^2 - \mathcal{E}_T(x, y)).$$

In particular,

$$\mathcal{E}_T(x, y) \geq -\frac{2n}{T} + |u_0|^2.$$

The reinterpretation of this last estimate for the classical Schrödinger problem is that under $\text{CD}(0, n)$ is that for all μ, ν sufficiently regular we have

$$\mathcal{E}_T(\mu, \nu) \geq -\frac{2n}{T} + |v_0 + \nabla^{W_2} \mathcal{H}(\mu_0 | \mathbf{m})|_{\mathcal{T}_{\mu_0} \mathcal{P}_2}^2.$$

Expressing the above inequality in terms of the fg -decomposition and the Γ, Γ_2 operators, and specializing it to pairs (μ, ν) with $\mu(dx) = \delta_y(x)$ for some $y \in M$ we finally arrive at the Li-Yau inequality [166],[12]

$$\frac{\Gamma(P_T g)}{(P_T g)^2} - \frac{\Delta P_T g}{P_T g} \leq \frac{n}{2T},$$

where I recall that P_T is the semigroup generated by \mathcal{L} .

4.4.2 Mean Field Schrödinger problem

It is time to finally present quantitative versions of the turnpike property for Schrödinger bridges. We work here in the context of **(MFSP)** but, as I have said above and as I will recall below, similar results hold for **(CSP)** under **CD**($\kappa, +\infty$). Also, I will only discuss in this manuscript turnpike estimates for entropic functionals, such as the relative entropy or the free energy (2.3.5). However, for each of the theorems appearing in the next sections, one could have an analogous weaker statement in terms of the Wasserstein distance $W_2(\cdot, \cdot)$.

4.4.2.1 Exponential convergence to equilibrium and the turnpike property

We are here in the setting of section (2.3). In addition to (2.3.4),(2.3.3) we also assume that the interaction potential W is strongly κ -convex for some $\kappa > 0$ and that the two marginals have the same mean, i.e.

$$\forall z, v \in \mathbb{R}^d, \quad \langle \nabla^2 W(z) \cdot v, v \rangle \geq \kappa |v|^2, \quad \int_{\mathbb{R}^d} z \mu(dz) = \int_{\mathbb{R}^d} z \nu(dz). \quad (4.4.3)$$

This is a classical assumption ensuring the ergodicity of the McKean-Vlasov dynamics for (2.3.1). Moreover, McCann [171] showed that, for any fixed $m \in \mathbb{R}^d$, there exists a unique minimiser of the free energy functional $\tilde{\mathcal{F}}$ among all probability measures with mean m . We call μ_∞ the minimiser associated with the value $m = \int z \mu(dz)$. Asking that μ and ν have the same mean is not really necessary and we could remove this assumption with some extra work. However, I prefer to keep it there to simplify the presentation. In order to express the turnpike property we shift the functional $\tilde{\mathcal{F}}$ in such a way that its minimum value is 0 and therefore consider $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ defined by

$$\mathcal{F}(\mu) := \tilde{\mathcal{F}}(\mu) - \tilde{\mathcal{F}}(\mu_\infty). \quad (4.4.4)$$

An important difference between **(CSP)** and **(MFSP)** is that for the latter no uniqueness result appears to be known. Therefore the total conserved energy, that is defined as in (4.3.2), depends a priori on the optimizer P and we shall denote it $\mathcal{E}_{T,P}(\mu, \nu)$. The preparation is now finished and I can state the following turnpike theorem, that I extract from [7].

Theorem 4.4.4. *Assume (2.3.3),(2.3.4),(4.4.3) and let P be a Schrödinger bridge. For all $t \in [0, T]$ we have*

$$\mathcal{F}(P_t) \leq \frac{\sinh(2\kappa(T-t))}{\sinh(2\kappa T)} \left(\mathcal{F}(\mu) - \frac{\mathcal{E}_{T,P}(\mu, \nu)}{2\kappa} \right) + \frac{\sinh(2\kappa t)}{\sinh(2\kappa T)} \left(\mathcal{F}(\nu) - \frac{\mathcal{E}_{T,P}(\mu, \nu)}{2\kappa} \right) + \frac{\mathcal{E}_{T,P}(\mu, \nu)}{2\kappa}. \quad (4.4.5)$$

Moreover, for all fixed $\theta \in (0, 1)$ there exists a decreasing function $B(\cdot)$ such that

$$\mathcal{F}(P_{\theta T}) \leq B(\kappa) (\mathcal{F}(\mu) + \mathcal{F}(\nu)) \exp(-2\kappa \min\{\theta, 1 - \theta\}T) \quad (4.4.6)$$

uniformly in $T \geq 1$.

Let me also state the corresponding result for **(CSP)** in a very informal way.

Theorem 4.4.5. *For **(CSP)**, the estimate (4.4.6) holds under **CD**($\kappa, +\infty$) replacing \mathcal{F} with the relative entropy $\mathcal{H}(\cdot | m)$.*

Theorem 4.4.4 contains, as a special case and up to a multiplicative constant, the same estimate for the rate of convergence to equilibrium of the McKean-Vlasov dynamics that is found in [61]. Therefore, the exponential rates in (4.4.5) are sharp. To the best of my knowledge, there are very few available results on the ergodic behavior of McKean-Vlasov control problems. As a notable exception I mention the articles [51, 52] on the long time behavior of mean field games that contain comparable results, although in a very different setting. Consequently, the proof methods are also different. The proof technique we adopted in [7] hinges on a corrector estimate and on an energy-transport inequality, and virtually extends to a much broader class of systems. Some examples of potential applications can be found in [136, 158]. A comment on the hypothesis: only (4.4.3) has a conceptual relevance for the turnpike property: the other hypothesis are technical are there only to make sure that we can work with optimality conditions in the form of the FBSDE system (2.3.10).

Stochastic calculus proof of the corrector estimate I now sketch the proof of the corrector estimate on which the proof of Theorem 4.4.4 is built. This time, it is stochastic calculus rather than Otto calculus that will serve our purposes. Using the notation of Theorem (2.3.2) we define the corrector as

$$c_f(t) = \frac{1}{2} \mathbb{E} \left[\int_0^t |Y_s|^2 ds \right], \quad \text{with } Y_s = \Psi^P(X_s).$$

Using the FBSDE (2.3.10) and denoting by $\langle Y \rangle_t$ the quadratic variation, we find

$$\begin{aligned} d|Y_t|^2 &= 2Y_t dY_t + d\langle Y \rangle_t \\ &= 2Y_t \cdot \tilde{\mathbb{E}}[\nabla^2 W(X_t - \tilde{X}_t) \cdot (Y_t - \tilde{Y}_t)] dt + Y_t \cdot Z_t \cdot dB_t + d\langle Y \rangle_t \end{aligned}$$

Since quadratic variation is an increasing process, taking expectations on both sides in the above display and using some standard tricks one gets the estimate

$$\frac{d^2}{dt^2} c_f(t) \geq \mathbb{E}[Y_t \cdot \tilde{\mathbb{E}}[\nabla^2 W(X_t - \tilde{X}_t) \cdot (Y_t - \tilde{Y}_t)]] \geq \kappa \mathbb{E}[|Y_t|^2] = 2\kappa \frac{d}{dt} c_f(t),$$

that once integrated yields an estimate resembling (4.1.1). Note that that the same proof would work for (**CSP**), thus giving a way to do computations analogous to those of Bakry-Émery's Γ -calculus by means of stochastic calculus. For classical SDEs, analogies between these two formalisms have been pointed out in [127, 153]. What the above calculation reveals is that the parallelism extends to *controlled* diffusion processes. A similar estimate for the backward corrector can be obtained exploiting results on time reversal for diffusion processes under a finite entropy condition [125].

4.4.3 Kinetic Schrödinger problem

This section is devoted to a turnpike estimate for the kinetic Schrödinger problem (**KSP**) obtained in collaboration with A.Chiarini and G.Greco, who is working on his PhD dissertation under our joint supervision. Proving the turnpike property for Schrödinger bridges in this context is harder than in (**CSP**), and we need to work under stronger assumptions on the potential U than its strong convexity. This is not a surprise. Indeed, proving the exponential convergence to equilibrium for the kinetic Fokker Planck equation is a difficult problem that has been, and still is, intensively studied by means of either a probabilistic or an analytic approach, see [65, 111, 200, 145] for some references on the probabilistic approach. To put it simply, what makes the turnpike property difficult to establish is the fact that corrector estimates of the form (4.1.1) do not hold in the present context. Following the terminology introduced by Villani in his monograph [66], this obstruction is a manifestation of the *hypocoercive* nature of the Kinetic Fokker Planck equation. (**KSP**) may indeed be regarded as the prototype of an hypocoercive stochastic control problem. For the moment, we have been able to show the turnpike property under the following *quasi-linearity* assumption:

$$\alpha \mathbf{Id} \leq \nabla^2 U(x) \leq \beta \mathbf{Id} \quad \forall x \in \mathbb{R}^d, \text{ and for some } 0 < \alpha < \beta < +\infty, \text{ with } \gamma \geq \sqrt{\beta} - \sqrt{\alpha}. \quad (4.4.7)$$

Assumptions of this type, where the friction parameter has to be in some sense large in comparison with the spectrum of U are commonly encountered in the literature. In the language of probability, they ensure that the synchronous coupling is contracting for the Langevin dynamics [38, 175]. On the other hand, from an analytical standpoint, Assumption (4.4.7) implies *local* gradient bounds for the semigroup generated by the Langevin dynamics [15]. Here is one of the main results we have established so far. In there, $(\mu_t)_{t \in [0, T]}$ denote the entropic interpolation for (**KSP**).

Theorem 4.4.6. *Fix $\delta \in (0, 1)$, assume that U is smooth, $\mathcal{H}(\mu|\mathbf{m}) + \mathcal{H}(\nu|\mathbf{m}) < +\infty$ and that (4.4.7) holds. Then, there exist $T_0 = T_0(\alpha, \beta, \kappa, \delta, d)$ such that as soon as $T > T_0$ and $t \in [\delta, T - \delta]$, it holds*

$$\mathcal{H}(\mu_t|\mathbf{m}) \leq C_{d, \alpha, \beta, \gamma} \delta^{-3} e^{-2\kappa[t \wedge (T-t)]} \mathcal{C}_T(\mu, \nu), \quad (4.4.8)$$

where $C_{d, \alpha, \beta, \gamma}$ is a numerical constant. Moreover, we have

$$\mathcal{H}(\mu_t|\mathbf{m}) \leq C_{d, \alpha, \beta, \gamma} \delta^{-3} e^{-2\kappa[t \wedge (T-t)]} \left[\mathcal{H}(\mu|\mathbf{m}_X) + \mathcal{H}(\nu|\mathbf{m}_X) \right], \quad (4.4.9)$$

where \mathbf{m}_X stands for the spatial marginal of \mathbf{m} .

As in the classical case, we managed to show that

$$\lim_{T \rightarrow \infty} \mathcal{C}_T(\mu, \nu) = \mathcal{H}(\mu | \mathbf{m}_X) + \mathcal{H}(\nu | \mathbf{m}_X) < +\infty. \quad (4.4.10)$$

In fact, we have been able to show that the convergence in (4.4.10) happens exponentially fast and with the same exponential rate κ appearing in (4.4.8) and that this implies an entropic Talagrand inequality: I do not report on these results for the sake of brevity and refer the interested reader to the preprint [?].

Proof strategy A general idea to obtain exponential speed of convergence to equilibrium for hypocoercive equations systematically exploited in [66] is that of modifying the "natural" Lyapunov function of the system by adding some extra terms in such a way that proving exponential dissipation becomes an easier task. For the Langevin dynamics, a suitable modification of the natural Lyapunov functional, that is the relative entropy $\mathcal{H}(\cdot | \mathbf{m})$, is obtained considering

$$\mu \mapsto a\mathcal{H}(\mu | \mathbf{m}) + \mathcal{I}(\mu)$$

for a carefully chosen constant $a > 0$. Emulating Bakry-Émery Γ -calculus [15] it is possible to show that the modified Lyapunov functional decays exponentially along solutions of the kinetic Fokker-Planck equation. Our proof of the turnpike property consists in implementing this abstract idea on the fg -decomposition of the entropic interpolation, as we now briefly explain. Indeed, in order to bound $\mathcal{H}(\mu_t^T | \mathbf{m})$ one is naturally led to consider the backward and forward correctors as usual

$$c_b(t) := \int_{\mathbb{R}^{2d}} |\nabla \log f_s|^2 f_s g_s \, d\mathbf{m}, \quad (4.4.11a)$$

$$c_f(t) := \int_{\mathbb{R}^{2d}} |\nabla \log g_s|^2 f_s g_s \, d\mathbf{m}. \quad (4.4.11b)$$

However, it is not clear how to obtain a differential inequality ensuring exponential (forward) dissipation of (4.4.11a) and exponential (backward) dissipation of (4.4.11b) as I did in the context of (CSP). However, we showed that it is possible to find two norms $|\cdot|_{M^{-1}}$ and $|\cdot|_{N^{-1}}$, that are equivalent to the Euclidean norm and such that if we define

$$\tilde{c}_b(t) := \int_{\mathbb{R}^{2d}} |\nabla \log f_s|_{N^{-1}}^2 f_s g_s \, d\mathbf{m} \quad \text{and} \quad \tilde{c}_f(t) := \int_{\mathbb{R}^{2d}} |\nabla \log g_s|_{M^{-1}}^2 f_s g_s \, d\mathbf{m}, \quad (4.4.12)$$

then $\tilde{c}_b(t)$ and $\tilde{c}_f(t)$ satisfy the desired exponential estimates. To complete the proof, one needs to take care of the boundary conditions. This part is non trivial as it demands to prove certain regularity properties of the fg -decomposition and it is accomplished in two steps: we first show a regularizing property of entropic interpolations, namely that if $\mathcal{H}(\mu | \mathbf{m}_X), \mathcal{H}(\nu | \mathbf{m}_X)$ are finite, then the Fisher information $\mathcal{I}(\mu_t^T)$ is finite for any $t \in (0, T)$. The proof of this property is based on a gradient bound obtained in [146] and is of independent interest. The second step consists in showing that for a fixed small δ , $\varphi^T(\delta)$ and $\psi^T(T - \delta)$ can be controlled with by the sum of $\mathcal{I}(\mu_\delta^T)$ and $\mathcal{I}(\mu_{T-\delta}^T)$. We prove this estimate adapting an argument used in [206] in the analysis of deterministic finite dimensional control problems.

Beyond quasilinear potentials Exponential dissipation of the entropy for the kinetic Fokker Planck equation is known to hold under weaker assumptions than (4.4.7). For example, [66, 15] obtain exponential estimates assuming bounded Hessian only, and in [65] this assumption is further relaxed. It is natural to wonder whether the turnpike property holds under a weaker assumption (4.4.7). So far we have obtained some partial results in this direction. For example, we know that if we impose some conditions on the target measure ν than we can work assuming only a bounded Hessian. However, the picture is not fully clear yet and for this reason I prefer not to discuss these improvements here.

4.5 Turnpike estimates in Wasserstein distance for non uniformly convex potentials

The entropic turnpike estimates obtained in the former sections hold under "pointwise" assumptions, namely the Bakry-Émery curvature condition for (CSP), the convexity of the interaction potential W in (MFSP), and the

convexity of U in **(KSP)**. It is natural to wonder whether or not the turnpike property still holds if we drop these rather strong local convexity (curvature) assumptions and replace them with weaker notions of convexity that we may call "averaged" or "integrated". A first difficulty in bringing affirmative answers is that we showed in [77] that the corrector estimate (4.2.4) is indeed *equivalent* to the Bakry-Émery condition. This suggests that entropic turnpike estimates are not to be expected in this more general context and that we have to find another suitable way to measure the distance from the invariant measure \mathbf{m} that allows for exponential turnpike estimates. A possible remedy to this issue is to draw inspiration from the recent advances in the understanding of the trend to equilibrium of the Fokker-Planck equation obtained by means of *coupling by reflection*, see e.g. [110, 111]. Let me take some time to illustrate the construction of coupling by reflection.

Coupling by reflection is a probabilistic technique to prove contraction estimates in (modified) Wasserstein distance of order one for uncontrolled stochastic differential equations in absence of local pointwise convexity. For simplicity, I focus on Kolmogorov diffusions on \mathbb{R}^d of the form

$$dX_t = -\nabla U(X_t)dt + dB_t, \quad (4.5.1)$$

even though the method has wider scope and applicability. For given probability measures μ, ν with finite first moment, the coupling by reflection of two solutions of (4.5.1) with initial distributions μ and ν respectively is a diffusion process $(X_t^\mu, X_t^\nu)_{t \geq 0}$ on $\mathbb{R}^d \times \mathbb{R}^d$ such that $X_0^\mu \sim \mu$, $X_0^\nu \sim \nu$ and that there exists a Brownian motion B_t with the property that

$$\begin{aligned} dX_t^\mu &= -\nabla U(X_t^\mu)dt + dB_t \\ dX_t^\nu &= -\nabla U(X_t^\nu)dt + (\mathbf{Id} - 2e_t e_t^T) dB_t \end{aligned}$$

where

$$e_t := Z_t / |Z_t|, \quad Z_t := X_t^\mu - X_t^\nu.$$

Since the process

$$\tilde{B}_t = \int_0^t (\mathbf{Id} - 2e_s e_s^T) dB_s$$

is a Brownian motion, coupling by reflection is indeed a coupling, in the sense that its first component has the law of the diffusion process (4.5.1) started at μ and the second component has the law of the diffusion process (4.5.1) started at ν . Loosely speaking, the increments of the Brownian motion \tilde{B} are obtained by decomposing the Brownian increments of B into a vector parallel to $X_t^\mu - X_t^\nu$ and a vector orthogonal to it. The orthogonal component is kept as it is, whereas the parallel component is taken with the opposite sign. As I said above, the main interest of coupling by reflection is that it helps in obtaining contraction estimates even when U is not convex. To make this more precise define

$$\kappa(r) := \inf \left\{ 2 \frac{\langle \nabla U(x) - \nabla U(y), x - y \rangle}{|x - y|^2} : x, y \in \mathbb{R}^d, |x - y| = r \right\}. \quad (4.5.2)$$

In a very interesting article [110] A. Eberle has shown that if the conditions

$$\liminf_{r \rightarrow +\infty} \kappa(r) > 0, \quad \int_0^1 r \kappa^-(r) dr < +\infty.$$

are met, then there exist $\kappa > 0$ and a concave increasing function f which is equivalent to the identity and such that for any μ, ν with finite first moment we have

$$\forall t > 0, \quad \frac{d}{dt} \mathbb{E} [f(|X_t^\mu - X_t^\nu|)] \leq -\kappa \mathbb{E} [f(|X_t^\mu - X_t^\nu|)].$$

With some work, it is possible to deduce from this bound there exist a universal constant C such that

$$W_1(\text{Law}(X_t^\mu), \text{Law}(X_t^\nu)) \leq C \exp(-\kappa t) W_1(\mu, \nu),$$

where W_1 is the Wasserstein distance of order 1. In particular this last estimate implies the exponential convergence of the semigroup generated by (4.5.1) towards \mathbf{m} .

