Essays in transportation inequalities, entropic gradient flows and mean field approximations

Lane Chun Lanston Yeung

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

Abstract

Essays in transportation inequalities, entropic gradient flows and mean field approximations

Lane Chun Lanston Yeung

This thesis consists of four chapters. In Chapter 1, we focus on a class of transportation inequalities known as the transportation-information inequalities. These inequalities bound optimal transportation costs in terms of relative Fisher information, and are known to characterize certain concentration properties of Markov processes around their invariant measures. We provide a characterization of the quadratic transportation-information inequality in terms of a dimension-free concentration property for i.i.d. copies of the underlying Markov process, identifying the precise high-dimensional concentration property encoded by this inequality. We also illustrate how this result is an instance of a general convex-analytic tensorization principle.

In Chapter 2, we study the entropic gradient flow property of McKean–Vlasov diffusions via a stochastic analysis approach. We formulate a trajectorial version of the relative entropy dissipation identity for these interacting diffusions, which describes the rate of relative entropy dissipation along every path of the diffusive motion. As a first application, we obtain a new interpretation of the gradient flow structure for the granular media equation. Secondly, we show how the trajectorial approach leads to a new derivation of the HWBI inequality.

In Chapter 3, we further extend the trajectorial approach to a class of degenerate diffusion equations that includes the porous medium equation. These equations are posed on a bounded domain and are subject to no-flux boundary conditions, so that their corresponding probabilistic representations are stochastic differential equations with normal reflection on the boundary. Our

stochastic analysis approach again leads to a new derivation of the Wasserstein gradient flow property for these nonlinear diffusions, as well as to a simple proof of the HWI inequality in the present context.

Finally, in Chapter 4, we turn our attention to mean field approximation – a method widely used to study the behavior of large stochastic systems of interacting particles. We propose a new approach to deriving quantitative mean field approximations for any strongly log-concave probability measure. Our framework is inspired by the recent theory of nonlinear large deviations, for which we offer an efficient non-asymptotic perspective in log-concave settings based on functional inequalities. We discuss three implications, in the contexts of continuous Gibbs measures on large graphs, high-dimensional Bayesian linear regression, and the construction of decentralized near-optimizers in high-dimensional stochastic control problems.

Table of Contents

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisors, Professors Daniel Lacker and Ioannis Karatzas, for their invaluable guidance and generous support throughout my doctoral studies. They are exemplary scholars, brilliant researchers, and excellent teachers. Many thanks for accepting me into your groups, introducing me to fascinating topics in probability, and your mentorship over the years. It has been a privilege to be your student.

I am indebted to Dan for his immense support during every stage of my doctoral studies. I am very grateful that he has shared with me his vast and deep knowledge of probability. I have learned a lot from him in research and academic careers. Despite his busy life, he has been generous with his time, provided much incisive feedback, and helped me navigate many difficult times.

I am also indebted to Ioannis for all his advice and huge encouragement throughout the years. I will never forget his exceptionally captivating lectures, which sparked my interest in probability. Thank you for your beautiful papers, which provided the basis for my research. Many thanks for guiding me on my research, and for sharing your wisdom and insights.

I would like to thank my thesis committee members, Professors Sumit Mukherjee, Kaizheng Wang, and Wenpin Tang for their valuable comments and suggestions on my research. Also, thanks to Sumit for your collaboration and to Sumit and Kaizheng for your wonderful lectures on large deviations and high-dimensional probability. Furthermore, I am very grateful to Professor Walter Schachermayer for his kind support and enlightening talks during his visits.

I am also thankful for being part of the monthly research meetings led by Ioannis and Dr. E. Robert Fernholz. I greatly benefited from the fascinating presentations made by Ioannis and other participants, including Professors Tomoyuki Ichiba, Walter Schachermayer, Kasper Larsen, Kim Weston, Martin Larsson, Andrey Sarantsev, and Dr. Donghan Kim.

I am fortunate to have fantastic coauthors, Bertram Tschiderer and Donghan Kim. Thanks for the collaboration on the papers that constitute part of this thesis. Thanks also go to Fuzhong Zhou for the project we are working on. I thank Donghan and Dr. Jiacheng Zhang for sharing with me their academic experiences. During my time at Columbia, I benefited from various seminars organized by Georgy Gaitsgori and Richard Groenewald, and from discussions with my peers, including Miguel Ángel Garrido, Luc Le Flem, Agathe Soret, Zhi Li, Camilo Hernández, Hindy Drillick, Abhishek Tilva, Xujia Liu, Xiao Lei, Luofeng Liao, Haoxian Chen, Ruizhe Jia, and Shihao Yu.

I am grateful to the professors at the IEOR Department, particularly Professors Agostino Capponi, Xunyu Zhou, Emanuel Derman, and Ton Dieker. I also thank Professor Raymond Chan and Dr. Alfred Ma for their tremendous support and encouragement over the years. I would also like to thank the staff members in the IEOR department, especially Lizbeth Morales, Winsor Yang, and Kristen Maynor.

Thanks are also due to the friends I met in Columbia and New York, including Albert Law, Yu Fu Wong, Lauren Tang, Brian Ling, George Chu, Kevin Kwan, Pak-Hin Lee, and Jack Cheng, for the fun times we spent together. I also thank my roommate Daniel Fawcett and our friend Andy Tan for our great memories.

Last but most importantly, I thank my family for their unconditional support and their understanding.

Dedication

To my family

Introduction

In this thesis, we present new results in transportation inequalities, entropic gradient flows and mean field approximations, important topics in probability and stochastic analysis. We provide the relevant background and a summary of contributions below.

1. Dimension-free characterization of transportation inequalities

Since Marton [1] and Talagrand [2] introduced the celebrated *transportation-entropy* inequalities, transportation inequalities have been a very active area in probability and functional analysis. Subsequent work by Bobkov–Götze [3] and Gozlan–Léonard [4] then unfolds their profound connections to the concentration of measure phenomenon and large deviation principles. The first topic studied in this thesis is a newer class of inequalities that bound Wasserstein distances in terms of Fisher information. These *transportation-information inequalities* were introduced in [5] and studied further in [6, 7, 8, 9, 10, 11]. They are known to characterize the concentration of Markov processes around their invariant measures.

We show in Chapter 1 that, for an ergodic Markov process on a Polish space, the quadratic transportation-information inequality for the invariant measure is equivalent to a dimension-free rate of convergence to equilibrium for the product Markov process, in which each coordinate evolves independently according to the original process. Remarkably, our characterization exactly parallels Gozlan's characterization of the quadratic transportation-entropy inequality in [12]. The proof is based on a new Laplace-type principle for the operator norms of Feynman-Kac semigroups, which is of independent interest.

Lastly, we illustrate how both our theorem and (a form of) Gozlans are instances of a general convex-analytic tensorization principle. This chapter is based on the paper [13] joint with Daniel Lacker.

2. Trajectorial approach to Otto calculus

In the study of the long-time behavior of Markov processes, an important property we can analyze is the trend to equilibrium, or more precisely, the convergence of the time-marginal laws to the stationary distribution. For certain diffusions, when we view their time-marginal distributions geometrically on the Wasserstein space, they actually evolve in the steepest possible direction. For example, Jordan, Kinderlehrer and Otto showed in their seminal work [14] that the Fokker-Planck diffusion is the Wasserstein gradient flow of steepest descent. Recently, a trajectorial approach to the result of [14] and to the resulting "Otto calculus" was developed in [15]. This approach operates at the level of the particles, and is based on a detailed perturbation analysis, as well as on tools from time reversal in stochastic calculus.