Coupling by reflection and turnpike property In this section, I report on ongoing work with L.Pfeiffer [92], where we develop a probabilistic approach to exponential turnpike estimates for stochastic control problems. In order to keep a conceptual coherence throughout the manuscript, I shall focus in this section only on applications to Schrödinger-like problems, and briefly discuss other kinds of stochastic control problems in the next section, dedicated to illustrate possible future research directions. To fix ideas, consider the classical Schrödinger problem (**CSP**) in which the terminal constraint $X_T \sim \nu$ is replaced by a terminal Lipschitz cost $\varphi(\cdot)$. That is to say, I am now looking at the following problem

$$\begin{aligned} \inf_{\alpha} \mathbb{E} \left[\int_0^T \frac{1}{2} |\alpha_t|^2 + \varphi(X_T^\alpha) \right] \\ \text{s.t.} \begin{cases} dX_t^\alpha = -\nabla U(X_t^\alpha) dt + \alpha_t dt + dB_t \\ X_0^\alpha \sim \mu \end{cases} \end{aligned} \quad (4.5.3)$$

Concerning the potential U , I relax strong convexity assuming that there exist $R, \kappa', L > 0$ such that

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq \begin{cases} \kappa' |x - y|^2, & \text{if } |x - y| \geq R \\ -L |x - y|^2 & \text{if } |x - y| \leq R \end{cases} \quad (4.5.4)$$

Note that, in contrast with the control problems considered in Chapter 1, problem (4.5.3) is in strong formulation, i.e. minimization is over controls and not path probability measures. This choice serves better the upcoming presentation and although there are in general some subtleties related to the different formulations, in the present context there is no issue and the two formulations are equivalent. Moreover, let me underline that the reason why I consider this modification of (**CSP**), which is interesting in its own right, is because we are not yet able to deal with the original Schrödinger problem. Indeed, although (**CSP**) can be seen to be in the form (4.5.3) for some penalty function φ , it cannot be guaranteed that φ is Lipschitz for an arbitrarily chosen target measures ν . Here is an example of the types of results we have been able to obtain so far.

Theorem 4.5.1. *Let μ admit first moment. Moreover, let U satisfy (4.5.4), φ be a smooth Lipschitz function. Then there exist universal constants $C, \kappa > 0$ depending on U and φ only such that if $\bar{\alpha}^T$ is the optimal control in (4.5.3), then we have*

$$\forall t > 0, \quad W_1(\text{Law}(X_t^{\bar{\alpha}^T}), \mathbf{m}) \leq C(\exp(-\kappa t) + \exp(-\kappa(T-t))). \quad (4.5.5)$$

The proof of this theorem is in two steps, each one involving a coupling argument. Let me just present the main ideas: the standard approach to (4.5.3) via the dynamic programming principle and the Hamilton-Jacobi-Bellman equation yields that the optimal control is Markovian and given by the gradient of the value function, i.e.

$$\bar{\alpha}_t^T(x) = -\nabla \varphi_t(x),$$

where $\varphi_t(x)$ is the optimal value of the problem that one obtains choosing $\mu = \delta_x$ and replacing T with $T - t$ in (4.5.3). Whereas a rough upper bound for $|\nabla \varphi_t(x)|$ is classically obtained by means of synchronous coupling, such estimate is not enough in view of the turnpike property. We observed that replacing synchronous coupling with a suitable slight modification of coupling by reflection in this classical argument yields the bound

$$\forall x \in \mathbb{R}^d, t \leq T, \quad |\nabla \varphi_t(x)| \lesssim \exp(-\kappa(T-t)), \quad (4.5.6)$$

that is the first important step towards establishing the turnpike property. In the above, we used the notation \lesssim to indicate that the inequality has to be understood up to an universal multiplicative constant. The second step of the proofs consists in using once again coupling by reflection between the optimally controlled dynamics corresponding to $\alpha_t(x) = -\nabla \varphi_t(x)$, and the uncontrolled dynamics corresponding to $\alpha_t(x) \equiv 0$. The analysis of this coupling cannot be done exactly as in [110] because the two diffusion processes do not have the same drift field, but thanks to (4.5.6) an adaptation of the original argument yields (4.5.5). The proof scheme I have just sketched is in fact quite robust and does not only apply to (4.5.3): I will rapidly discuss other control problems that are susceptible to be analysed

in an analogous way in the next section. Restricting the attention to Schrödinger and in particular to **(MFSP)** when the diffusion coefficient is constant and the drift is of the form

$$b(\mu, x) = -\nabla W * \mu(x) - \nabla U(x)$$

we can obtain results analogous to (4.5.1) in the following scenarios

- When $U = 0$ and W is strongly convex outside a ball, or when both potential are strongly convex outside a ball.
- When W is and its Hessian are bounded, U is strongly convex outside a ball and the bound on the Hessian of W is small enough.

4.6 Perspectives

Beyond the toy model A large part of the techniques developed to establish the results in this chapter are arguably applicable to a number of other interesting situations, but some analogy between the stochastic control problem at hand and the toy model must be present. The simplest and probably most commonly encountered stochastic control problem whose different geometric structure than **(TSP)** is the following generalization of the classical LQ setting:

$$\begin{aligned} \inf_{\alpha} \mathbb{E} \left[\int_0^T \frac{1}{2} |\alpha_t|^2 + V(X_t^\alpha) dt + g(X_T^\alpha) \right] \\ \text{s.t. } dX_t^\alpha = \alpha_t dt + dB_t, \quad X_0^\alpha \sim \mu. \end{aligned} \tag{4.6.1}$$

For this problem, it is not too difficult to show that an exponential turnpike property holds as soon as the running cost V is uniformly convex. Interestingly enough, the proof in this case is quite different than the one based on the corrector estimate (4.2.4) that underlies most of the turnpike estimates presented in this chapter and seems closer in spirit to the proof strategy of [206] in the deterministic case. At the general level, one expects that, because of the presence of the Brownian motion, for stochastic control problems global turnpike estimates (i.e. valid for a very large class of initial and terminal conditions) should be more accessible than for the deterministic case. However, how to make this rigorous is not so clear. In [92] we began to address this problem. By now we have understood that, at least for stochastic control problems of the form (4.6.1), one can establish the exponential turnpike property even if V is not necessarily convex using an argument based on coupling by reflection. More precisely, the turnpike property holds as soon as $V = V_0 + V_1$, where V_0 is strictly convex and V_1 is Lipschitz. This is rather encouraging and we are currently developing this research line. One of the next steps would be to include a mean field term of the form $W * \text{Law}(X_t^\alpha)(X_t^\alpha)$ in the running cost and try to see what kind of turnpike property, if any, can be established. It is worth pointing out that the use of coupling by reflection we made in the analysis of (4.6.1) is quite different from the one I explained in section 4.5 that is effective to treat Schrödinger like problems. In particular, for (4.6.1) what one aims at showing is not that "the optimal control is small" as in (4.5.6), but rather that "it is the gradient of an almost convex function".

Lattice gases In [72] we provided a formal argument for the convexity of the free energy along entropic interpolations under the generalized McCann's condition of [62]. Even though the convexity estimate we obtained is rather qualitative than quantitative, it is still an encouraging result in view of establishing turnpike estimates and novel functional inequalities.

Part II

Markov chains

Convex entropy decay for Markov chains

Contents

5.1	Functional inequalities for discrete Markov chains	59
5.1.1	Trying to imitate Bakry-Émery	60
5.1.2	Discrete convex Sobolev inequalities	61
5.2	Coupling rates and convex entropy decay	63
5.2.1	Setup and main assumptions	63
5.2.2	Coupling rates and second derivative of the entropy	64
5.3	A sufficient condition for interacting random walks on \mathbb{N}^d	65
5.3.1	Equivalence with Wasserstein contraction	69
5.3.2	Examples	69
5.4	Hardcore models	71
5.5	Perspectives	72
5.5.1	Overcoming the limitations of convex entropy decay	72
5.5.2	Relations with entropic curvature and $\text{CD}(\kappa, +\infty)$	72

5.1 Functional inequalities for discrete Markov chains

A recurrent theme of this manuscript is that "entropy is convex". In different occasions, we have shown that various kinds of entropy (Boltzmann entropy, free energy,...) are convex along various kinds of curves (geodesics, gradient flows, Schrödinger bridges...). Many of these results rely on *pointwise* comparisons between differential operators, such as $\Gamma_2(f, f) \geq \kappa \Gamma(f, f)$. Given how well Γ -calculus and optimal transport techniques work for proving convex interpolation estimates, it is a most natural question to ask whether similar ideas can be successfully applied in the context of Markov chains on countable spaces in particular because convexity of the entropy implies many important functional inequalities, such as the (modified) logarithmic Sobolev inequality, henceforth MLSI. It turns out that this task is rather complicated and many open problems still remain. In the article [82], I have developed a method to prove convexity of the entropy along the marginal flow of a continuous time Markov chain. In a nutshell, this method attempts at injecting the probabilistic intuition behind the construction of a contractive coupling into the algebraic problem of comparing the multitude of terms generated by differentiating twice the entropy. With respect to convex entropy decay, to the best of my knowledge, the sufficient conditions I found match and sometimes improve on the best available results. Convex entropy decay implies lower bounds on the best constants in convex Sobolev inequalities, such as the modified logarithmic Sobolev inequality and Beckner inequalities. To the best of my knowledge, these bounds are sometimes better than the best bounds available in the literature, and this especially true in some *non perturbative* situations, as I will explain below. On the other hand, perturbative criteria, that typically apply to models coming statistical mechanics in the high temperature regime, often outperform the bounds I obtained. This is not surprising: convexity of the entropy is a sufficient, but by far not necessary condition for the modified logarithmic Sobolev inequality, and is therefore natural that it holds under more restrictive assumptions.

Convexity of the entropy for Markov Chains The first articles I am aware of where convex entropy decay for Markov chains is established have been written by Caputo, Dai Pra, Posta and coauthors [41, 45, 103]. In these works they developed a method based on a discrete analogous of Bochner's identity and obtained estimates on the spectral gap and the best constant in the modified logarithmic Sobolev inequality for a large class of non local dynamics, whereas the more general problem of defining a notion of discrete Ricci curvature has been tackled in [181, 173, 142, 196]. In particular, the notion of entropic Ricci curvature put forward in [170, 115] has deep implications in terms of functional inequalities. Explicit lower bounds for the entropic Ricci curvature in concrete examples have been recently obtained in [117, 119] and [114]. More recently similar ideas have been applied in [113] in the context of non linear Markov chains. Entropic curvature has also been used to obtain convexity bounds for general ϕ -entropies in [152].

5.1.1 Trying to imitate Bakry-Émery

My main motivation for writing [82] was twofold.

- One of my goals was to find an analogous statement for probability measures on the lattice \mathbb{N}^d of the fact that strongly log-concave probability measures on \mathbb{R}^d satisfy the logarithmic Sobolev inequality with a positive constant, i.e.

$$\mathbf{m} = \exp(-U), \quad \nabla^2 U \geq \kappa \Rightarrow \mathbf{LSI}(\kappa). \quad (5.1.1)$$

Strong convexity is a geometric criterion that is non perturbative, in the sense that it is satisfied by probability measures that may be *far* from being product measures. On the contrary, most results for continuous time Markov chains are perturbative in spirit, ensuring positive lower bounds on the MLSI constant only if the interaction is small and \mathbf{m} is almost a product measure. In light of these observations, it is very natural to seek for non perturbative sufficient conditions on the generator of a continuous time Markov chain on \mathbb{N}^d implying MLSI. To the best of my knowledge, such results have only been obtained for $d = 1$, with the exception of some two dimensional examples treated in [103]. It turns out that the use of coupling rates enables to lift the obstacles that have limited non perturbative criteria in a number of interesting situations. In particular I am presenting below a sufficient non perturbative condition for MLSI and general convex Sobolev inequalities for random walks on \mathbb{N}^d that is valid for any value of d . As a corollary, we obtain that multiplying a multidimensional Poisson distribution by a density of the form $\exp(-V)$ yields a probability measure satisfying MLSI if a local condition at the origin holds and the (discrete) Hessian of the potential V has non negative entries, see condition (5.3.11). This creates a curious parallelism with the above mentioned result for probability measures on \mathbb{R}^d , where it is the non negativity of $\text{Hess } V$ as a quadratic form that plays an essential role. In section 5.5 I will present a natural conjecture on the relation between $\text{CD}(\kappa, +\infty)$ and the condition I found.

- The second main motivation that drove me was to show that if one is able to construct a contractive coupling, then MLSI holds. Once again, this is well known and understood for diffusions on manifolds. More precisely, one has the following, that I take and adapt from [211].

Theorem 5.1.1. *Let M be a smooth compact Riemannian manifold without boundary and $\mathcal{L} = \frac{1}{2}\Delta - \nabla U \cdot \nabla$. The following are equivalent*

- (i) *For any $x, y \in M$, we can construct a coupling $(X_t^x, X_t^y)_{t \geq 0}$ of the diffusion process with generator \mathcal{L} started at x and of the diffusion process with generator \mathcal{L} started at y in such a way that*

$$\mathbb{E}[d_M(X_t^x, X_t^y)] \leq \exp(-\kappa t) d_M(x, y) \quad \forall t > 0.$$

- (ii) *For any $x, y \in M$, we can construct a coupling $(X_t^x, X_t^y)_{t \geq 0}$ of the diffusion process with generator \mathcal{L} started at x and of the diffusion process with generator \mathcal{L} started at y in such a way that*

$$\text{almost surely, } d_M(X_t^x, X_t^y) \leq \exp(-\kappa t) d_M(x, y) \quad \forall t > 0.$$

- (iii) *The condition $\text{CD}(\kappa, +\infty)$.*

(iv) Let \mathbf{m} be the reversible measure for \mathcal{L} . Then the relative entropy $\mathcal{H}(\cdot|\mathbf{m})$ is κ -convex along displacement interpolations (4.2.2).

As a corollary, (i) implies $\mathbf{LSI}(\kappa)$.

The above result is quite beautiful, and also quite surprising. Indeed, at first glance one might think that (i) is a much weaker condition than (ii) and this is indeed the case in the context of Markov chains, at least for any reasonable choice of distance one might think of. Theorem 5.1.1 served as an inspiration for designing various notions of Ricci curvature for Markov chains. The main difficulty in finding the good definition is that of striking a good balance between the desire to have at the same time a wide range of applicability and powerful implications. For example, the notion of coarse Ricci curvature [181] is based on item (i) and therefore it is relatively easy to produce lower bounds but it is not known to imply MLSI or Beckner inequalities, although it does imply Poincaré inequality. On the other hand, the notion of entropic curvature [115], based on item (iv) takes care of this issue but it is much harder to come up with positive lower bounds. In particular, the distance that is used there to construct geodesics between probability measures is not a very simple object to handle. Although I did not even attempt at defining a new notion of curvature, I managed to establish that, at least in the setting of random walks on \mathbb{N}^d , a geometric condition on the jump rates that may be seen as analogous to item (iii) implies MLSI and general convex Sobolev inequalities and is equivalent to the possibility of constructing a certain coupling whose contractive properties can be placed somehow in between (i) and (ii). The literature about general convex Sobolev inequalities for Markov chains is not abundant, see [152, 36]. Therefore, in many of the examples I analyzed, the lower bounds that I obtained seem to be new and the best available so far. Concerning MLSI, I can sometimes improve on the best known estimates I am aware of and provide new sufficient non-perturbative conditions that were not known before. However, as I have already said, one can in many situations obtain better bounds on the MLSI constant by means of other methods and this is essentially because convex entropy decay is a much stronger condition than the MLSI.

I now move on to show some of the results I obtained in a more quantitative and less descriptive fashion.

5.1.2 Discrete convex Sobolev inequalities

Let me spend some time to introduce ϕ -entropies and discrete convex Sobolev inequalities for Markov chains. I begin by considering a continuous time Markov chain on a countable state space Ω , whose infinitesimal generator \mathcal{L} takes the form

$$\mathcal{L}f(\eta) = \sum_{\sigma \in G} c(\eta, \sigma) \nabla_{\sigma} f(\eta), \quad (5.1.2)$$

where G is a collection of maps $\sigma : \Omega \rightarrow \Omega$ called moves, $c : \Omega \times G \rightarrow \mathbb{R}_{\geq 0}$ are the transition rates and $\nabla_{\sigma} f(\eta)$ denotes the discrete gradient $\nabla_{\sigma} f(\eta) = f(\sigma\eta) - f(\eta)$. I will also assume that the Markov chain is reversible and denote \mathbf{m} the reversible measure. Given a convex function $\phi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ and a positive function $f : \Omega \rightarrow \mathbb{R}_{> 0}$ the ϕ -entropy $\mathcal{H}^{\phi}(f|\mathbf{m})$ of f is defined as follows:

$$\mathcal{H}^{\phi}(f|\mathbf{m}) = \sum_{\eta \in \Omega} \phi(f(\eta)) \mathbf{m}(\eta) - \phi\left(\sum_{\eta \in \Omega} f(\eta) \mathbf{m}(\eta)\right). \quad (5.1.3)$$

A typical question of interest is to estimate the best constant κ_{ϕ} such that

$$\mathcal{H}^{\phi}(P_t f|\mathbf{m}) \leq \exp(-\kappa_{\phi} t) \mathcal{H}^{\phi}(f|\mathbf{m}) \quad (5.1.4)$$

holds uniformly on $f > 0$ and $t > 0$. In the above, I denoted by P_t the Markovian semigroup generated by \mathcal{L} . It is well known that (5.1.4) is equivalent to the *convex Sobolev inequality*

$$\forall f > 0, \quad \kappa_{\phi} \mathcal{H}^{\phi}(f|\mathbf{m}) \leq \mathcal{E}(\phi'(f), f), \quad (5.1.5)$$

where $\mathcal{E}(f, g)$ is the Dirichlet form

$$\mathcal{E}(f, g) = - \sum_{\eta \in \Omega} f(\eta) (\mathcal{L}g)(\eta) \mathbf{m}(\eta).$$

The family of convex Sobolev inequalities is quite rich. Indeed, defining

$$\phi_\alpha(a) = \begin{cases} \frac{1}{\alpha-1}(a^\alpha - a) - a + 1, & \text{if } \alpha \in (1, 2], \\ a \log a - a + 1, & \text{if } \alpha = 1. \end{cases}$$

we get that (5.1.5) is the *modified logarithmic Sobolev inequality* (henceforth MLSI) for $\alpha = 1$,

$$\forall f > 0, \quad \kappa_1 \mathcal{H}^{\phi_1}(f|\mathbf{m}) \leq \mathcal{E}(\log f, f). \quad (5.1.6)$$

For $\alpha = 2$ we recover the *Poincaré inequality*, whereas for $\alpha \in (1, 2)$ we find the family of (discrete) *Beckner inequalities* [16, 17].

$$\forall f > 0, \quad \kappa_\alpha \mathcal{H}^{\phi_\alpha}(f|\mathbf{m}) \leq \frac{\alpha}{\alpha-1} \mathcal{E}(f^{\alpha-1}, f). \quad (5.1.7)$$

In [82], I have endeavoured to establish the estimate

$$\frac{d^2}{dt^2} \mathcal{H}^\phi(P_t f|\mathbf{m}) \geq \kappa_\phi \mathcal{E}(\phi'(P_t f), P_t f) \quad (5.1.8)$$

that implies (5.1.4). To see why this is the case we recall that, at least formally we have

$$\frac{d}{dt} \mathcal{H}^\phi(P_t f|\mathbf{m}) = -\mathcal{E}(\phi'(P_t f), P_t f).$$

Therefore, we obtain from (5.1.8) and Gronwall's lemma that $\frac{d}{dt} \mathcal{H}^\phi(P_t f|\mathbf{m}) \rightarrow 0$. From this, (5.1.5) follows integrating (5.1.8) over $[t, \infty)$ provided that $\mathcal{H}^\phi(P_t f|\mathbf{m}) \rightarrow 0$.