2.1 McKean-Vlasov diffusions

This trajectorial analysis was generalized in Chapter 2, where we consider a class of stochastic differential equations (SDEs) with mean field drift interaction. Theses SDEs are non-local (or non-linear) in the sense that the drift term depends on the *distribution* of the state variable. Non-local equations of this form arise in the modeling of weakly interacting diffusions, which gained prominence after the work of McKean [16]. As is well known from [17, 18, 19], the McKean-Vlasov diffusion can be characterized as a *gradient flow*. This is an optimality property stating that the curve of time-marginal distributions evolves in the steepest possible direction of the free energy functional with respect to the quadratic Wasserstein distance.

In Chapter 2, we provide a *trajectorial* approach to deriving this entropic steepest descent property. To this end, we first formulate a trajectorial analogue of the relative entropy dissipation identity, which describes exactly the rate of relative entropy dissipation along every single path of the diffusion, rather than at the level of their ensembles. Specifically, it is formulated in terms of the semimartingale decomposition of a relative entropy process into a martingale and a compensator, the latter of which can be viewed as a stochastic counterpart of the deterministic rate of relative entropy dissipation. Our stochastic analysis approach is based on time reversal of diffusions and Lions differential calculus over Wasserstein space.

A direct perturbation analysis then reveals the steepest descent property of the McKean–Vlasov diffusion. This is accomplished by considering a "perturbed" diffusion, constructed by adding a gradient potential to the drift of the original diffusion. By performing the same trajectorial analysis as before, we deduce that the Wasserstein metric slope in the original, unperturbed setting is always steeper than the one in the perturbed setting.

Additionally, our trajectorial approach also leads to a simple proof of the *HWBI inequality* [20, Theorem 4.2], which is an extension of the famous HWI inequality of Otto-Villani [21]. This chapter is based on the paper [22] joint with Bertram Tschiderer.

2.1 Porous medium diffusions

In Chapter 3, we further extend the trajectorial approach to a class of degenerate parabolic equations, that includes the *porous medium equation*. Different from Chapter 2, we pose them on a bounded domain and impose no-flux boundary condition. The corresponding probabilistic representations turn out to be SDEs with normal reflection on the boundary. Our stochastic analysis approach again leads to a new derivation of the Wasserstein gradient flow property for these nonlinear diffusions, as well as to a simple proof of the HWI inequality in the present context. A key difficulty in adapting the trajectorial approach to this setting stems from the degenerate parabolicity of the diffusion. This is tackled by restricting the initial condition to be nondegenerate, which ensures that the solution to the parabolic equation is smooth. This chapter is based on the paper [23] joint with Donghan Kim.

3. Mean field approximations

In Chapter 4, we shift our focus to mean field approximation–a powerful method widely used to study the behavior of large stochastic systems of interacting particles. The main reason that these systems are difficult to analyze is due to the interactions between particles, which propagate throughout the systems and make them complex and intractable very rapidly as the number of particles gets large. Mean field approximation substantially reduces this high dimensionality by approximating the original system with a closely related one, in which particles evolve independently and interactions take place simply with the average state of the system. While it was originally used as a heuristic to make rough predictions, a large body of work has since been devoted to rigorously justifying the asymptotic validity of such an approximation in various contexts.

In the discrete setting, this problem was studied in the groundbreaking paper of Chatterjee-Dembo [24], where they showed that the mean field approximation holds if the Gibbs measure has *low gradient complexity*, as measured by the metric entropy of the range of the gradient of the Hamiltonian. This result, along with a number of subsequent papers [25, 26, 27], resolved an open problem regarding subgraph counts in sparse random graphs, and had important applications to Ising models [28, 29, 30]. Alternative and typically more convenient measures of gradient complexity have since appeared, based on the notion of Gaussian-width [31, 32, 33] or Rademacher-width [34].

In this chapter, we propose an alternative approach to deriving quantitative mean field approximations for *continuous* Gibbs measures. It turns out that a natural condition for the mean field approximation to be valid in our setting is the *strong log-concavity* of the Gibbs measure. This provides us with the suitable regularity, such that when combined with functional inequalities (specifically, the log-Sobolev and the Poincaré inequalities), allows us to obtain simple bounds on the mean field approximation error. Our bound is typically simpler to work with then prior bounds involving covering numbers.

We discuss three applications of our framework. The first application concerns Gibbs measures with pairwise heterogeneous interactions. This class of Gibbs measures appears as invariant measures of locally interacting diffusion process. Under suitable concavity assumptions, our framework implies that the mean field approximation is valid for this class of Gibbs measures. Furthermore, if the interaction matrix is doubly stochastic or converges in cut metric, some additional precise asymptotic results and a law of large numbers are available. The second application is concerned with high-dimensional Bayesian linear regression, where we show that if the prior distribution is log-concave, then the posterior is mean field, i.e., close to a product distribution. Leveraging this, we also derive a law of large numbers for the posterior. The final application is concerned with a class of high-dimensional stochastic control problems, in which a large number of players cooperatively choose their drifts to maximize an expected reward minus a quadratic running cost. For a broad class of potentially asymmetric rewards, we show that there exist approximately optimal controls which are decentralized, in the sense that each player's control depends only on its own state and not the states of the other players. Moreover, the optimal decentralized controls can be constructed non-asymptotically, without reference to any mean field limit. This chapter is based on the paper [35] joint with Daniel Lacker and Sumit Mukherjee.

Chapter 1: A characterization of transportation-information inequalities for Markov processes in terms of dimension-free concentration

Inequalities between transportation costs and Fisher information are known to characterize certain concentration properties of Markov processes around their invariant measures. This chapter provides a new characterization of the quadratic transportation-information inequality $\mathcal{W}_2 I$ in terms of a dimension-free concentration property for i.i.d. (conditionally on the initial positions) copies of the underlying Markov process. This parallels Gozlan's characterization of the quadratic transportation-entropy inequality $\mathcal{W}_2 H$. The proof is based on a new Laplace-type principle for the operator norms of Feynman-Kac semigroups, which is of independent interest. Lastly, we illustrate how both our theorem and (a form of) Gozlan's are instances of a general convex-analytic tensorization principle. This chapter is based on the paper [13] joint with Daniel Lacker.

1.1 Introduction

There is by now a vast literature on the connections between concentration of measure and *transportation-entropy inequalities*, which bound Wasserstein distances in terms of relative entropy. See [36, 37, 38] for thorough discussions. In this chapter, we focus on a somewhat newer class of inequalities between Wasserstein distances and Fisher information, introduced in [5] and studied further in [6, 7, 8, 9, 10, 11]. These *transportation-information inequalities* characterize the concentration of Markov processes around their invariant measures. We show in this chapter that, for an ergodic Markov process on a Polish space E with invariant measure μ, the quadratic transportation-information inequality for μ is equivalent to a dimension-free rate of convergence to equilibrium for the natural Markov process associated with invariant measure $\mu^{\otimes n}$. We present this main result first, and then we discuss its close analogy with Gozlan's characterization of the quadratic transportation-entropy inequality [12], along with other related literature.

Let us first fix notation. Fix throughout the chapter a complete separable metric space (E, d). Denote by $B(E)$ the set of measurable and bounded real-valued functions on E. Let $\mathcal{P}(E)$ be the space of Borel probability measures on E, equipped with the topology of weak convergence.

For $\nu, \mu \in \mathcal{P}(E)$, the p-order Wasserstein distance is defined as usual by

$$\mathcal{W}_p(\nu, \mu) := \left(\inf_{\pi} \int_{E\times E} d^p(x, y)\, \pi\,(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}, \tag{1.1}$$

where the infimum is over all couplings π of ν and μ. (The value $+\infty$ is allowed.)

Fix a Borel probability measure μ on E. We work with a continuous-time E-valued Markov process, governed by the family $(\Omega, \mathcal{F}, (X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in E})$. The transition semigroup is denoted by $(\mathrm{P}_t)_{t\geq 0}$, defined as usual by $\mathrm{P}_t f(x) := \mathbb{E}_x[f(X_t)]$.