Probabilistic approach to convex entropy decay As I said above, the method I developed for establishing (5.1.8) is based on the notion *coupling rates*, that I am now going to define. To do so, it is convenient to augment the set G with a null element e , defined by $e\eta = \eta \forall \eta \in \Omega$. I also set $G^* = G \cup \{e\}$,

Definition 5.1.1. Let $\eta, \bar{\eta} \in \Omega$ and \mathcal{L} as in (5.1.2). We say that $\mathbf{c}^{\text{cpl}}(\eta, \bar{\eta}, \cdot, \cdot) : G^* \times G^* \rightarrow \mathbb{R}_{\geq 0}$ are coupling rates for $(\eta, \bar{\eta})$ if

$$\begin{aligned} \forall \gamma \in G, \quad \sum_{\bar{\gamma} \in G^*} \mathbf{c}^{\text{cpl}}(\eta, \bar{\eta}, \gamma, \bar{\gamma}) &= c(\eta, \gamma), \\ \forall \bar{\gamma} \in G, \quad \sum_{\gamma \in G^*} \mathbf{c}^{\text{cpl}}(\eta, \bar{\eta}, \gamma, \bar{\gamma}) &= c(\bar{\eta}, \bar{\gamma}). \end{aligned} \quad (5.1.9)$$

If coupling rates are available for any pair $(\eta, \bar{\eta})$ then one can define a Markov generator \mathcal{L}^{cpl} acting on $F : \Omega \times \Omega \rightarrow \mathbb{R}$ as follows

$$\mathcal{L}^{\text{cpl}} F(\eta, \bar{\eta}) = \sum_{\gamma, \bar{\gamma} \in G^*} \mathbf{c}^{\text{cpl}}(\eta, \bar{\eta}, \gamma, \bar{\gamma}) \nabla_{\gamma, \bar{\gamma}} F(\eta, \bar{\eta}),$$

where $\nabla_{\gamma, \bar{\gamma}} F(\eta, \bar{\eta}) := F(\gamma\eta, \bar{\gamma}\bar{\eta}) - F(\eta, \bar{\eta})$. A Markov chain on $\Omega \times \Omega$ with generator \mathcal{L}^{cpl} started at $(\eta, \bar{\eta})$ indeed realizes a coupling of a Markov chain with generator \mathcal{L} started at η and of a Markov chain with generator \mathcal{L} started at $\bar{\eta}$. The vast majority of convexity estimates for the entropy in the context of Markov chain rests on establishing some kind of discrete versions of Bochner's identity following the technique put forward [41, 45]. Doing so, typically requires some non trivial algebraic manipulations. What I proposed to do in [82] is indeed a probabilistic *ansatz* to exploit the probabilistic intuition we have on the dynamics of the Markov chain to perform these algebraic passages. Although I do not yet have a general algorithm for producing an efficient coupling in view of obtaining (5.1.8), there

are some general guidelines. In particular, as one may expect, it is often convenient to construct the coupling rates in such a way that the associated Markov chain on $\Omega \times \Omega$ reaches as quickly as possible the diagonal $\{(\eta, \eta) : \eta \in \Omega\}$ and if it starts from the set $\{\eta, \bar{\eta} : \exists \sigma \in G \text{ s.t. } \bar{\eta} = \sigma\eta\}$, it never leaves it. Very recently, coupling techniques have found important applications in the area of discrete functional inequalities and I would like to mention two results that I find particularly interesting. The first one is due to Hermon and Salez [148] who used couplings in combination with the so called martingale method in order to obtain the best known estimate on the MLSI constant for inhomogeneous zero range processes. In the context of spin systems, a strong link between the existence of contractive couplings and the best constant in MLSI has been established [34]. Both these results are perturbative in spirit as they rely on an approximate tensorization property of the relative entropy.

5.2 Coupling rates and convex entropy decay

In this section I give some basic definitions and present simple but rather general results on how to use coupling rates to obtain convexity estimates for the evolution of entropic functionals.

5.2.1 Setup and main assumptions

Given a state space Ω that is at most countable, a finite set of moves G , and non negative transition rates $c(\eta, \sigma)$ I consider the formal generator (5.1.2) and make the basic assumption that

- The set G is finite. \mathcal{L} is irreducible and admits an invariant probability measure $\mathbf{m} \in \mathcal{P}(\Omega)$ that satisfies

$$\sum_{\eta \in \Omega, \sigma \in G} c(\eta, \sigma) \mathbf{m}(\eta) < +\infty.$$

It is well known, see for instance [179], that under this hypothesis the invariant measure is unique and for any initial $\eta \in \Omega$ there exists a continuous time Markov chain $(X_t^\eta)_{t \geq 0}$ whose infinitesimal generator is \mathcal{L} and such that $X_0^\eta = \eta$. Moreover, $(X_t^\eta)_{t \geq 0}$ is non-explosive. Following closely [103] we also assume that \mathbf{m} is reversible for \mathcal{L} and that each move admits an "inverse". More precisely

- There exists an involution

$$\begin{aligned} G &\longrightarrow G \\ \sigma &\longmapsto \sigma^{-1} \end{aligned}$$

such that $\sigma^{-1}(\sigma(\eta)) = \eta$ holds whenever $\mathbf{m}(\eta)c(\eta, \sigma) > 0$ and

$$\sum_{\substack{\eta \in \Omega \\ \sigma \in G}} F(\eta, \sigma) c(\eta, \sigma) \mathbf{m}(\eta) = \sum_{\substack{\eta \in \Omega \\ \sigma \in G}} F(\sigma\eta, \sigma^{-1}) c(\eta, \sigma) \mathbf{m}(\eta) \quad (5.2.1)$$

holds for all bounded $F : \Omega \times G \longrightarrow \mathbb{R}$.

Using reversibility, it is not hard to see that for a given $\phi : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$, the convex Sobolev inequality (5.1.5) holds with constant κ_ϕ if and only if

$$\forall f > 0, \quad \mathcal{H}^\phi(f|\mathbf{m}) \leq \frac{\kappa_\phi}{2} \sum_{\substack{\eta \in \Omega \\ \sigma \in G}} \nabla_\sigma(\phi' \circ f) \nabla_\sigma f(\eta) c(\eta, \sigma) \mathbf{m}(\eta). \quad (5.2.2)$$

In view of this expression, it is convenient to introduce the function Φ defined by

$$\Phi : \mathbb{R}_{>0}^2 \longrightarrow \mathbb{R}_{>0}, \quad \Phi(a, b) := (\phi'(b) - \phi'(a))(b - a).$$

A natural assumption for the method to work is the following

$$\phi \text{ is convex and the function } \Phi \text{ is also convex.} \quad (5.2.3)$$

When $\phi = \phi_\alpha$, the function Φ is denoted Φ_α . It is not hard to see that Φ_α satisfies (5.2.3).

5.2.2 Coupling rates and second derivative of the entropy

Let me now show how one can use coupling rates to organize the terms originating from differentiating $\mathcal{E}(\phi'(P_t f), P_t f)$ and find appropriate upper bounds in view of establishing (5.1.8). In what follows, for a differentiable function $(a, b) \mapsto \Phi(a, b)$ I denote by $D\Phi(a, b)$ the Jacobian, i.e. the 1×2 matrix $[\partial_a \Phi(a, b), \partial_b \Phi(a, b)]$. I also use \cdot for the standard matrix-vector product. In the next Lemma one can see that, for each admissible choice of coupling rates we have a different equivalent way of rewriting the second derivative of the entropy. Using the convexity of Φ , we then obtain a first general lower bound. To write many of the upcoming expressions in a more compact way it is convenient to introduce the set

$$\mathcal{S} = \{(\eta, \sigma) : c(\eta, \sigma) > 0\}.$$

Lemma 5.2.1. *Let $\{\mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \cdot, \cdot)\}_{(\eta, \sigma) \in \mathcal{S}}$ be coupling rates. For all $f > 0$ define f^ϕ as*

$$f^\phi : \Omega \times \Omega \longrightarrow \mathbb{R}_{\geq 0}, \quad f^\phi(\eta, \bar{\eta}) = \Phi(f(\eta), f(\bar{\eta})).$$

We have:

$$\frac{d}{dt} 2\mathcal{E}(\phi'(P_t f), P_t f) \Big|_{t=0} = \sum_{\substack{(\eta, \sigma) \in \mathcal{S} \\ \gamma, \bar{\gamma} \in G^*}} c(\eta, \sigma) \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma}) D\Phi(f(\eta), f(\sigma\eta)) \cdot \begin{bmatrix} \nabla_\gamma f(\eta) \\ \nabla_{\bar{\gamma}} f(\sigma\eta) \end{bmatrix} \mathbf{m}(\eta). \quad (5.2.4)$$

Consequently,

$$\begin{aligned} \frac{d}{dt} 2\mathcal{E}(\phi'(f_t), f_t) \Big|_{t=0} &\leq \sum_{\substack{(\eta, \sigma) \in \mathcal{S} \\ \gamma, \bar{\gamma} \in G^*}} c(\eta, \sigma) \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma}) \nabla_{\gamma, \bar{\gamma}} f^\phi(\eta, \sigma\eta) \mathbf{m}(\eta) \\ &\quad - \sum_{\substack{(\eta, \sigma) \in \mathcal{S} \\ \gamma, \bar{\gamma} \in G^* \\ \gamma\eta = \bar{\gamma}\sigma\eta}} c(\eta, \sigma) \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma}) \left(\nabla_{\gamma, \bar{\gamma}} f^\phi(\eta, \sigma\eta) - D\Phi(f(\eta), f(\sigma\eta)) \cdot \begin{bmatrix} \nabla_\gamma f(\eta) \\ \nabla_{\bar{\gamma}} f(\sigma\eta) \end{bmatrix} \right) \mathbf{m}(\eta). \end{aligned} \quad (5.2.5)$$

From the above lemma we immediately deduce a sufficient condition for convex entropy decay.

Corollary 5.2.1. *Let $\{\mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \cdot, \cdot)\}_{(\eta, \sigma) \in \mathcal{S}}$ be coupling rates. If*

- *There exists $\kappa' \geq 0$ such that*

$$\frac{1}{2} \sum_{\substack{(\eta, \sigma) \in \mathcal{S} \\ \gamma, \bar{\gamma} \in G^*}} c(\eta, \sigma) \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma}) \nabla_{\gamma, \bar{\gamma}} f^\phi(\eta, \sigma\eta) \mathbf{m}(\eta) \leq -\kappa' \mathcal{E}(\phi'(f), f) \quad (5.2.6)$$

holds uniformly on $f > 0$.

- *There exist $\kappa'', \kappa''' \geq 0$ such that*

$$\inf_{(\eta, \sigma) \in \mathcal{S}} \min\{\mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \sigma, \text{e}), \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \text{e}, \sigma^{-1})\} \geq \kappa'' \quad (5.2.7)$$

and

$$\inf_{(\eta, \sigma) \in \mathcal{S}} \sum_{\substack{\gamma, \bar{\gamma} \in G^* \\ \gamma\eta = \bar{\gamma}\sigma\eta}} \mathbf{c}^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma}) \geq \kappa''' \quad (5.2.8)$$

hold.

Then

- (i) *The convex Sobolev inequality (5.1.5) holds with $\kappa_\phi = \kappa'$ for all Φ satisfying (5.2.3).*

(ii) The modified logarithmic Sobolev inequality (5.1.6) holds with $\kappa_1 = \kappa' + 2\kappa''$.

(iii) For $\alpha \in (1, 2]$, the discrete Beckner inequality (5.1.7) holds with $\kappa_\alpha = \kappa' + (\alpha - 1)\kappa'''$.

Of course, this result does not say how to obtain (5.2.6). In order to do so, for most examples I treated, I used a general strategy that works in two steps. The first one is to find coupling rates $c^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma})$ such that for any η, σ the support of

$$(\gamma, \bar{\gamma}) \mapsto c^{\text{cpl}}(\eta, \sigma\eta, \gamma, \bar{\gamma})$$

can be partitioned into three subsets (see figure (5.2) for a visual explanation in a concrete example)

- The set of *parallel* terms, i.e. $\{\gamma, \bar{\gamma} : \bar{\gamma} = \gamma\}$
- The set of *rotational* terms that is divided into two subsets, the first being $\{\sigma\} \times G$ and the second being $G \times \{\sigma^{-1}\}$. We refer to figure (5.2) for a justification of this terminology.
- The set of *good* terms, $\{(\sigma, \mathbf{e}), (\mathbf{e}, \sigma^{-1})\}$.

The second step consists in using reversibility to show that the parallel terms in (5.2.5) give an overall non negative contribution and that the same holds for rotational terms. Finally, I show that the overall contribution of good terms dominates the Dirichlet form. At this point (5.2.6) is proven. Estimating precisely the contribution of rotational terms seems to be crucial to obtain on the MLSI constant in a non perturbative setting. There are some relevant examples (e.g. zero range dynamics on the complete graph) where a fourth class of terms had to be included in the analysis. In this case, the strategy I outlined needs to be adapted and typically does not produce the best possible bound on the MLSI constant. Let me now move on to illustrate what kind of results this technique can give by applying it to two concrete examples: interacting random walks and hardcore models.

5.3 A sufficient condition for interacting random walks on \mathbb{N}^d

I use the term "interacting random walk" for a general continuous time Markov chain on \mathbb{N}^d whose generator is of the form

$$\mathcal{L}f(\eta) = \sum_{i=1}^d c(\eta, \gamma_i^+) \nabla_i^+ f(\eta) + c(\eta, \gamma_i^-) \nabla_i^- f(\eta), \quad (5.3.1)$$

where for all $\eta \in \mathbb{N}^d$ and $1 \leq i \leq d$ the move γ_i^+ (resp. γ_i^-) correspond to an upward (resp. downward) jump in the i -th coordinate and the jump rates $c(\eta, \gamma_i^\pm)$ are defined by

$$c(\eta, \gamma_i^+) = \exp(-\nabla_i^+ V^+(\eta)), \quad c(\eta, \gamma_i^-) = \exp(-\nabla_i^- V^-(\eta)).$$

where $V^-, V^+ : \mathbb{N}^d \rightarrow \mathbb{R}$ are two given potentials. The reversible measure for \mathcal{L} is then given by

$$\mathbf{m}(\cdot) = \frac{1}{Z} \exp(-(V^+ + V^-(\cdot))).$$

I established convex entropy decay under the following geometric condition.

Assumption 5.3.1. For all $\eta \in \Omega$, $1 \leq i \leq d$ we have $\kappa^+(\eta, i) \geq 0$ and $\kappa^-(\eta, i) \geq 0$, where

$$\kappa^+(\eta, i) = -\nabla_i^+ c(\eta, \gamma_i^+) - \sum_{\substack{\bar{\gamma} \in G \\ \bar{\gamma} \neq \gamma_i^+, \gamma_i^-}} \max\{\nabla_i^+ c(\eta, \bar{\gamma}), 0\}, \quad (5.3.2)$$

and

$$\kappa^-(\eta, i) = \nabla_i^+ c(\eta, \gamma_i^-) - \sum_{\substack{\gamma \in G \\ \gamma \neq \gamma_i^+, \gamma_i^-}} \max\{-\nabla_i^+ c(\eta, \gamma), 0\}. \quad (5.3.3)$$

The probabilistic meaning of this condition may seem a bit difficult to catch at first sight. It will become more clear by having a look at figure 5.1 and reading the statement of Theorem 5.3.3, where I will provide an equivalent characterization in terms of couplings.

Theorem 5.3.2. *Let $V^-, V^+ : \mathbb{N}^d \rightarrow \mathbb{R}$ be given and the generator \mathcal{L} be defined by (5.3.1). Moreover, assume that Assumption (5.3.1) holds. Next, define*

$$\kappa = \inf_{\substack{\eta \in \mathbb{N}^d \\ 1 \leq i \leq d}} \kappa^+(\eta, i) + \kappa^-(\eta, i) \quad (5.3.4)$$

The following holds

- (i) *For any ϕ satisfying (5.2.3) the convex Sobolev inequality (5.2.2) holds with $\kappa_\phi = \kappa$. In particular MLSI holds with $\kappa_1 = \kappa$.*
- (ii) *For $\alpha \in (1, 2]$, the Beckner inequality (5.1.7) holds with $\kappa_\alpha = \alpha\kappa$.*

Concerning MLSI and the spectral gap, the results of Theorem 5.3.2 are well known for $d = 1$. Comparable results can be found e.g. in [119, 173, 45, 41]. For Beckner inequalities and $d = 1$, we refer to [152]. When $d > 1$ much less appear to be known. For MLSI and Poincaré inequalities perturbative sufficient conditions are given in [41, 103]. In [103], a non perturbative two dimensional example is also treated. Erbar et al. gave in [114, Thm 3.9] a general abstract sufficient condition implying positive lower bounds for the entropic Ricci curvature [115, 170]. It can be checked that this criterion, when applied to the setting of Theorem 5.3.2 provides a bound for κ_1 that is worse or the same as what Theorem 5.3.2 gives, and in some cases it may give no positive lower bounds. However, although the authors only apply their result in the weak interaction/high temperature regime, it seems that its validity extends to the non perturbative setup. In particular, it could be used to provide lower bounds for some of the examples we are going to present at section 5.3.2. To conclude this short overview of the existing literature, let me mention the work of Johnson [150], who obtains estimates on the spectral gap and MLSI constant in $d = 1$, as well as some estimates on the spectral gap (but no estimates on the MLSI constant) for $d \geq 2$. Finally, for $d \geq 2$ the results of Theorem 5.3.2 about Beckner's inequalities and general convex Sobolev inequalities seem to be new and bring some answers to the questions raised in [150, Sec. 9] and [103] about sufficient conditions for MLSI in dimension larger than one.

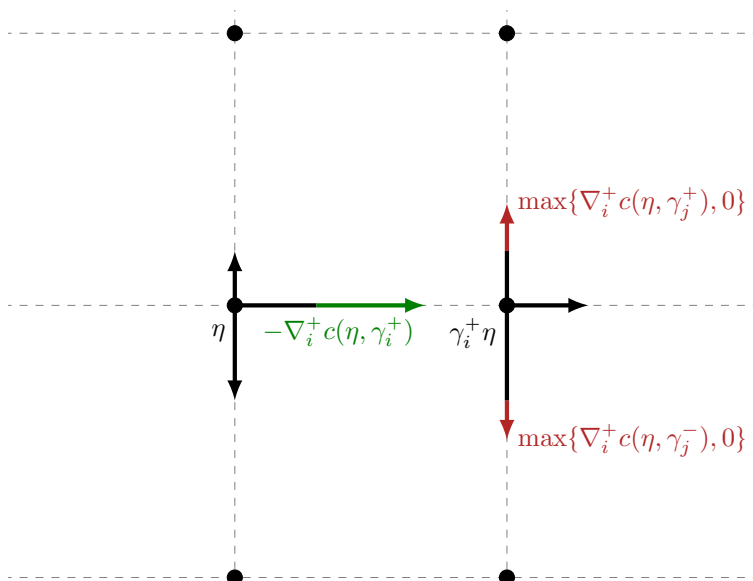


Figure 5.1 – The condition $\kappa^+(\eta, i) \geq 0$ imposes that the length of the green arrow is at least as much as the total length of the red arrows. The coupling interpretation of this condition is that the random walker starting at η can use his/her larger probability to make the γ_i^+ move in order to run after the walker starting at $\gamma_i^+ \eta$ whenever he/she tries to get at distance two from η using the moves γ_j^+, γ_j^- .

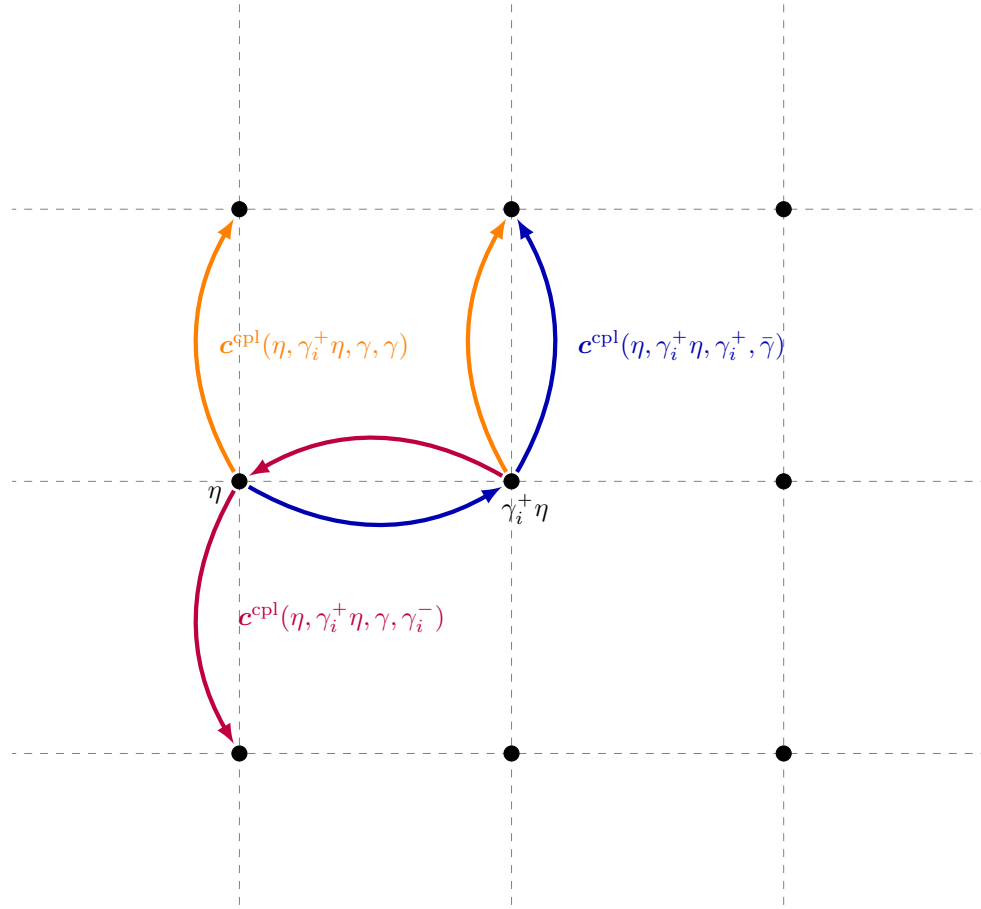


Figure 5.2 – The coupling rates in the proof of Theorem 5.3.2 couple the dynamics of two random walkers $X_t^\eta, X_t^{\gamma_i^+ \eta}$ starting at $\eta, \gamma_i^+ \eta$ respectively in such a way that, on a short time interval $[0, \varepsilon]$ only one of the following movements can be observed:

- The two walkers meet at η (resp. $\gamma_i^+ \eta$) with probability $\varepsilon \kappa^-(\eta, i)$ (resp. $\varepsilon \kappa^+(\eta, i)$). These are the *good terms*.
- The two walkers move in parallel making the same move γ with probability $\varepsilon c^{\text{spl}}(\eta, \gamma_i^+ \eta, \gamma, \gamma)$ (orange lines). These are the *parallel terms*.
- The walker $X_t^{\gamma_i^+ \eta}$ runs after the walker X_t^η with probability $\varepsilon c^{\text{spl}}(\eta, \gamma_i^+ \eta, \gamma, \gamma_i^-)$ (purple lines) or the walker X_t^η runs after the walker $X_t^{\gamma_i^+ \eta}$ with probability $\varepsilon c^{\text{spl}}(\eta, \gamma_i^+ \eta, \gamma_i^+, \gamma)$ (blue lines). These are the *rotational terms*.

5.3.1 Equivalence with Wasserstein contraction

This section is devoted to the equivalence between the condition (5.3.1) and some contraction properties of the Wasserstein distance constructed with respect to the natural graph distance of the lattice \mathbb{N}^d . Note that this distance does not coincide with the restriction to probability measures supported on \mathbb{N}^d of the usual Wasserstein distance, that is constructed using the Euclidean metric. For a given exponent $p \geq 1$ and $\eta, \bar{\eta} \in \mathbb{N}^d$, the graph distance is given by

$$d(\eta, \bar{\eta}) = \sum_{i=1}^d |\eta_i - \bar{\eta}_i|.$$

For given $\mu, \nu \in \mathcal{P}_p(\mathbb{N}^d)$ we define the Wasserstein distance of order p as

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \sum_{\eta', \eta''} d^p(\eta', \eta'') \pi(\eta', \eta'') \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is, as usual, the set of all couplings of μ and ν . Here is the announced result: to simplify the presentation, I removed an inessential technical assumption from the original statement.

Theorem 5.3.3. *Let $V^-, V^+ : \mathbb{N}^d \rightarrow \mathbb{R}$ be given and \mathcal{L} be defined by (5.3.1). The following statements are equivalent for any $\kappa > 0$*

(i) *The estimate*

$$W_p(\mu_t, \nu_t) \leq \exp\left(-\frac{\kappa}{p}t\right) W_p(\mu, \nu) \tag{5.3.5}$$

holds uniformly on $\mu, \nu \in \mathcal{P}_p(\mathbb{N}^d)$, $t > 0$ and $p \geq 1$, where μ_t (resp. ν_t) is the law at time t of a Markov chain with generator \mathcal{L} started at μ (resp. ν).

(ii) *V^-, V^+ satisfy Assumption (5.3.1) and*

$$\inf_{\substack{\eta \in \mathbb{N}^d \\ i \leq d}} \kappa^+(\eta, i) + \kappa^-(\eta, i) \geq \kappa, \tag{5.3.6}$$

where $\kappa^+(\eta, i)$ and $\kappa^-(\eta, i)$ are defined at (5.3.2) and (5.3.3).

The following Corollary is immediately obtained concatenating Theorem 5.3.2 and Theorem 5.3.3.

Corollary 5.3.1. *Let $V^-, V^+ : \mathbb{N}^d \rightarrow \mathbb{R}$ be given and \mathcal{L} be defined by (5.3.1). If the contraction estimate (5.3.5) holds uniformly on $\mu, \nu \in \mathcal{P}_p(\mathbb{N}^d)$, $t > 0$ and $p \geq 1$, then for all ϕ satisfying (5.2.3) the convex Sobolev inequality holds with constant $\kappa_\phi = \kappa$ and for any $\alpha \in [1, 2]$ the Beckner inequality holds with $\kappa_\alpha = \alpha\kappa$.*

The proof of (ii) \Rightarrow (i) in Theorem 5.3.3 is done by directly constructing coupling rates that give a contractive coupling between two points of \mathbb{N}^d at distance one. Figure 5.2 explains some properties of the support of this coupling. In particular, these coupling rates have the property that the support of $(\gamma, \bar{\gamma}) \mapsto c^{\text{pl}}(\eta, \gamma_i^+ \eta, \gamma, \bar{\gamma})$ can be partitioned into the set of *parallel* terms, *rotational* terms and *good* terms, as I explained in the introduction of this chapter.

5.3.2 Examples

Theorem 5.3.2 is still a rather abstract result. For this reason, let me spend some time to discuss some specific examples where explicit lower bounds on κ_ϕ can be obtained. In [103, Sec 3.2], the authors managed to establish MLSI for a two dimensional non perturbative example corresponding to $V^+(\eta) = h(\eta_1 + \eta_2)$, $V^-(\eta) = \sum_{j=1}^d \log(\eta_j!)$ with h convex and increasing. There, they raised the question of how to generalize this result to a dimension $d > 2$. Thanks to Theorem 5.3.2 we can answer this question in the next Corollary. In order to state this result, and in the remainder of this chapter, for a function $h : \mathbb{R} \rightarrow \mathbb{R}$ we use the notation $\nabla^+ h(m)$ to indicate the increment $h(m+1) - h(m)$ and the potential V^- appearing in (5.3.1) is chosen to be

$$V^-(\eta) = \sum_{i=1}^d \log(\lambda)\eta_i + \log(\eta_i!)$$

so that the measure $\exp(-V^-)$ is an unnormalized multidimensional Poisson distribution. and we will abbreviate V^+ with V .

Corollary 5.3.2. *Let $|\eta| = \sum_{i=1}^d \eta_i$, $h : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function and set $V(\eta) = \beta h(|\eta|)$. Consider the generator \mathcal{L} given by (5.3.1). If*

$$\inf_{m \in \mathbb{N}} \lambda - (d-1)[\exp(-\beta \nabla^+ h(m)) - \exp(-\beta \nabla^+ h(m+1))] \geq 0, \quad (5.3.7)$$

then the conclusion Theorem 5.3.2 holds with

$$\kappa = \inf_{m \in \mathbb{N}} \lambda - (d-2)[\exp(-\beta \nabla^+ h(m)) - \exp(-\beta \nabla^+ h(m+1))]. \quad (5.3.8)$$

In particular, if h is strictly increasing and

$$\beta \geq \frac{\log(d-1) - \log(\lambda)}{h(1) - h(0)}, \quad (5.3.9)$$

then the conclusion of Theorem 5.3.2 holds with

$$\kappa = \lambda - (d-2) \exp(-\beta \nabla_+ h(0)). \quad (5.3.10)$$

Perturbative criteria typically assert that a probability measure of the form $\mathbf{m} = \exp(-\beta V)\boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is a product measure satisfy MLSI provided β is small enough. This is often called a weak interaction/high temperature condition. On the contrary, (5.3.9) asks for a lower bound on β and is therefore a non perturbative condition. I now discuss a more general example that contains the one above as a special case.

Corollary 5.3.3. *Let $V : \mathbb{N}^d \rightarrow \mathbb{R}$ be such that*

$$\nabla_i^+ \nabla_j^+ V(\eta) \geq 0, \quad \forall \eta \in \mathbb{N}^d, i, j \leq d \quad (5.3.11)$$

holds uniformly in $x \in \mathbb{N}^d$. Consider the generator \mathcal{L} given by (5.3.1), and assume that

(i) If

$$\inf_{\eta \in \mathbb{N}^d, i=1, \dots, d} \lambda - \sum_{\substack{j=1 \\ j \neq i}}^d [\exp(-\nabla_j^+ V(\eta)) - \exp(-\nabla_j^+ V(\gamma_i^+ \eta))] \geq 0. \quad (5.3.12)$$

Then the conclusion of Theorem 5.3.2 holds with κ given by

$$\inf_{\eta \in \mathbb{N}^d, i=1, \dots, d} \lambda + [\exp(-\nabla_i^+ V(\eta)) - \exp(-\nabla_i^+ V(\gamma_i^+ \eta))] - \sum_{\substack{j=1 \\ j \neq i}}^d [\exp(-\nabla_j^+ V(\eta)) - \exp(-\nabla_j^+ V(\gamma_i^+ \eta))].$$

(ii) If

$$\min_{i=1, \dots, d} \lambda - \sum_{\substack{j=1 \\ j \neq i}}^d \exp(-\nabla_j^+ V(\mathbf{0})) \geq 0, \quad (5.3.13)$$

where $\mathbf{0} = (0, \dots, 0) \in \mathbb{N}^d$, then the conclusion of Theorem 5.3.2 holds with κ given by (5.3.13).

I did not introduce an "inverse temperature" parameter β in Corollary 5.3.3. If I had done so, we could have seen that, as for Corollary 5.3.2, the local condition (5.3.13) is always satisfied in the low temperature regime $\beta \rightarrow +\infty$.

5.4 Hardcore models

I conclude the presentation of the results of [82] discussing hardcore models and refer the interested reader to the article for more examples of application. To set up the model, I begin by fixing a finite undirected graph (V, E) that is also simple ($(x, x) \notin E$) and connected. As usual, if $(x, y) \in E$ I write $x \sim y$ and say that x, y are neighbors. The state space of the classical hardcore model is

$$\Omega = \{\eta : V \rightarrow \{0, 1\} \text{ s.t. } \eta_x \eta_y = 0, \forall x \sim y\}.$$

For $x \in V$ we define its neighborhood as $N_x = \{y \neq x : y \sim x\}$. A configuration $\eta \in \Omega$ is such that if a site x is occupied, then all sites in its neighborhood are empty. For any $x \in V$ I define $\delta_x \in \Omega$ as the configuration

$$(\delta_x)_y = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{otherwise.} \end{cases}$$

The set of moves is $G = \{\gamma_x^+, \gamma_x^- : x \in V\}$, where

$$\gamma_x^+(\eta) = \begin{cases} \eta + \delta_x, & \text{if } \eta + \delta_x \in \Omega, \\ \eta & \text{otherwise.} \end{cases}, \quad \gamma_x^-(\eta) = \begin{cases} \eta - \delta_x, & \text{if } \eta - \delta_x \in \Omega, \\ \eta & \text{otherwise.} \end{cases}.$$

The generator is given by

$$\mathcal{L}f(\eta) = \sum_{x \in V} c(\eta, \gamma_x^-) \nabla_x^- f(\eta) + c(\eta, \gamma_x^+) \nabla_x^+ f(\eta) \quad (5.4.1)$$

where

$$c(\eta, \gamma_x^-) = \eta_x, \quad c(\eta, \gamma_x^+) = \rho \prod_{y \in N_x} (1 - \eta_y).$$

for some constant $\rho > 0$. The above means that a new particle arrives at rate ρ on an empty site $x \in V$ if the neighborhood of x is empty and that each occupied site $x \in V$ is emptied at rate 1. In particular, for all $x \in V$ we have $(\gamma_x^+)^{-1} = \gamma_x^-$ and the reversible measure for the hardcore model is known to be (see [103] for example)

$$\pi(\eta) = \frac{1}{Z} \mathbf{1}_{\eta \in \Omega} \prod_{x \in V} \rho^{\eta_x},$$

where Z is the normalization constant.

Theorem 5.4.1. *Let Δ be the maximum degree of (V, E) , $\Delta = \sup_{x \in V} |N_x|$. Assume that*

$$\rho \Delta \leq 1 \quad (5.4.2)$$

and set

$$\kappa = 1 - \rho(\Delta - 1), \quad \bar{\kappa} = \min\{\rho, (1 - \rho\Delta)\} \quad (5.4.3)$$

Then the following hold

- (i) If ϕ satisfies (5.2.3), then the convex Sobolev inequality (5.2.2) holds with $\kappa_\phi = \kappa$.
- (ii) The modified log Sobolev inequality (5.1.6) holds with $\kappa_1 = \kappa + 2\bar{\kappa}$.
- (iii) For $\alpha \in (1, 2]$, the Beckner inequality (5.1.7) holds with $\kappa_\alpha = \alpha\kappa$

The hardcore model and its generalizations have been intensively studied, see the discussion in [165, Sec. 22.4]. Mixing times have been studied in [168],[109],[208] among others. The best estimates for the MLSI of the hardcore model I am aware of have been obtained in [103, 114]. For instance, in [114, Cor 4.8] MLSI is shown to hold with constant $1 - \rho(\Delta - 1)$ under assumption (5.4.2). Therefore Theorem 5.4.1 improves on this result. I am not aware of previously known results about Beckner inequalities or general convex Sobolev inequalities for the hardcore model.

5.5 Perspectives

There are many open questions awaiting for an answer in the field. Let me just mention a couple of those I am most interested in.

5.5.1 Overcoming the limitations of convex entropy decay

The methodology proposed in [82] and more generally all works attempting at establishing convexity of the entropy for Markov chains are not particularly efficient when applied to *conservative* systems, such as inhomogeneous zero-range processes and inhomogeneous exclusion processes on the complete graph. For these models it is known that unless a perturbative condition on the rates is satisfied, the entropy is *not* convex in general, even though a contractive coupling, called monotone coupling, can be constructed. This tells that a discrete version of Theorem 5.1.1 cannot be established in full generality. However, one can still ask the question whether or not the best MLSI constant coincides with what couplings "predicts". Despite recent improvements on the MLSI constant such as in [148], an answer is still missing.

Tilting the Dirichlet form A possible remedy to overcome the lack of convexity of the entropy in conservative systems could be to introduce a weighted Dirichlet form

$$\mathcal{E}^w(f, \phi'(f)) = \sum_{\substack{\eta \in \Omega \\ \sigma \in G}} \nabla_\sigma(\phi' \circ f) \nabla_\sigma f(\eta) w(\eta, \sigma) c(\eta, \sigma) \mathbf{m}(\eta).$$

where the weights $w(\eta, \sigma)$ are uniformly bounded above and bounded away from zero. The interest of deforming the Dirichlet form this way is that the differential inequality

$$\frac{d}{dt} \mathcal{E}^w(P_t f, \phi'(P_t f)) \leq -\kappa \mathcal{E}^w(P_t f, \phi'(P_t f))$$

implies a lower bound for the best MLSI constant.

Beyond reversibility The last years have seen a surge of interest around functional inequalities for non reversible Markov processes and question of how to adapt the techniques I developed to handle non reversible chains is compelling. Obviously, the notion of coupling rates makes perfect sense even if the Markov chain is not reversible and there are several non reversible Markov chains for which one can construct contractive couplings. Moreover, some of the sufficient conditions I gave in [82] remain valid even if the underlying Markov chain is not reversible, see (5.3.2), (5.3.3) for example. However, I often made use of reversibility in the proofs in order to obtain cancellations between certain cross-terms and it is not yet clear to me how to do so without this assumption.

5.5.2 Relations with entropic curvature and $\text{CD}(\kappa, +\infty)$

I suspect that the sufficient conditions I found in [82] are indeed sufficient for having positive *entropic* curvature in the sense of [115, 170]. It would quite nice to advance in this direction for two reasons. In first place because positive lower bounds on the entropic curvature have other consequences than MLSI. For example, they imply transport entropy inequalities. A second reason is that doing so would establish a clear connection between coupling rates and entropic curvature. As far as I know, this link is still missing. It is also very natural to wonder what Assumption (5.3.1) "converges" to when we make the mesh of lattice finer and finer, i.e. we work on $\frac{1}{N}\mathbb{N}^d$ and let $N \rightarrow +\infty$, and in particular if it relates to (5.1.1) and the contraction properties of

$$dX_t = -\nabla U(X_t)dt + dB_t. \tag{5.5.1}$$

My personal feeling is that the "limit" of (5.3.1) is related to the following statement

- For any $x, y \in \mathbb{R}^d$, we can construct a coupling $(X_t^x, X_t^y)_{t \geq 0}$ of the diffusion process (5.5.1) started at x and y respectively in such a way that

$$\text{a.s. } d_{\ell^1}(X_t^x, X_t^y) \leq \exp(-\kappa t) d_{\ell^1}(x, y) \quad \forall t > 0,$$

where d_{ℓ^1} is the ℓ^1 distance on \mathbb{R}^d , $d_{\ell^1}(x, y) = \sum_i |x_i - y_i|$, that is obviously related to the graph distance on \mathbb{N}^d used at Theorem 5.3.3. Moreover, it seems that contraction in d_{ℓ^1} implies a stronger condition on U than its convexity, although this affirmation needs to be carefully verified.