Assumptions 1.1.1. Throughout this chapter, we assume the following two conditions:

(1) The probability measure μ is ergodic and reversible for $(X_t)_{t\geq 0}$.

(2) The semigroup $(\mathrm{P}_t)_{t\geq 0}$ is strongly continuous on $L^2(\mu)$.

Let $\mathcal{L}$ denote the infinitesimal generator of $(\mathrm{P}_t)_{t\geq 0}$ with domain denoted by $\mathbb{D}(\mathcal{L}) \subset L^2(\mu)$. The corresponding Dirichlet form is defined by

$$\mathcal{E}(g, g) := -\int_E g\mathcal{L}g\, \mathrm{d}\mu, \quad \text{for } g \in \mathbb{D}(\mathcal{L}). \tag{1.2}$$

Under our standing assumptions, $\mathcal{E}$ is closable in the Hilbert space $L^2(\mu)$ and its closure $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ has domain $\mathbb{D}(\mathcal{E}) = \mathbb{D}(\sqrt{-\mathcal{L}})$ in $L^2(\mu)$. For $\nu \in \mathcal{P}(E)$, the *Fisher information* of ν with respect to μ is defined by

$$I(\nu \,|\, \mu) := \begin{cases} \mathcal{E}\left(\sqrt{f}, \sqrt{f}\right) & \text{if } \nu \ll \mu,\ \frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f, \text{ and } \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty & \text{otherwise.} \end{cases} \tag{1.3}$$

Example 1.1.2. A classical example comes from diffusion processes: Let E be a complete connected (finite-dimensional) Riemannian manifold equipped with its volume measure $\mathrm{d}x$. Let $V \in C^1(E)$ be such that $\mu(\mathrm{d}x) = e^{-V(x)}\mathrm{d}x$ defines a probability measure. Let $\mathcal{L} = \Delta - \nabla V \cdot \nabla$, where ∇ and Δ denote the usual gradient and Laplace-Beltrami operators, respectively. Then

$$\mathcal{E}(g,g) = \int_E |\nabla g|^2 \,\mathrm{d}\mu, \qquad \text{for all } g \in \mathbb{D}(\mathcal{E}) = H^1(E,\mu),$$

where $|\cdot|$ is the Riemannian norm, and $H^1(E,\mu)$ is the closure of the space of infinitely differentiable functions on E with respect to the Sobolev norm $g \mapsto \sqrt{\int_E (|g|^2 + |\nabla g|^2)\,\mathrm{d}\mu}$. In this case,

$$I(f\mu\,|\,\mu) = \int_E |\nabla\sqrt{f}|^2 \,\mathrm{d}\mu = \frac{1}{4}\int_E |\nabla \log f|^2 f \,\mathrm{d}\mu.$$

The transportation-information inequalities of interest in this chapter are the following:[1] For $C > 0$ and $p \geq 1$, we say that μ *satisfies the $\mathcal{W}_p I(C)$ inequality* if

$$\mathcal{W}_p^2(\mu,\nu) \leq C I(\nu\,|\,\mu), \quad \text{for all } \nu \in \mathcal{P}(E). \tag{1.4}$$

1.1.1 A known characterization of $\mathcal{W}_1 I$

In [5], characterizations of $\mathcal{W}_1 I$ are provided in terms of concentration properties for the Markov process. They make use of the Feynman-Kac semigroups $(\mathrm{P}_t^f)_{t\geq 0}$, defined (as in [5, 39, 40]) for $f \in B(E)$ by

$$\mathrm{P}_t^f g(x) \coloneqq \mathbb{E}_x\left[g(X_t)\exp\left(\int_0^t f(X_s)\mathrm{d}s\right)\right], \quad x \in E,$$

[1] Take note that different authors adopt different conventions regarding the constant C. For instance, [5] uses $4c^2$ where we use C.