Part III

Interpolation in the space of measures

Contents

6.1	Motivation and problem formulation	78
6.2	Problem formulation	78
6.2.1	Splines in \mathbb{R}^d	78
6.2.2	Splines in $\mathcal{P}_2(\mathbb{R}^d)$	79
6.3	The structure of measure-valued splines	79
6.3.1	Decomposition of optimal solutions	80
6.3.2	Formulation of the problem in phase space	80
6.3.3	Monge solutions in phase space	82
6.4	Fluid dynamical formulation of (6.2.3) and connections with Otto calculus	83
6.4.1	A fluid dynamic formulation for (6.2.3)	83
6.4.2	A Riemannian approach	83
6.5	Perspectives and numerical experiments	85

This chapter is devoted to the presentation of the results of the article [68], written in collaboration with Y.Chen and T.Georgiou: deviating from the main themes of this manuscript, the problem under consideration here is that of constructing a smooth interpolation for a collection of measure-valued data indexed over time. The angle of attack that we chose was that of lifting the concept of cubic spline curve from the setting of points in a Euclidean space to that of probability measures, using the framework of multimarginal optimal transport. The results we obtained are to be considered as a first step towards developing a toolbox for interpolation in the space of probability measures; the basic elements of the theory were proven rigorously whereas we only gave formal arguments for some other claims. The general problem of interpolating in Wasserstein space is a fascinating one. In fact, in parallel and independent work, Benamou, Gallouët and Vialard [26] proposed the same model; more recently, a new model has been proposed in [70]. The structure of this chapter is as follows

- In Section 6.2, I define the notion of splines in Wasserstein space by emulating its well known Euclidean counterpart.
- Section 6.3 explores the structure of measure-valued splines and, in particular. It also presents alternative formulations in phase space as well as addresses the question of Monge solutions. Besides certain expected parallels to classical splines, measure-valued splines enjoy a number of interesting structural properties which mirror other well known properties of optimal transport. In particular, we showed that the construction of measure-valued splines relates to a multimarginal optimal transportation problem, and I discuss the existence of Monge-like solutions for an extended (relaxed) formulation of the multimarginal optimal transport problem.
- Section 6.4 is devoted to a fluid dynamic formulation of measure valued spline curves, and to understand its relation with an alternative approach to the spline problem inspired by Otto calculus.
- The last section 6.5 has two parts. In the first I hint at possible possible future research directions. In the second part I report some numerical experiments on Gaussian measures taken from our article that clearly demonstrate the interest of replacing piecewise geodesic interpolation with splines.

6.1 Motivation and problem formulation

Consider a collection of (empirical) probability distributions

$$(\rho_i)_{i=0,1,\dots,N},$$

that are specified at a number of successive points in time $0 = t_0 < t_1 < \dots < t_N = 1$. From an engineering standpoint, such distributions may represent density of particles, concentration of pollutants, image intensity, power distribution, etc., associated with some underlying time-varying physical process. In pertinent application areas, invariably, the goal is to interpolate the available data-set so as, e.g., to estimate the spread of a particle beam or the potential spread of pollutants in-between reference points, to resolve features between successive slices in magnetic resonance imaging, and so on. Thus, our aim was to construct in a systematic manner a measure-valued curve which interpolates *smoothly* a data-set that consists of successive probability distributions. In a classical setting, where the data-set consists of points $(x_i)_{i=0,1,\dots,N}$ in \mathbb{R}^d , a natural choice is to interpolate with a smooth curve such as a cubic spline. This motivates us to seek a suitable generalization of spline curves from the Euclidean setting to *measure-valued spline curves* on the Wasserstein space of probability measures. To do this, we drew inspiration from a variational formulation of splines due to Holladay [149], asserting that the spline-curve in Euclidean space minimizes mean-squared acceleration among all other interpolants, and lifted this notion from points to measures.

A bit of notation I introduce here some notation that is specific to this chapter. The set of functions which are k times differentiable and whose k -th derivative is square-integrable is denoted by $H^k([0, 1]; \mathbb{R}^d)$, and abbreviated by H^k . Classical splines are, by definition, twice continuously differentiable and piecewise cubic polynomials. Thus, for a fixed sequence $\mathcal{T} := (t_i)_{i=0,\dots,N}$ with $0 = t_0 < t_1 < \dots < t_N = 1$ we denote by $\Pi_3([t_i, t_{i+1}])$ the set of \mathbb{R}^d -valued cubic polynomials defined on the interval $[t_i, t_{i+1}]$ and the corresponding set of splines

$$\mathcal{S}_3 := \left\{ X \in \mathcal{C}_2([0, 1]; \mathbb{R}^d) : X|_{[t_i, t_{i+1}]} \in \Pi_3([t_i, t_{i+1}]) \quad \forall i = 0, \dots, N-1 \right\}.$$

In the sequel I work on $\mathcal{P}(\Omega)$, where $\Omega = \mathcal{C}([0, 1]; \mathbb{R}^d)$. Ω is equipped with the canonical sigma algebra generated by the projection maps $(X_t)_{t \in [0, 1]}$, defined by

$$\forall \omega \in \Omega, \quad X_t(\omega) = \omega_t. \quad (6.1.1)$$

If $\mathcal{T} = (t_i)_{i=0,\dots,N}$ is a finite set of times, I denote $X_{\mathcal{T}}$ the vector $(X_{t_0}, X_{t_1}, \dots, X_{t_N})$.

6.2 Problem formulation

Let me now draw the analogy between curve fitting in finite-dimensions and interpolation in the Wasserstein space that lead to our definition of \mathbb{R}^d valued splines.

6.2.1 Splines in \mathbb{R}^d

Let $\mathcal{T} = (t_i)_{i=0,\dots,N}$ with $0 = t_0 < t_1 < \dots < t_N = 1$ be an array of time-data, and $(x_i)_{i=0,\dots,N}$ be a sequence of spatial data in \mathbb{R}^d . The natural interpolating spline for the data is the only $S \in \mathcal{S}_3$ such that $S_{t_i} = x_i$ for $0 \leq i \leq N$ and whose second derivative vanishes at $t = 0, 1$. Holladay's Theorem [149] tells that the variational problem

$$\inf_X \int_0^1 |\ddot{X}_t|^2 dt \quad (6.2.1a)$$

$$X \in H^2, \quad (6.2.1b)$$

$$X_{t_i} = x_i, \quad i = 0, \dots, N. \quad (6.2.1c)$$

admits as unique solution the natural interpolating spline for the data $(t_i, x_i)_{i=0,\dots,N}$, which I denote $S(x_0, \dots, x_N)$ without emphasizing the dependence on the time data \mathcal{T} , as they are kept fixed in all what follows. Also, I denote $\mathcal{S}_3^0 \subset \mathcal{S}_3$ the set of all natural splines

$$\mathcal{S}_3^0 = \{S(x_0, \dots, x_N) : (x_0, \dots, x_N) \subseteq \mathbb{R}^{d \times (N+1)}\}.$$

6.2.2 Splines in $\mathcal{P}_2(\mathbb{R}^d)$

Let us now turn the attention to measures: for given $(t_i, \rho_i)_{i=0, \dots, N}$, with $0 = t_0 < t_1 \dots < t_N = 1$ and $\{\rho_0, \dots, \rho_N\} \subseteq \mathcal{P}_2(\mathbb{R}^d)$, inspired by Holladay's theorem and with an optimal transport viewpoint, we viewed the problem of interpolating smoothly the data as

" the problem of transporting the mass configuration ρ_0 into the mass configuration ρ_i at time t_i while minimizing mean-squared acceleration."

To better justify the model it is useful to record the following observations, motivated by the above informal description of the problem.

- A transport plan is a probability measure $P \in \mathcal{P}(\Omega)$, where $\Omega = \mathcal{C}([0, T]; \mathbb{R}^d)$ and for $A \subseteq \Omega$, $P[A]$ represents the total mass which flows along the paths in A .
- For a plan to be admissible, it must be that at time t_i , the mass configuration induced by P is ρ_i .

$$(X_{t_i})_{\#}P = \rho_i, \quad i = 0, \dots, N.$$

- Since we consider acceleration, we ask that an admissible plan P is such that $P[H^2] = 1$.
- Since we penalize acceleration, we need to consider the mean-square acceleration¹

$$\int_0^1 \int_{\Omega} |\partial_{tt} X_t|^2 dP dt \tag{6.2.2}$$

of an (admissible) plan P .

We are now in the position to define measure-valued spline curves.

Definition 6.2.1. *Let $(t_i, \rho_i)_{i=0, \dots, N} \subset [0, 1] \times \mathcal{P}_2(\mathbb{R}^d)$ be given data. Consider the problem*

$$\inf_P \int_0^1 \int_{\Omega} |\partial_{tt} X_t|^2 dP dt \tag{6.2.3a}$$

$$P \in \mathcal{P}(\Omega), P[H^2] = 1 \tag{6.2.3b}$$

$$(X_{t_i})_{\#}P = \rho_i, \quad i = 0, \dots, N. \tag{6.2.3c}$$

An interpolating spline for the data $(t_i, \rho_i)_{i=0, \dots, N}$ is defined to be the marginal flow $(\rho_t)_{t \in [0, 1]}$ of an optimal measure for (6.2.3).

Note that if instead of taking the second derivative in (6.2.3a) we take the first derivative, then problem (6.2.3) is an equivalent formulation of Monge-Kantorovich problems within each time interval $[t_i, t_{i+1}]$. Also we note that, in general, we cannot guarantee uniqueness for the optimal measure in (6.2.3).

6.3 The structure of measure-valued splines

In this section I describe some basic properties of measure-valued spline curves by linking the problem (6.2.3) with a multi-marginal optimal transport problem. In particular, Theorem 6.3.2 contains an extended formulation of (6.2.3), which is more appealing from a computational standpoint. Moreover, it allows to better understand the support of optimal plans, see Theorem 6.3.3 .

¹When $(X_t)_{t \in [0, 1]}$ is the canonical process, we denote the acceleration $\partial_{tt} X_t$ instead of \ddot{X}_t .

6.3.1 Decomposition of optimal solutions

The following theorem asserts that at least an optimal solution for (6.2.3) exists and gives details about the structure of the solution. In the article we did not prove uniqueness of the measure-valued spline through a given data set; this interesting question remains open for further investigation. In words, Theorem 6.3.1 says that any optimal solution is supported on splines of \mathbb{R}^d , and that its joint distribution at times t_0, \dots, t_N solves a multimarginal optimal transport problem whose cost function \mathcal{C} is the optimal value in (6.2.1), i.e.

$$\mathcal{C}(x_0, \dots, x_N) := \int_0^1 |\partial_{tt} S_t(x_0, \dots, x_N)|^2 dt. \quad (6.3.1)$$

Thus a spline curve on $\mathcal{P}_2(\mathbb{R}^d)$ is found by pushing forward through splines of \mathbb{R}^d the solution of a multimarginal optimal problem. This is in analogy with the well known fact that the geodesics of $\mathcal{P}_2(\mathbb{R}^d)$ are constructed pushing forward the optimal coupling of the Monge-Kantorovich problem through geodesics of \mathbb{R}^d . In the statement of the theorem, extending the notations used so far in a natural way, I call $\Pi(\rho_0, \rho_1, \dots, \rho_N)$ the set of all couplings of the probability measures ρ_0, \dots, ρ_N .

$$\Pi(\rho_0, \rho_1, \dots, \rho_N) = \{ \pi \in \mathcal{P}(\mathbb{R}^d \times \dots \times \mathbb{R}^d) : (X_i)_\# \pi = \rho_i \},$$

where I denoted by X_i the i -th coordinate map on $(\mathbb{R}^d)^{N+1}$, i.e. $X_i(x_0, \dots, x_N) = x_i$.

Theorem 6.3.1. *Let $\{\rho_0, \dots, \rho_N\} \subseteq \mathcal{P}_2(\mathbb{R}^d)$. Then there exists at least an optimal solution for (6.2.3). Moreover, the following are equivalent*

(i) \hat{P} is an optimal solution for (6.2.3).

(ii) $\hat{P}[\mathcal{S}_3^0] = 1$ and $\hat{\pi} := (X_T)_\# \hat{P}$ is an optimal solution for

$$\begin{aligned} \inf_{\pi} \int \mathcal{C}(x_0, x_1, \dots, x_N) d\pi \\ \pi \in \Pi(\rho_0, \rho_1, \dots, \rho_N), \end{aligned} \quad (6.3.2)$$

where \mathcal{C} has been defined at (6.3.1).

A suitable entropic regularization of multimarginal optimal transport problems, such as the one in (6.3.2), can be solved numerically using iterative Bregman projections [25]. In the special case where the marginals are Gaussian distributions, a numerically efficient semidefinite programming (SDP) formulation is possible.

6.3.2 Formulation of the problem in phase space

One aspect of the cost \mathcal{C} which complicates the tractability of (6.3.2) is that, to the best of our knowledge, there is no closed form expression valid for any N . For this reason, we propose a second, equivalent formulation of (6.3.2) in a larger space with an explicit cost function. We achieved this by looking into "phase space": that is, we introduce auxiliary variables playing the role of velocities. This is a classical procedure in physics used to rewrite second order equations as systems of first order equations. Here we consider probability measures on the product space $\Omega \times \Omega$, and we denote $(X_t, V_t)_{t \in [0,1]}$ the canonical process on it. As in (6.1.1), the canonical projection maps (X_t, V_t) are defined for all $(\omega^1, \omega^2) \in \Omega \times \Omega$ by

$$(X_t, V_t)(\omega^1, \omega^2) = (\omega_t^1, \omega_t^2).$$

Then,

$$\inf_Q \int_0^1 \int_{\Omega \times \Omega} |\partial_t V_t|^2 dQ dt \quad (6.3.3a)$$

$$Q \in \mathcal{P}(\Omega \times \Omega), Q[H_1 \times H_1] = 1, \quad (6.3.3b)$$

$$Q[\partial_t X_t = V_t \ \forall t \in [0, 1]] = 1, \quad (6.3.3c)$$

$$(X_{t_i})_\# Q = \rho_i \quad i = 0, \dots, N, \quad (6.3.3d)$$

is easily seen to be equivalent to (6.2.3). The interesting fact is that, the multimarginal optimal transport problem associated with (6.3.3) has an explicit cost function. All relies on the following representation of \mathcal{C} as the solution of a minimization problem.

Lemma 6.3.1. *Let $(x_i, v_i)_{i=0, \dots, N} \subset \mathbb{R}^d \times \mathbb{R}^d$ be given. The optimal value of the problem*

$$\inf_{(X, V)} \int_0^1 |\dot{V}_t|^2 dt \quad (6.3.4a)$$

$$(X, V) \in H_1 \times H_1, \quad (6.3.4b)$$

$$\dot{X}_t = V_t, \quad \forall t \in [0, 1], \quad (6.3.4c)$$

$$X_{t_i} = x_i, \quad i = 0, \dots, N, \quad (6.3.4d)$$

$$V_{t_i} = v_i, \quad i = 0, \dots, N. \quad (6.3.4e)$$

is given by

$$\sum_{i=0}^{N-1} (t_{i+1} - t_i)^{-1} c(x_i, x_{i+1}, v_i, v_{i+1}) \quad (6.3.5)$$

where

$$c(x_i, x_{i+1}, v_i, v_{i+1}) = 12|x_{i+1} - x_i - v_i|^2 - 12\langle x_{i+1} - x_i - v_i, v_{i+1} - v_i \rangle + 4|v_{i+1} - v_i|^2. \quad (6.3.6)$$

In particular,

$$\mathcal{C}(x_0, \dots, x_N) = \inf_{v_0, \dots, v_N \in \mathbb{R}^d} \sum_{i=0}^{N-1} (t_{i+1} - t_i)^{-1} c(x_i, x_{i+1}, v_i, v_{i+1}) \quad (6.3.7)$$

and the infimum in (6.3.7) is attained and is unique.

In analogy with the above formula, it is worth observing that multimarginal optimal transport problems for a cost of the form

$$\mathcal{C}(x_0, \dots, x_N) = \inf_{y \in Y} \sum_{i=0}^N c_i(x_i, y).$$

have been studied in [54]. The main difference with the above is that c in (6.3.7) depends on both x_i and x_{i+1} , which somewhat complicates the analysis.

Theorem 6.3.2. *Let $\{\rho_0, \dots, \rho_N\} \subseteq \mathcal{P}_2(\mathbb{R}^d)$. Then there exists at least an optimal solution for (6.3.3). Moreover, for an admissible plan \hat{Q} the following are equivalent*

(i) \hat{Q} is an optimal solution for (6.3.3).

(ii) $\hat{Q}[X \in \mathcal{S}_3^0] = 1$ and $\hat{\gamma} := (X_{\mathcal{T}}, V_{\mathcal{T}})_{\#} \hat{Q}$ is an optimal solution for

$$\inf_{\gamma} \sum_{i=0}^{N-1} (t_{i+1} - t_i)^{-1} \int_{\mathbb{R}^{2d \times 2(N+1)}} c(x_i, x_{i+1}, v_i, v_{i+1}) d\gamma, \quad (6.3.8)$$

$$\gamma \in \Gamma(\rho_0, \dots, \rho_N),$$

where $\Gamma(\rho_0, \dots, \rho_N)$ is defined by

$$\Gamma(\rho_0, \dots, \rho_N) := \left\{ \gamma \in \mathcal{P}(\mathbb{R}^{d \times (N+1)} \times \mathbb{R}^{d \times (N+1)}) : (X_i)_{\#} \gamma = \rho_i \quad \forall i = 0, \dots, N \right\}.$$

The set $\Gamma(\rho_0, \dots, \rho_N)$ should be thought of as the lift to the phase space of the set of couplings $\Pi(\rho_0, \dots, \rho_N)$. Indeed the projection onto the space variables of any plan in $\Gamma(\rho_0, \dots, \rho_N)$ is a coupling in $\Pi(\rho_0, \dots, \rho_N)$. We made use of this to prove the next proposition about the equivalence between the two multimarginal problems. There, we denote \mathbb{V} the maps that associates to (x_0, \dots, x_N) the optimal solution of (6.3.7). It is not hard to see that \mathbb{V} is a linear map.

Proposition 6.3.1. *The problem (6.3.8) is equivalent to the problem (6.3.2) in the following sense:*

(i) *If $\hat{\gamma}$ is optimal for (6.3.8) then*

$$\hat{\pi} := (X_0, \dots, X_N) \# \hat{\gamma}$$

is optimal for (6.3.2).

(ii) *If $\hat{\pi}$ is optimal for (6.3.2), then*

$$\hat{\gamma} := (X_0, \dots, X_N, \mathbb{V}(X_0, \dots, X_N)) \# \hat{\pi}$$

is optimal for (6.3.8).

6.3.3 Monge solutions in phase space

There is something that can be said about Monge, or graphical, solutions to the extended formulation. Roughly speaking, an optimal plan is said to be “Monge” if it can be written as the push forward of the first marginal through a deterministic map. The existence of Monge solutions for the two-marginals problems is rather well-understood. The multimarginal case is considerably harder to study (see [188] for more details). Unfortunately, we were not able to provide a complete existence result. However, we show that if an optimal solution has some regularity properties, then it is of Monge type (in phase space).

Theorem 6.3.3. *Let $\hat{\gamma}$ be an optimal solution for (6.3.8) such that for all $i = 0, \dots, N-1$ the measure $\hat{\gamma}_i \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ defined by*

$$\hat{\gamma}_i = (X_i, V_i) \# \hat{\gamma}$$

is absolutely continuous w.r.t to the Lebesgue measure. Then there exist a map

$$\Phi = (\varphi_1, \dots, \varphi_N, \psi_1, \dots, \psi_N) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times N} \times \mathbb{R}^{d \times N}$$

such that $\hat{\gamma}$ is concentrated on the graph of Φ , i.e.

$$\hat{\gamma} = (\mathbf{id}, \varphi_1, \dots, \varphi_N, \mathbf{id}, \psi_1, \dots, \psi_N) \# \hat{\gamma}_0,$$

or equivalently

$$\hat{\gamma} \left[\bigcap_{i=1}^N \{X_i = \varphi_i(X_0, V_0), V_i = \psi_i(X_0, V_0)\} \right] = 1. \quad (6.3.9)$$

It would be very desirable to derive the conclusion assuming just regularity of the (ρ_i) instead of the $\hat{\gamma}_i$. Theorem 6.3.3 suggests that Monge solutions for the original formulation (6.3.2) are not to be expected; the support of an optimal solution should be locally of dimension $2d$. I also believe that the assumptions of the Theorem can be largely relaxed. The next proposition is to be seen as a first step in this direction for the case when $N = 2$ (i.e. we interpolate three measures), using the general results of [187].

Proposition 6.3.2. *Let $N = 2$, $\hat{\pi}$ an optimal solution for (6.3.2), and (x_0, x_1, x_2) a point in the support of $\hat{\pi}$. Then there is a neighborhood O of (x_0, x_1, x_2) such that the intersection of the support of $\hat{\pi}$ with O is contained in a Lipschitz submanifold of dimension $2d$.*

Let us note that this proposition does not yield the existence of Monge solutions for (6.3.8); however it proves that optimal solutions of (6.3.2) have a support which is locally of dimension $2d$, without making any further regularity assumption on the optimal coupling.

6.4 Fluid dynamical formulation of (6.2.3) and connections with Otto calculus

Together with YongXin and Tryphon, we investigated the fluid-dynamic formulation for (6.2.3). The main difference with respect to the one for the Monge-Kantorovich problem is that the "action" to be minimized contains an acceleration in term instead of a kinetic energy term.

To better understand what follows, let me recall the fluid dynamic formulation of the Monge-Kantorovich problem, which is due to Benamou and Brenier. In [24] they showed that the optimal value for

$$\inf_{\mu, v} \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \mu_t(x) dx dt \quad (6.4.1a)$$

$$\partial_t \mu_t(x) + \nabla \cdot (v_t \mu_t)(x) = 0 \quad (6.4.1b)$$

$$\mu_0 = \rho_0, \mu_1 = \rho_1$$

is the squared Wasserstein distance $W_2^2(\rho_0, \rho_1)$ and that the optimal curve is the displacement interpolation [171]. Here, $\mu_t(x)$ may be thought of as the density of a fluid evolving in time and $v_t(x)$ as a velocity variable, representing variation of density. Then $\mu_t(x)$ and $v_t(x)$ are tied together via the continuity equation (6.4.1b).

6.4.1 A fluid dynamic formulation for (6.2.3)

Inspired by (6.3.8), one can consider the following problem

$$\inf_{\mu, a} \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}^d} |a_t(x, v)|^2 \mu_t(x, v) dx dv dt \quad (6.4.2a)$$

$$\partial_t \mu_t(x, v) + \langle \nabla_x \mu_t(x, v), v \rangle + \nabla_v \cdot (a_t \mu_t)(x, v) = 0, \quad (6.4.2b)$$

$$\int_{\mathbb{R}^d} \mu_{t_i}(x, v) dv = \rho_{t_i}(x), \quad i = 0, \dots, N, \quad (6.4.2c)$$

where I denote by ∇_x (resp. ∇_v) the gradient taken w.r.t. the x (resp. v) variables, so that $\nabla_x \cdot$ stands for the divergence taken w.r.t. the x variables, and similarly for $\nabla_v \cdot$. In contrast to (6.4.1a), one should observe that the variable $a_t(x, v)$ represents an acceleration. In [68] we made the following

Claim 6.4.1. *The two problems (6.2.3) and (6.4.2) are equivalent.*

Moreover, we provided a formal calculation to justify it. However, the argument we used does not constitute a rigorous proof as it rests on assuming existence of Monge-like solutions for (6.3.8), which we only proved under certain assumptions. Also, we took derivatives formally without showing that this can actually be done and, if so in what sense.

6.4.2 A Riemannian approach

The goal of this section is to propose an alternative viewpoint on the spline problem for measures which is more geometric in spirit and to see that the resulting optimization problem admits (6.2.3) as a relaxation. There exist different approaches to the problem of interpolating smoothly data on a Riemannian manifold; in the upcoming discussion I am following the *intrinsic approach*, see [178], [44]. Given data $(t_i, x_i)_{i=0, \dots, N} \subseteq [0, 1] \times M$, where M is a Riemannian manifold, Holladay's theorem suggests to define the interpolating spline as the optimizer for

$$\inf_X \int_0^1 \left\langle \frac{\mathbf{D}}{dt} \dot{X}_t, \frac{\mathbf{D}}{dt} \dot{X}_t \right\rangle_{T_{X_t} M} dt \quad (6.4.3)$$

$$X \in H^2([0, 1]; M), \quad X_{t_i} = x_i \quad i = 0, \dots, N,$$

where, for a curve X_t on M , we denote \dot{X}_t its velocity, and by $\frac{\mathbf{D}}{\mathbf{d}t}\dot{X}_t$ the acceleration, i.e. the covariant derivative of the velocity along the curve. When turning to the Riemannian manifold of optimal transport, we shall use the notion of covariant derivative given by Otto calculus. While I refer to section 2.2.2 for more detailed constructions, I recall here that given a regular curve $(\rho_t)_{t \in [0,1]}$, whose velocity field is (v_t) , then the covariant derivative of (v_t) along (ρ_t) is given by the formula (see e.g. [4, Example 6.7])

$$\frac{\mathbf{D}}{\mathbf{d}t}v_t = \partial_t v_t + \frac{1}{2}\nabla|v_t|^2 \in \mathcal{T}_{\rho_t}\mathcal{P}_2.$$

and the norm of the covariant derivative on the tangent space $\mathcal{T}_{\rho_t}\mathcal{P}_2$ is given by

$$\left\langle \frac{\mathbf{D}}{\mathbf{d}t}v_t, \frac{\mathbf{D}}{\mathbf{d}t}v_t \right\rangle_{\mathcal{T}_{\rho_t}\mathcal{P}_2} = \int_{\mathbb{R}^d} \left| \partial_t v_t + \frac{1}{2}\nabla|v_t|^2 \right|^2 \rho_t(x) dx. \quad (6.4.4)$$

At this point, it is natural to define measure valued spline curves via the following problem

$$\inf_{\rho, v} \int_0^1 \int_{\mathbb{R}^d} \left| \partial_t v_t + \frac{1}{2}\nabla|v_t|^2 \right|^2(x) \rho_t(x) dx dt \quad (6.4.5a)$$

$$\partial_t \rho_t(x) + \nabla \cdot (v_t \rho_t)(x) = 0, \quad v_t \in \mathcal{T}_{\rho_t}\mathcal{P}_2 \quad (6.4.5b)$$

$$(\rho_t) \in H_2([0, 1]; \mathcal{P}_2(\mathbb{R}^d)), \quad \rho_{t_i} = \rho_i, \quad i = 0, \dots, N, \quad (6.4.5c)$$

where the space $H_2([0, 1]; \mathcal{P}_2(\mathbb{R}^d))$ should be properly defined using the notions of absolutely continuous and regular curve. Clearly, the problem (6.4.5) looks rather different from (6.4.2), and therefore, it should not be equivalent to (6.2.3). However, although different, the two problems are strongly related: in [26] the authors prove that the optimal values of the two problems coincide. Moreover, I would like to provide a heuristic calculation showing that (6.2.3) can be viewed as a relaxation of (6.4.5). More precisely, one may view in some sense (6.4.5) as the Monge problem for (6.2.3). In order to avoid confusion with the concept of Monge solutions encountered at section 6.3.3, let me clarify that a Monge solution for (6.2.3) is a plan $P \in \mathcal{P}(\Omega)$ for which there exist a family of maps $\mathbb{X}_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

$$P := ((\mathbb{X}_t)_{t \in [0,1]})_{\#} \rho_0. \quad (6.4.6)$$

Consider now a solution (ρ, v) for (6.4.5), and define the maps \mathbb{X}_t via

$$\partial_t \mathbb{X}_t(x) = v_t(\mathbb{X}_t(x)), \quad \mathbb{X}_0(x) = x. \quad (6.4.7)$$

and P through (6.4.6). These are the flow maps for the velocity field (v_t) on \mathbb{R}^d and satisfy

$$(\mathbb{X}_t)_{\#} \rho_0 = \rho_t \quad (6.4.8)$$

Therefore, P is admissible for (6.2.3) and we have

$$\begin{aligned} \int_0^1 \int_{H_2} |\partial_{tt} X_t|^2 dP dt &\stackrel{(6.4.6)}{=} \int_0^1 \int_{\mathbb{R}^d} |\partial_{tt} \mathbb{X}_t(x)|^2 \rho_0(x) dx dt \\ &\stackrel{(6.4.7)}{=} \int_0^1 \int_{\mathbb{R}^d} \left| \partial_t v_t(\mathbb{X}_t(x)) \right|^2 \rho_0(x) dx dt \\ &\stackrel{(6.4.7)}{=} \int_0^1 \int_{\mathbb{R}^d} \left| \partial_t v_t + Dv_t \cdot v_t \right|^2(\mathbb{X}_t(x)) \rho_0(x) dx dt \\ &= \int_0^1 \int_{\mathbb{R}^d} \left| \partial_t v_t + \frac{1}{2}\nabla|v_t|^2 \right|^2(\mathbb{X}_t(x)) \rho_0(x) dx dt \\ &\stackrel{(6.4.8)}{=} \int_0^1 \int_{\mathbb{R}^d} \left| \partial_t v_t + \frac{1}{2}\nabla|v_t|^2 \right|^2(x) \rho_t(x) dx dt, \end{aligned}$$

Note that in the above calculations we exploited the fact that $v_t \in \mathcal{T}_{\rho_t}\mathcal{P}_2$ and therefore is (almost) a gradient vector field implying that $Dv_t \cdot v_t = \frac{1}{2}\nabla|v_t|^2$. Thus, we have seen that, to a solution of (6.4.5) we can associate a Monge solution for (6.2.3) and the cost of the two solutions in their respective problems is identical.

To conclude that the two problems are equivalent, we should reverse this last statement. But to do this, we should know that we can w.l.o.g consider Monge solutions (6.4.6) where the maps \mathbb{X}_t are the flow maps for a *gradient* vector field. It is known that non-Monge optimal solutions exist in specific examples. Although this does not exclude a priori that some optimal Monge solution exist, it is very likely that this does not happen in general. Therefore (6.4.5) and (6.2.3), even though they have the same optimal value, are not to be thought as equivalent problems.

6.5 Perspectives and numerical experiments

There are, in my opinion, many open questions around the topic of smooth interpolation in the space of probability measures that are quite interesting. Let me present a few.

Theoretical questions Firstly, one would like to understand whether the existence of Monge solutions for the extended formulation in phase space of (6.3.2), that we proved conditionally on having a regular optimal solution, can be proven without this restrictive assumption. At the moment it is not clear to me if regularity of the data can be translated into some kind of regularity for the solutions: we verified that this is true when the data are Gaussian measures, but it may well fail to be true in general. Answering this question is of particular interest from an applied standpoint as Monge solutions are much easier to interpret and visualize. Another natural research direction to explore concerns the geometry of optimizers. In a nutshell, the main question here is to understand what is the right counterpart to the statement that optimal solutions of (2.1.1) are constructed pushing forward the first marginal through the gradient of a convex function. What replaces the notion of "convex function" in the present context? Note that cyclical monotonicity is known to be a necessary condition for optimality in the multimarginal case as well [187] and in the two marginal case the fact that the optimal solution is concentrated on the graph of a convex function can easily be guessed (not so easily proven) from the monotonicity condition. However, in the multimarginal case, such condition is less readable and we were not able to extract from some it some meaningful geometric property of optimizers. Finally, we do not know whether uniqueness holds for (6.2.3). All the sufficient conditions for uniqueness I am aware of (see e.g. [188] for an overview) are not verified by this variational problem.

Numerical methods The numerical analysis of 6.3.2 is not straightforward. The problem is of course solvable in principle as a very large linear program after discretization of the marginals and the cost. However, as it is often the case for multimarginal problems, this approach may be of limited interest in concrete situations. A general possibility popularized by Cuturi [102] that has had a strong impact in the applied optimal transport community is that of solving an *entropic regularization* of the problem at hand for which efficient algorithms such as Sinkhorn's algorithm are available. As one may expect, the entropic regularization of Monge Kantorovich problem essentially coincides the Schrödinger problem. In the work [26], the authors implement this idea on a version of (6.3.2) and performed a wide range of numerical experiments. From a theoretical perspective it would be very nice to justify the addition of the entropic term made there by means of a large deviations principle of Sanov type. My feeling is that it would very hard to come up with a particle system that does the job. However, I suspect that if another regularization is made one could connect (6.2.3) with the action functional of the kinetic Schrödinger problem (**KSP**). Making clear this connection may not be merely of theoretical interest, but could lead to a new, and performing algorithm for solving (6.3.2). The authors of [70] make progresses towards obtaining a numerically tractable notion of smooth interpolating curve in $\mathcal{P}_2(\mathbb{R}^d)$. In particular for their model optimal solution are of Monge type. However, the model considered there replaces the multimarginal problem (6.3.2) with a sequence of standard Monge–Kantorovich problems between two marginals and the relation between the resulting optimization problem and a measure valued spline is less apparent to me.

Some numerical experiments In [68] we also worked towards obtaining an efficient algorithm for the construction of splines. I do not report on this here. Nevertheless, I still believe it is worth to present some numerical experiments that visually explain the interest of working with splines. In the first picture I display four Gaussian probability measures on \mathbb{R}^2 , that are the ones we want to interpolate at times $\mathcal{T} = \{0, 1/3, 2/3, 1\}$. The second picture shows what

one would get with a "piecewise geodesic" interpolation. That is, we compute the displacement interpolation between any pair of consecutive measures and glue the geodesics together to get an interpolant. Obviously, the resulting curve has some sharp angles. In the third picture I show what the interpolant obtained solving (6.3.2) numerically looks like.

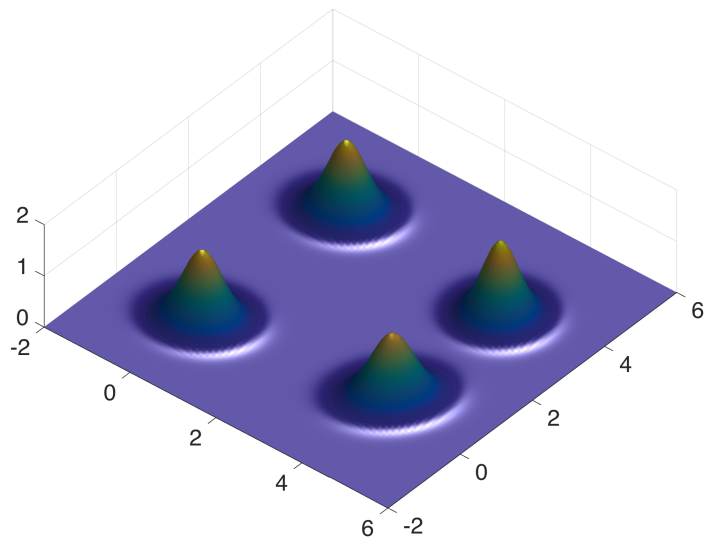


Figure 6.1 – Successive one-time marginal distributions

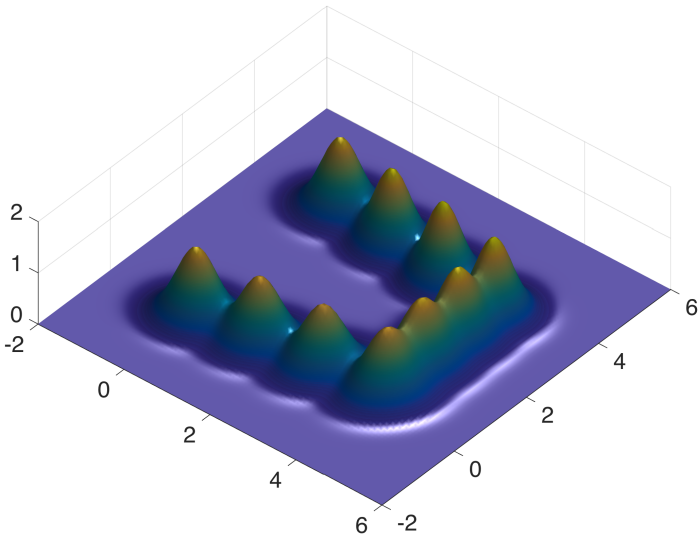


Figure 6.2 – Optimal transport interpolation

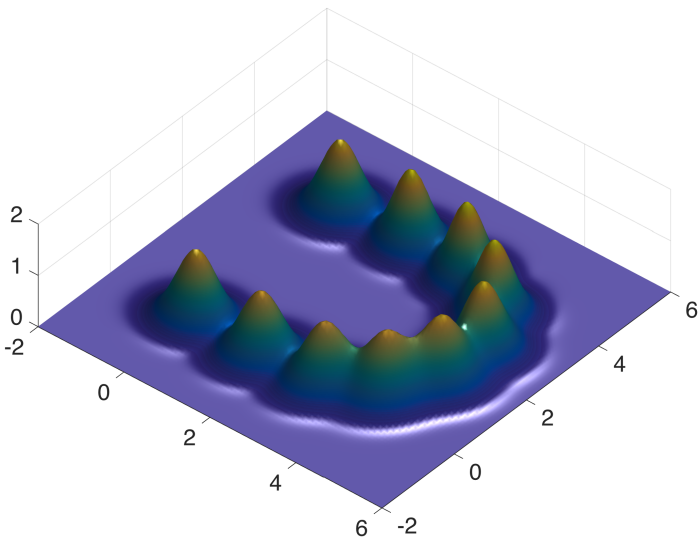


Figure 6.3 – Spline interpolation

Frequently used notation

$\mathcal{C}(M)$ The space of continuous functions over M

$\mathcal{C}_k^c(M)$ The elements of $\mathcal{C}_k(M)$ with compact support

$\mathcal{C}_c^\infty(M)$ The set of smooth compactly supported functions on the Riemannian manifold M .

$\mathcal{C}_+^\infty(M)$ The set of smooth positive functions over the manifold M .

$\mathcal{C}_b^\infty(M)$ The set of smooth and bounded functions on the Riemannian manifold M .

$\mathcal{C}_k(M)$ The space of k -times continuously differentiable functions over M

Ω The space of continuous paths $\mathcal{C}([0, T]; M)$, where M is a Riemannian manifold. In many situations, $M = \mathbb{R}^d$.

\sim Equivalence for probability measures

Bibliography

- [1] Stefan Adams, Nicolas Dirr, Mark A. Peletier, and Johannes Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Communications in Mathematical Physics*, 307(3):791–815, 2011. (Cited on pages [47](#) et [48](#).)
- [2] Luigi Ambrosio and Jin Feng. On a class of first order Hamilton-Jacobi equations in metric spaces. *Journal of Differential Equations*, 256(7):2194 – 2245, 2014. (Cited on pages [33](#) et [35](#).)
- [3] Luigi Ambrosio and Wilfrid Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Communications on Pure and Applied Mathematics*, 61(1):18–53, 2008. (Cited on page [24](#).)
- [4] Luigi Ambrosio and Nicola Gigli. A user’s guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013. (Cited on pages [12](#), [17](#), [20](#), [21](#), [22](#) et [84](#).)
- [5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008. (Cited on pages [18](#), [22](#), [32](#), [33](#), [34](#) et [35](#).)
- [6] Luigi Ambrosio and Aldo Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003. (Cited on page [12](#).)
- [7] Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, 178(1):475–530, 2020. (Cited on pages [25](#), [28](#) et [51](#).)
- [8] Julio Backhoff-Veraguas, Mathias Beiglböck, and Giovanni Conforti. A non-linear monotonicity principle and applications to Schrödinger type problems. *arXiv preprint arXiv:2101.09975, in revision at Bulletin of the London Mathematical Society*, 2021. (Cited on page [11](#).)
- [9] Julio Backhoff-Veraguas, Mathias Beiglböck, Martin Huesmann, and Sigrid Källblad. Martingale Benamou–Brenier: a probabilistic perspective. *Annals of Probability*, 48(5):2258–2289, 2020. (Cited on page [15](#).)
- [10] Julio Backhoff-Veraguas, Mathias Beiglböck, and Gudmund Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):203, 2019. (Cited on page [15](#).)
- [11] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX 1983/84*, pages 177–206. Springer, 1985. (Cited on pages [42](#) et [48](#).)
- [12] D. Bakry and M. Ledoux. A logarithmic Sobolev form of the Li-Yau parabolic inequality. *Rev. Mat. Iberoam.*, 22(2):683–702, 2006. (Cited on page [50](#).)
- [13] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013. (Cited on pages [42](#) et [50](#).)
- [14] Elena Bandini, Andrea Cosso, Marco Fuhrman, and Huyên Pham. Randomized filtering and bellman equation in wasserstein space for partial observation control problem. *Stochastic Processes and their Applications*, 129(2):674–711, 2019. (Cited on page [32](#).)
- [15] Fabrice Baudoin. Bakry–Émery meet villani. *Journal of Functional Analysis*, 273(7):2275–2291, 2017. (Cited on pages [52](#) et [53](#).)
- [16] William Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, pages 159–182, 1975. (Cited on page [62](#).)

- [17] William Beckner. A generalized Poincaré inequality for Gaussian measures. *Proceedings of the American Mathematical Society*, pages 397–400, 1989. (Cited on page 62.)
- [18] M. Beiglböck and C. Griessler. A land of monotone plenty. *Annali della SNS*, Vol. XIX, issue 1, April 2019. (Cited on page 15.)
- [19] Mathias Beiglböck, Manu Eder, Christiane Elgert, and Uwe Schmock. Geometry of distribution-constrained optimal stopping problems. *Probability Theory and Related Fields*, 172(1-2):71–101, 2018. (Cited on page 15.)
- [20] Mathias Beiglböck, Martin Goldstern, Gabriel Maresch, and Walter Schachermayer. Optimal and better transport plans. *Journal of Functional Analysis*, 256(6):1907–1927, 2009. (Cited on page 12.)
- [21] Mathias Beiglböck and Nicolas Juillet. On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42–106, 2016. (Cited on page 15.)
- [22] Mathias Beiglböck, Marcel Nutz, and Florian Stebegg. Fine properties of the optimal skorokhod embedding problem. *Journal of the European Mathematical Society*, 2021. (Cited on page 15.)
- [23] Mathias Beiglböck, Marcel Nutz, and Nizar Touzi. Complete duality for martingale optimal transport on the line. *The Annals of Probability*, 45(5):3038–3074, 2017. (Cited on page 15.)
- [24] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. (Cited on pages 5, 19 et 83.)
- [25] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015. (Cited on page 80.)
- [26] Jean-David Benamou, Guillaume Carlier, Simone Di Marino, and Luca Nenna. An entropy minimization approach to second-order variational mean-field games. *Mathematical Models and Methods in Applied Sciences*, 29(08):1553–1583, 2019. (Cited on pages 77, 84 et 85.)
- [27] Alain Bensoussan, Jens Frehse, and Phillip Yam. *Mean field games and mean field type control theory*, volume 101. Springer, 2013. (Cited on page 33.)
- [28] Alain Bensoussan, P Jameson Graber, and Sheung Chi Phillip Yam. Control on hilbert spaces and application to mean field type control theory. *arXiv preprint arXiv:2005.10770*, 2020. (Cited on page 32.)
- [29] Espen Bernton, Promit Ghosal, and Marcel Nutz. Entropic optimal transport: Geometry and large deviations. *arXiv preprint arXiv:2102.04397*, 2021. (Cited on pages 15 et 47.)
- [30] Lorenzo Bertini, Alberto De Sole, Davide Gabrielli, Giovanni Jona-Lasinio, and Claudio Landim. Macroscopic fluctuation theory for stationary non-equilibrium states. *Journal of Statistical Physics*, 107(3-4):635–675, 2002. (Cited on page 32.)
- [31] Lorenzo Bertini, Alberto De Sole, Davide Gabrielli, Giovanni Jona-Lasinio, and Claudio Landim. Macroscopic fluctuation theory. *Reviews of Modern Physics*, 87(2):593, 2015. (Cited on pages 8, 28, 29 et 32.)
- [32] Arne Beurling. An automorphism of product measures. *Annals of Mathematics*, pages 189–200, 1960. (Cited on page 13.)
- [33] S. Bianchini and L. Caravenna. On optimality of c -cyclically monotone transference plans. *C. R. Math. Acad. Sci. Paris*, 348(11-12):613–618, 2010. (Cited on page 12.)
- [34] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. *arXiv preprint arXiv:2103.07459*, 2021. (Cited on page 63.)

- [35] Sergey Bobkov, Ivan Gentil, and Michel Ledoux. Hypercontractivity of Hamilton Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001. (Cited on page 46.)
- [36] Sergey Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. (Cited on page 61.)
- [37] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. *Archive for Rational Mechanics and Analysis*, 208(2):429–445, 2013. (Cited on page 25.)
- [38] François Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent vlasov-fokker-planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):867–884, 2010. (Cited on page 52.)
- [39] Jonathan M Borwein and Adrian Stephen Lewis. Decomposition of multivariate functions. *Canadian journal of mathematics*, 44(3):463–482, 1992. (Cited on page 13.)
- [40] Jonathan M Borwein, Adrian Stephen Lewis, and Roger Nussbaum. Entropy minimization, DAD problems, and doubly stochastic kernels. *Journal of Functional Analysis*, 123(2):264–307, 1994. (Cited on page 13.)
- [41] Anne-Severine Boudou, Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Spectral gap estimates for interacting particle systems via a Bochner-type identity. *Journal of Functional Analysis*, 232(1):222–258, 2006. (Cited on pages 60, 62 et 66.)
- [42] Amarjit Budhiraja, Paul Dupuis, and Markus Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *The Annals of Probability*, 40(1):74–102, 2012. (Cited on page 7.)
- [43] Matteo Burzoni, Vincenzo Ignazio, Max Reppen, and Mete Soner. Viscosity solutions for controlled McKean–Vlasov jump-diffusions. *SIAM Journal on Control and Optimization*, 58(3):1676–1699, 2020. (Cited on page 33.)
- [44] Margarida Camarinha, Silva Leite, and Peter Crouch. On the geometry of Riemannian cubic polynomials. *Differential Geometry and its Applications*, 15(2):107–135, 2001. (Cited on page 83.)
- [45] Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Convex entropy decay via the Bochner–Bakry–Émery approach. *Annales de l’IHP Probabilités et statistiques*, 45(3):734–753, 2009. (Cited on pages 60, 62 et 66.)
- [46] Pierre Cardaliaguet. Long time average of first order mean field games and weak KAM theory. *Dynamic Games and Applications*, 3(4):473–488, 2013. (Cited on page 42.)
- [47] Pierre Cardaliaguet. Notes on mean field games. Technical report, available at <https://www.ceremade.dauphine.fr/~cardaliaguet/>, 2013. (Cited on page 42.)
- [48] Pierre Cardaliaguet, Guillaume Carlier, and Bruno Nazaret. Geodesics for a class of distances in the space of probability measures. *Calculus of Variations and Partial Differential Equations*, 48(3-4):395–420, 2013. (Cited on page 30.)
- [49] Pierre Cardaliaguet, Marco Cirant, and Alessio Porretta. Splitting methods and short time existence for the master equations in mean field games. *arXiv preprint arXiv:2001.10406*, 2020. (Cited on page 33.)
- [50] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The Master Equation and the Convergence Problem in Mean Field Games:(AMS-201)*, volume 201. Princeton University Press, 2019. (Cited on page 33.)
- [51] Pierre Cardaliaguet, Jean-Michel Lasry, Pierre-Louis Lions, and Alessio Porretta. Long time average of mean field games. *Networks and Heterogeneous Media*, 7(2):279–301, 2012. (Cited on pages 42 et 51.)
- [52] Pierre Cardaliaguet, Jean-Michel Lasry, Pierre-Louis Lions, and Alessio Porretta. Long time average of mean field games with a nonlocal coupling. *SIAM Journal on Control and Optimization*, 51(5):3558–3591, 2013. (Cited on pages 42 et 51.)

- [53] Pierre Cardaliaguet and Alessio Porretta. Long time behavior of the master equation in mean field game theory. *Analysis & PDE*, 12(6):1397–1453, 2019. (Cited on page 42.)
- [54] Guillaume Carlier and Ivar Ekeland. Matching for teams. *Economic theory*, 42(2):397–418, 2010. (Cited on page 81.)
- [55] Guillaume Carlier and Maxime Laborde. A differential approach to the multi-marginal Schrödinger system. *SIAM Journal on Mathematical Analysis*, 52(1):709–717, 2020. (Cited on page 13.)
- [56] René Carmona and François Delarue. Mean field forward-backward stochastic differential equations. *Electronic Communications in Probability*, 18, 2013. (Cited on page 26.)
- [57] René Carmona and François Delarue. Probabilistic analysis of mean-field games. *SIAM Journal on Control and Optimization*, 51(4):2705–2734, 2013. (Cited on pages 25 et 26.)
- [58] René Carmona and François Delarue. Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *Annals of Probability*, 43(5):2647–2700, 2015. (Cited on pages 25, 26 et 32.)
- [59] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2018. (Cited on pages 25 et 26.)
- [60] René Carmona, François Delarue, and Aimé Lachapelle. Control of McKean–Vlasov dynamics versus mean field games. *Mathematics and Financial Economics*, 7(2):131–166, 2013. (Cited on page 33.)
- [61] José Carrillo, Robert McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana*, 19(3):971–1018, 2003. (Cited on pages 25 et 51.)
- [62] José Antonio Carrillo, Stefano Lisini, Giuseppe Savaré, and Dejan Slepčev. Nonlinear mobility continuity equations and generalized displacement convexity. *Journal of Functional Analysis*, 258(4):1273–1309, 2010. (Cited on pages 40 et 56.)
- [63] Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021.
- [64] Patrick Cattiaux, Arnaud Guillin, and Florent Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability Theory and Related Fields*, 140(1-2):19–40, 2008. (Cited on page 25.)
- [65] Patrick Cattiaux, Arnaud Guillin, Pierre Monmarché, and Chaoren Zhang. Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *Journal of Functional Analysis*, 277(11):108288, 2019. (Cited on pages 52 et 53.)
- [66] C Cedric Villani. *Hypoconvexity*. Number 949-951. American Mathematical Soc., 2009. (Cited on pages 6, 52 et 53.)
- [67] Y. Chen, T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *preprint arXiv:1412.4430*, 2014. (Cited on page 24.)
- [68] Yongxin Chen, Giovanni Conforti, and Tryphon Georgiou. Measure-valued spline curves: An optimal transport viewpoint. *SIAM Journal on Mathematical Analysis*, 50(6):5947–5968, 2018. (Cited on pages 77, 83 et 85.)
- [69] Yongxin Chen, Giovanni Conforti, Tryphon Georgiou, and Luigia Ripani. Multi-marginal Schrödinger bridges. In *International Conference on Geometric Science of Information*, pages 725–732. Springer, 2019.
- [70] Sinho Chewi, Julien Clancy, Thibaut Le Gouic, Philippe Rigollet, George Stepaniants, and Austin Stromme. Fast and smooth interpolation on Wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3061–3069. PMLR, 2021. (Cited on pages 77 et 85.)

- [71] Alberto Chiarini, Giovanni Conforti, Giacomo Greco, and Zhenjie Ren. Entropic turnpike estimates for the kinetic schrödinger problem. *arXiv preprint arXiv:2108.09161*, 2021.
- [72] Alberto Chiarini, Giovanni Conforti, and Luca Tamanini. Schrödinger problem for lattice gases: A heuristic point of view. In *International Conference on Geometric Science of Information*, pages 891–899. Springer, 2021. (Cited on pages 28 et 56.)
- [73] Lenaïc Chizat, Pierre Roussillon, Flavien Léger, François-Xavier Vialard, and Gabriel Peyré. Faster Wasserstein distance estimation with the Sinkhorn divergence. *Advances in Neural Information Processing Systems*, 33, 2020. (Cited on page 48.)
- [74] Shui-Nee Chow, Wuchen Li, Chenchen Mou, and Haomin Zhou. Dynamical Schrödinger bridge problems on graphs. *Journal of Dynamics and Differential Equations*, pages 1–20, 2021. (Cited on page 24.)
- [75] Shui-Nee Chow, Wuchen Li, and Haomin Zhou. A discrete schrödinger equation via optimal transport on graphs. *Journal of Functional Analysis*, 276(8):2440–2469, 2019. (Cited on page 24.)
- [76] Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. Long-time behaviour of entropic interpolations. *forthcoming in Potential Analysis*, 2020.
- [77] Gauthier Clerc, Giovanni Conforti, and Ivan Gentil. On the variational interpretation of local logarithmic Sobolev inequalities. *forthcoming in Annales de la faculté des Sciences de Toulouse*, 2020. (Cited on pages 50 et 54.)
- [78] Maria Colombo, Luigi De Pascale, and Simone Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canadian Journal of Mathematics*, 67(2):350–368, 2015. (Cited on page 15.)
- [79] Giovanni Conforti. Bridges of Markov counting processes: quantitative estimates. *Electronic Communications in Probability*, 21, 2016.
- [80] Giovanni Conforti. Fluctuations of bridges, reciprocal characteristics and concentration of measure. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 54(3):1432–1463, 2018.
- [81] Giovanni Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probability Theory and Related Fields*, 174(1-2):1–47, 2019. (Cited on pages 23, 43, 46 et 47.)
- [82] Giovanni Conforti. A probabilistic approach to convex (ϕ) -entropy decay for Markov chains. *arXiv preprint arXiv:2004.10850*, *forthcoming in Annals of Applied Probability*, 2020. (Cited on pages 59, 60, 62, 71 et 72.)
- [83] Giovanni Conforti, Paolo Dai Pra, and Sylvie Roelly. Reciprocal class of jump processes. *Journal of Theoretical probability*, 30(2):551–580, 2017.
- [84] Giovanni Conforti, Stefano De Marco, and Jean-Dominique Deuschel. On small-noise equations with degenerate limiting system arising from volatility models. In *Large Deviations and Asymptotic Methods in Finance*, pages 473–505. Springer, 2015.
- [85] Giovanni Conforti, Anna Kazeykina, and Zhenjie Ren. Game on random environment, mean-field Langevin system and neural networks. *arXiv preprint arXiv:2004.02457*, *forthcoming in Mathematics of Operations Research*, 2020.
- [86] Giovanni Conforti, Tetiana Kosenkova, and Sylvie Roelly. Conditioned point processes with application to lévy bridges. *Journal of Theoretical Probability*, 32(4):2111–2134, 2019.
- [87] Giovanni Conforti, Richard Kraaij, and Daniela Tonon. Hamilton Jacobi equation for controlled gradient flows: the comparison principle. *In preparation*. (Cited on pages 31 et 32.)

- [88] Giovanni Conforti and Christian Léonard. Reciprocal classes of random walks on graphs. *Stochastic Processes and their Applications*, 127(6):1870–1896, 2017.
- [89] Giovanni Conforti and Christian Léonard. Time reversal of Markov processes with jumps under a finite entropy condition. *forthcoming in Stochastic processes and Applications*, 2021.
- [90] Giovanni Conforti, Christian Léonard, Rüdiger Murr, and Sylvie Roelly. Bridges of markov counting processes. reciprocal classes and duality formulas. *Electronic Communications in Probability*, 20, 2015.
- [91] Giovanni Conforti and Michele Pavon. Extremal flows in Wasserstein space. *Journal of Mathematical Physics*, 59(6):063502, 2018.
- [92] Giovanni Conforti and Laurent Pfeiffer. A coupling approach to the turnpike property in stochastic control. *In preparation*. (Cited on pages 55 et 56.)
- [93] Giovanni Conforti and Luigia Ripani. Around the entropic Talagrand inequality. *Bernoulli*, 26(2):1431–1452, 2020. (Cited on pages 45 et 46.)
- [94] Giovanni Conforti and Sylvie Roelly. Bridge mixtures of random walks on an Abelian group. *Bernoulli*, 23(3):1518–1537, 2017.
- [95] Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):964–1008, 2021. (Cited on page 23.)
- [96] Giovanni Conforti and Max Von Renesse. Couplings, gradient estimates and logarithmic Sobolev inequality for Langevin bridges. *Probability Theory and Related Fields*, 172(1):493–524, 2018.
- [97] Andrea Cosso, Marco Fuhrman, and Huyen Pham. Long time asymptotics for fully nonlinear Bellman equations: a backward SDE approach. *Stochastic Processes and their Applications*, 126(7):1932–1973, 2016. (Cited on page 42.)
- [98] M. G. Crandall and P.-L. Lions. Hamilton-Jacobi equations in infinite dimensions VI: Nonlinear A and Tataru’s method refined. In *Evolution equations, control theory, and biomathematics (Han sur Lesse, 1991)*, volume 155 of *Lecture Notes in Pure and Appl. Math.*, pages 51–89. Dekker, New York, 1994. (Cited on pages 33 et 34.)
- [99] Michael G Crandall and Pierre-Louis Lions. Hamilton-Jacobi equations in infinite dimensions, III. *Journal of Functional Analysis*, 68(2):214–247, 1986. (Cited on page 32.)
- [100] Ana Bela Cruzeiro and Jean Claude Zambrini. Malliavin calculus and Euclidean quantum mechanics. I. Functional calculus. *Journal of Functional Analysis*, 96(1):62–95, 1991. (Cited on pages 24 et 26.)
- [101] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975. (Cited on page 13.)
- [102] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013. (Cited on page 85.)
- [103] Paolo Dai Pra and Gustavo Posta. Entropy decay for interacting systems via the Bochner-Bakry-Émery approach. *Electronic Journal of Probability*, 18(52):1–21, 2013. (Cited on pages 60, 63, 66, 69 et 71.)
- [104] Donald Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(4):247–308, 1987. (Cited on pages 4 et 7.)
- [105] M.P. do Carmo. *Riemannian Geometry*. Mathematics (Boston, Mass.). Birkhäuser, 1992. (Cited on page 23.)
- [106] Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. A new class of transport distances between measures. *Calculus of Variations and Partial Differential Equations*, 34(2):193–231, 2009. (Cited on pages 30 et 40.)

- [107] MD Donsker and SRS Varadhan. Large deviations from a hydrodynamic scaling limit. *Communications on Pure and Applied Mathematics*, 42(3):243–270, 1989. (Cited on page 29.)
- [108] Manh Hong Duong, Vaios Laschos, and Michiel Renger. Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(4):1166–1188, 2013. (Cited on page 48.)
- [109] Martin Dyer and Catherine Greenhill. On Markov chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000. (Cited on page 71.)
- [110] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probability Theory and Related Fields*, 166(3-4):851–886, 2016. (Cited on pages 54 et 55.)
- [111] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability*, 47(4):1982–2010, 2019. (Cited on pages 52 et 54.)
- [112] Michel Émery. En cherchant une caractérisation variationnelle des martingales. In *Séminaire de Probabilités XXII*, pages 147–154. Springer, 1988. (Cited on page 26.)
- [113] Matthias Erbar, Max Fathi, and André Schlichting. Entropic curvature and convergence to equilibrium for mean-field dynamics on discrete spaces. *arXiv preprint arXiv:1908.03397*, 2019. (Cited on page 60.)
- [114] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali. Ricci curvature bounds for weakly interacting markov chains. *Electronic Journal of Probability*, 22, 2017. (Cited on pages 60, 66 et 71.)
- [115] Matthias Erbar and Jan Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Archive for Rational Mechanics and Analysis*, 206(3):997–1038, 2012. (Cited on pages 60, 61, 66 et 72.)
- [116] Matthias Erbar, Jan Maas, and Michiel Renger. From large deviations to wasserstein gradient flows in multiple dimensions. *Electronic Communications in Probability*, 20:12 pp., 2015. (Cited on page 48.)
- [117] Matthias Erbar, Jan Maas, and Prasad Tetali. Discrete Ricci curvature bounds for Bernoulli-Laplace and random transposition models. *Annales de la Faculté des sciences de Toulouse: Mathématiques*, 24(4):781–800, 2015. (Cited on page 60.)
- [118] Carlos Esteve, Borjan Geshkovski, Dario Pighin, and Enrique Zuazua. Large-time asymptotics in deep learning. *arXiv preprint arXiv:2008.02491*, 2020. (Cited on page 42.)
- [119] Max Fathi and Jan Maas. Entropic Ricci curvature bounds for discrete interacting systems. *The Annals of Applied Probability*, 26(3):1774–1806, 2016. (Cited on pages 60 et 66.)
- [120] Jin Feng. Large deviation for diffusions and hamilton-jacobi equation in hilbert spaces. *The Annals of Probability*, 34(1):321–385, 2006. (Cited on pages 32, 33, 34 et 40.)
- [121] Jin Feng and Markos Katsoulakis. A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions. *Archive for Rational Mechanics and Analysis*, 192(2):275–310, 2009. (Cited on pages 33 et 40.)
- [122] Jin Feng and Thomas G. Kurtz. *Large Deviations for Stochastic Processes*. American Mathematical Society, 2006. (Cited on pages 32 et 33.)
- [123] Jin Feng, Toshio Mikami, and Johannes Zimmer. A Hamilton-Jacobi PDE associated with hydrodynamic fluctuations from a nonlinear diffusion. *preprint; ArXiv:1903.00052*, 2019. (Cited on page 33.)
- [124] Markus Fischer. On the form of the large deviation rate function for the empirical measures of weakly interacting systems. *Bernoulli*, 20(4):1765–1801, 2014. (Cited on page 16.)

- [125] Hans Föllmer. Time reversal on Wiener space. In *Stochastic Processes—Mathematics and Physics*, pages 119–129. Springer, 1986. (Cited on page 52.)
- [126] Hans Föllmer. Random fields and diffusion processes. In *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pages 101–203. Springer, 1988. (Cited on page 4.)
- [127] Joaquin Fontbona and Benjamin Jourdain. A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *The Annals of Probability*, 44(1):131–170, 2016. (Cited on page 52.)
- [128] Robert Fortet. Résolution d’un système d’équations de M. Schrödinger. *J. Math. Pure Appl.* IX, 1:83–105, 1940. (Cited on page 13.)
- [129] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996. (Cited on page 12.)
- [130] Wilfrid Gangbo and Alpár R Mészáros. Global well-posedness of master equations for deterministic displacement convex potential mean field games. *arXiv preprint arXiv:2004.01660*, 2020. (Cited on page 33.)
- [131] Wilfrid Gangbo, Alpár R Mészáros, Chenchen Mou, and Jianfeng Zhang. Mean field games master equations with non-separable hamiltonians and displacement monotonicity. *arXiv preprint arXiv:2101.12362*, 2021. (Cited on page 33.)
- [132] Wilfrid Gangbo and Andrzej Świąch. Existence of a solution to an equation arising from the theory of mean field games. *Journal of Differential Equations*, 259(11):6573–6643, 2015. (Cited on page 33.)
- [133] Wilfrid Gangbo and Andrzej Świąch. Metric viscosity solutions of hamilton–jacobi equations depending on local slopes. *Calculus of Variations and Partial Differential Equations*, 54(1):1183–1218, 2015. (Cited on page 33.)
- [134] Wilfrid Gangbo and Adrian Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019. (Cited on page 33.)
- [135] Ivan Gentil, Christian Léonard, and Luigia Ripani. About the analogy between optimal transport and minimal entropy. *Annales de la faculté des sciences de Toulouse Sér. 6*, 26(3):569–700, 2017. (Cited on page 24.)
- [136] Ivan Gentil, Christian Léonard, and Luigia Ripani. Dynamical aspects of the generalized schrödinger problem via otto calculus—a heuristic point of view. *Revista Matemática Iberoamericana*, 36(4):1071–1112, 2020. (Cited on pages 24, 27 et 51.)
- [137] N. Gigli and L. Tamanini. Second order differentiation formula on $RCD^*(K, N)$ spaces. Accepted at JEMS, arXiv:1802.02463, (2018). (Cited on page 32.)
- [138] Nicola Gigli. *Second order analysis on $(P_2(M), W_2)$* . American Mathematical Soc., 2012. (Cited on pages 17, 19, 20, 21 et 22.)
- [139] Nicola Gigli and Luca Tamanini. Benamou–Brenier and duality formulas for the entropic cost on $RCD^*(k, N)$ spaces. *Probability Theory and Related Fields*, 176(1-2):1–34, 2020. (Cited on page 5.)
- [140] Nathael Gozlan and Nicolas Juillet. On a mixture of Brenier and Strassen theorems. *Proceedings of the London Mathematical Society*, 120(3):434–463, 2020. (Cited on page 15.)
- [141] Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Revista Matemática Iberoamericana*, 30(1):133–163, 2014. (Cited on page 32.)
- [142] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probability Theory and Related Fields*, 160(1-2):47–94, 2014. (Cited on page 60.)

- [143] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327–3405, 2017. (Cited on page 46.)
- [144] P Jameson Graber, Alpár R Mészáros, Francisco J Silva, and Daniela Tonon. The planning problem in mean field games as regularized mass transport. *Calculus of Variations and Partial Differential Equations*, 58(3):115, 2019. (Cited on pages 26 et 28.)
- [145] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. The kinetic Fokker-Planck equation with mean field interaction. *Journal de Mathématiques Pures et Appliquées*, 150:1–23, 2021. (Cited on page 52.)
- [146] Arnaud Guillin and Feng-Yu Wang. Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality. *Journal of Differential Equations*, (253):20–40, 2012. (Cited on page 53.)
- [147] Gaoyue Guo, Xiaolu Tan, and Nizar Touzi. On the monotonicity principle of optimal skorokhod embedding problem. *SIAM Journal on Control and Optimization*, 54(5):2478–2489, 2016. (Cited on page 15.)
- [148] Jonathan Hermon and Justin Salez. Entropy dissipation estimates for inhomogeneous zero-range processes. *arXiv preprint arXiv:1903.01410*, 2019. (Cited on pages 63 et 72.)
- [149] John C Holladay. A smoothest curve approximation. *Mathematical tables and other aids to computation*, 11(60):233–243, 1957. (Cited on page 78.)
- [150] Oliver Johnson. A discrete log-sobolev inequality under a bakry-émery type condition. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 53(4):1952–1970, 2017. (Cited on page 66.)
- [151] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998. (Cited on pages 8, 22, 24 et 27.)
- [152] Ansgar Jüngel and Wen Yue. Discrete Beckner inequalities via the Bochner-Bakry-Émery approach for markov chains. *The Annals of Applied Probability*, 27(4):2238–2269, 2017. (Cited on pages 60, 61 et 66.)
- [153] Ioannis Karatzas, Walter Schachermayer, and Bertram Tschiderer. Pathwise Otto calculus. *arXiv preprint arXiv:1811.08686*, 2018. (Cited on pages 27 et 52.)
- [154] Claude Kipnis, Stefano Olla, and Srinvasa Varadhan. Hydrodynamics and large deviation for simple exclusion processes. *Communications on Pure and Applied Mathematics*, 42(2):115–137, 1989. (Cited on pages 7 et 29.)
- [155] M. Knott and C. Smith. On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.*, 40(2):328–334, 1992. (Cited on page 12.)
- [156] A. J. Krener. Reciprocal diffusions and stochastic differential equations of second order. *Stochastics*, 107(4):393–422, 1988. (Cited on page 24.)
- [157] A.J. Krener. Reciprocal diffusions in flat space. *Probability Theory and Related Fields*, 107(2):243–281, 1997. (Cited on page 24.)
- [158] Flavien Léger and Wuchen Li. Hopf-Cole transformation via generalized Schrödinger bridge problem. *Journal of Differential Equations*, 274:788–827, 2021. (Cited on pages 24 et 51.)
- [159] C. Léonard. From the Schrödinger problem to the Monge-Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012. (Cited on pages 44 et 47.)
- [160] Christian Léonard. Minimizers of energy functionals. *Acta Mathematica Hungarica*, 93(4):281–325, 2001. (Cited on page 13.)
- [161] Christian Léonard. Entropic projections and dominating points. *ESAIM: Probability and Statistics*, 14:343–381, 2010. (Cited on page 13.)

- [162] Christian Léonard. Girsanov theory under a finite entropy condition. In C. Donati-Martin, A. Lejay, and A. Rouault, editors, *Séminaire de Probabilités XLIV*, volume 2046 of *Lecture Notes in Mathematics*, pages 429–465. Springer, 2012. (Cited on page 4.)
- [163] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. (Cited on pages 5, 13, 15 et 30.)
- [164] Christian Léonard. On the convexity of the entropy along entropic interpolations. In *Measure Theory in Non-Smooth Spaces*, pages 194–242. Sciendo Migration, 2017. (Cited on page 44.)
- [165] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017. (Cited on page 71.)
- [166] Peter Li and Shing-Tung Yau. On the parabolic kernel of the Schrödinger operator. *Acta Math.*, 156(3-4):153–201, 1986. (Cited on page 50.)
- [167] Wuchen Li, Jianfeng Lu, and Li Wang. Fisher information regularization schemes for Wasserstein gradient flows. *Journal of Computational Physics*, 416:109449, 2020. (Cited on page 24.)
- [168] Michael Luby and Eric Vigoda. Fast convergence of the Glauber dynamics for sampling independent sets. *Random Structures & Algorithms*, 15(3-4):229–241, 1999. (Cited on page 71.)
- [169] Jin Ma, J-M Morel, and Jiongmin Yong. *Forward-backward stochastic differential equations and their applications*. Number 1702. Springer Science & Business Media, 1999. (Cited on page 25.)
- [170] Jan Maas. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011. (Cited on pages 60, 66 et 72.)
- [171] Robert McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997. (Cited on pages 23, 40, 44, 51 et 83.)
- [172] Lionel McKenzie. Turnpike theorems for a generalized Leontief model. *Econometrica (pre-1986)*, 31(1, 2):165, 1963. (Cited on page 42.)
- [173] Alexander Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calculus of Variations and Partial Differential Equations*, 48(1-2):1–31, 2013. (Cited on pages 60 et 66.)
- [174] Toshio Mikami. Monge’s problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability Theory and Related Fields*, 129(2):245–260, 2004. (Cited on page 44.)
- [175] Pierre Monmarché. Almost sure contraction for diffusions on \mathbb{R}^d . application to generalised langevin diffusions. *arXiv preprint arXiv:2009.10828*, 2020. (Cited on page 52.)
- [176] Léonard Monsaingeon, Luca Tamanini, and Dmitry Vorotnikov. The dynamical Schrödinger problem in abstract metric spaces. *arXiv preprint arXiv:2012.12005*, 2020. (Cited on pages 24 et 46.)
- [177] Matteo Muratori and Giuseppe Savaré. Gradient flows and evolution variational inequalities in metric spaces. i: Structural properties. *Journal of Functional Analysis*, 278(4):108347, 2020. (Cited on pages 32 et 34.)
- [178] Lyle Noakes, Greg Heinzinger, and Brad Paden. Cubic splines on curved spaces. *IMA Journal of Mathematical Control and Information*, 6(4):465–473, 1989. (Cited on page 83.)
- [179] J.R. Norris. *Markov Chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1999. (Cited on page 63.)
- [180] Marcel Nutz and Florian Stebegg. Canonical supermartingale couplings. *The Annals of Probability*, 46(6):3351–3398, 2018. (Cited on page 15.)

- [181] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009. (Cited on pages 60 et 61.)
- [182] Felix Otto. The geometry of dissipative equations: the porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001. (Cited on page 8.)
- [183] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000. (Cited on pages 45, 48 et 50.)
- [184] Soumik Pal. On the difference between entropic cost and the optimal transport cost. *arXiv preprint arXiv:1905.12206*, 2019. (Cited on page 48.)
- [185] Soumik Pal and Ting-Kam Leonard Wong. The geometry of relative arbitrage. *Math. Financ. Econ.*, 10(3):263–293, 2016. (Cited on page 15.)
- [186] Soumik Pal and Ting-Kam Leonard Wong. Multiplicative Schrödinger problem and the Dirichlet transport. *Probability Theory and Related Fields*, 178(1):613–654, 2020. (Cited on page 8.)
- [187] Brendan Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 43(3):529–536, 2012. (Cited on pages 15, 82 et 85.)
- [188] Brendan Pass. Multi-marginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015. (Cited on pages 82 et 85.)
- [189] Teemu Pennanen and Ari-Pekka Perkkiö. Convex duality in nonlinear optimal transport. *Journal of Functional Analysis*, 277(4):1029–1060, 2019. (Cited on page 13.)
- [190] Huyên Pham and Xiaoli Wei. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM: Control, Optimisation and Calculus of Variations*, 24(1):437–461, 2018. (Cited on page 32.)
- [191] Aldo Pratelli. On the sufficiency of c -cyclical monotonicity for optimality of transport plans. *Mathematische Zeitschrift*, 258(3):677–690, 2008. (Cited on page 12.)
- [192] Luigia Ripani. Convexity and regularity properties for entropic interpolations. *Journal of Functional Analysis*, 277(2):368–391, 2019. (Cited on page 49.)
- [193] B Roynette, P Vallois, and Marc Yor. Some penalisations of the wiener measure. *Japanese Journal of Mathematics*, 1(1):263–290, 2006. (Cited on page 24.)
- [194] L. Rüschendorf. On c -optimal random variables. *Statist. Probab. Lett.*, 27(3):267–270, 1996. (Cited on page 12.)
- [195] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I-projections. *Statistics & probability letters*, 17(5):369–375, 1993. (Cited on page 13.)
- [196] Paul-Marie Samson. Entropic curvature on graphs along Schrödinger bridges at zero temperature. *arXiv preprint arXiv:2003.05179*, 2020. (Cited on page 60.)
- [197] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137(2):519–529, 2009. (Cited on page 12.)
- [198] Erwin Schrödinger. La théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst Henri Poincaré*, (2):269 – 310, 1932. (Cited on page 3.)
- [199] Michel Talagrand. Transportation cost for Gaussian and other product measures. *Geometric & Functional Analysis GAFA*, 6(3):587–600, May 1996. (Cited on page 45.)
- [200] Denis Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8(2):163–198, 2002. (Cited on page 52.)

- [201] Luca Tamanini. A generalization of Costa's entropy power inequality. *arXiv preprint arXiv:2012.12230*, 2020. (Cited on page 49.)
- [202] Hiroshi Tanaka. Limit theorems for certain diffusion processes with interaction. In *North-Holland Mathematical Library*, volume 32, pages 469–488. Elsevier, 1984. (Cited on page 25.)
- [203] Daniel Tataru. Viscosity solutions of Hamilton-Jacobi equations with unbounded nonlinear terms. *Journal of Mathematical Analysis and Applications*, 163(2):345–392, 1992. (Cited on pages 33 et 34.)
- [204] Daniel Tataru. Viscosity solutions for Hamilton-Jacobi equations with unbounded nonlinear term: a simplified approach. *Journal of Differential Equations*, 111(1):123–146, 1994. (Cited on pages 33 et 34.)
- [205] M. Thiell. Second order stochastic differential equations and non-Gaussian reciprocal diffusions. *Probability Theory and Related Fields*, 97(1-2):231–257, 1993. (Cited on page 24.)
- [206] Emmanuel Trélat and Enrique Zuazua. The turnpike property in finite-dimensional nonlinear optimal control. *Journal of Differential Equations*, 258(1):81–114, 2015. (Cited on pages 9, 42, 53 et 56.)
- [207] Ali Süleyman Üstünel and Moshe Zakai. Sufficient conditions for the invertibility of adapted perturbations of identity on the Wiener space. *Probability theory and related fields*, 139(1-2):207–234, 2007. (Cited on page 26.)
- [208] Eric Vigoda. A note on the Glauber dynamics for sampling independent sets. *The Electronic Journal of Combinatorics*, 8(1):1–8, 2001. (Cited on page 71.)
- [209] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003. (Cited on page 12.)
- [210] Max von Renesse. An optimal transport view of Schrödinger's equation. *Canadian Mathematical Bulletin*, 55(4):858–869, 2012. (Cited on pages 22 et 24.)
- [211] Max von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Communications on pure and applied mathematics*, 58(7):923–940, 2005. (Cited on pages 23, 44 et 60.)
- [212] Cong Wu and Jianfeng Zhang. Viscosity solutions to parabolic master equations and McKean–Vlasov SDEs with closed-loop controls. *Annals of Applied Probability*, 30(2):936–986, 2020. (Cited on page 33.)
- [213] Kunio Yasue. Stochastic calculus of variations. *Journal of functional Analysis*, 41(3):327–340, 1981. (Cited on page 24.)
- [214] Jean-Claude Zambrini. Variational processes and stochastic versions of mechanics. *Journal of Mathematical Physics*, 27(9):2307–2330, 1986. (Cited on page 24.)
- [215] Alexander Zaslavski. *Turnpike properties in the calculus of variations and optimal control*, volume 80. Springer Science & Business Media, 2005. (Cited on page 42.)
- [216] Alexander Zaslavski. *Turnpike conditions in infinite dimensional optimal control*. Springer, 2019. (Cited on page 42.)