for any measurable function g for which the expectation is well-defined. They make use also of the operator norm

$$\|\mathrm{P}_t^f\|_{L^2(\mu)} := \sup\left\{\left\|\mathrm{P}_t^f g\right\|_{L^2(\mu)} : g \geq 0,\ \int_E g^2 \,\mathrm{d}\mu \leq 1\right\}. \tag{1.5}$$

This coincides with the spectral radius of the bounded symmetric operator P_t^f. Moreover, a well known form of the Feynman-Kac formula states that $(\mathrm{P}_t^f)_{t\geq 0}$ is a strongly continuous semigroup with infinitesimal generator given by $g \mapsto \mathcal{L}g + fg$ with the same domain as $\mathcal{L}$; see, e.g., [41, Section III.19] or [42, Section 6.1].

Theorem 1.1.3 ([5, Corollary 2.5])**.** *Assume there exists $x_0 \in E$ such that $\int_E d^2(x, x_0)\,\mu(\mathrm{d}x) < \infty$. Let $C > 0$. The following are equivalent:*

(1) μ satisfies the $\mathcal{W}_1 I(C)$ inequality.

(2) For any $\lambda \in \mathbb{R}$, $t > 0$, and 1-Lipschitz function $f : E \to \mathbb{R}$,

$$\frac{1}{t}\log\|\mathrm{P}_t^{\lambda f}\|_{L^2(\mu)} \leq \lambda \int_E f\,\mathrm{d}\mu + \frac{C\lambda^2}{4}.$$

(3) For any $r, t > 0$, 1-Lipschitz function $f : E \to \mathbb{R}$, and $\nu \in \mathcal{P}(E)$ such that $\mathrm{d}\nu/\mathrm{d}\mu \in L^2(\mu)$,

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s \geq \int_E f\,\mathrm{d}\mu + r\right) \leq \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \exp\left(-\frac{tr^2}{C}\right),$$

where $\mathbb{P}_\nu(\cdot) := \int_E \mathbb{P}_x(\cdot)\,\nu(\mathrm{d}x)$.

In other words, Theorem 1.1.3 characterizes the $\mathcal{W}_1 I$ inequality in terms of (2) concentration inequalities for the operator norms of the Feynman-Kac semigroups and (3) deviation inequalities for the time-averages $\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s$ from the spatial averages $\int_E f\,\mathrm{d}\mu$ for Lipschitz f.

1.1.2 A new characterization of $\mathcal{W}_2 I$

Our main result, Theorem 1.1.4 below, provides a similar characterization for the $\mathcal{W}_2 I$ inequality, in which conditions (2) and (3) are replaced by dimension-free counterparts involving the product measures $\mu^{\otimes n}$. Different characterizations of the $\mathcal{W}_2 I$ inequality have been given in prior literature, which we discuss in more detail in Chapter 1.1.4, but none in terms of dimension-free properties.

We first need some notation. For any $n \in \mathbb{N}$, we define in the natural way the product Markov process $(X_t^1, \ldots, X_t^n)_{t \geq 0}$ on E^n in which each coordinate evolves according to the original process on E, conditionally independently given $(X_0^1, \ldots, X_0^n)$. We write $\mathbb{P}_x^n$ for the law of this process given $(X_0^1, \ldots, X_0^n) = x$ and $\mathbb{E}_x^n$ for expectation under $\mathbb{P}_x^n$.

The corresponding infinitesimal generator maps a suitable function f to the function $x \mapsto \sum_{i=1}^n \mathcal{L}f(\cdot, x_{-i})(x_i)$, where $f(\cdot, x_{-i}) := f(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_n)$ for $x = (x_1, \ldots, x_n) \in E^n$. That is, $\mathcal{L}$ acts on each coordinate separately, and we sum over the coordinates. The corresponding Dirichlet form is the so-called *sum-form*:

$$\mathcal{E}^{\oplus n}(g, g) = \int_{E^n} \sum_{i=1}^n \mathcal{E}(g(\cdot, x_{-i}), g(\cdot, x_{-i}))\, \mu^{\otimes n}(\mathrm{d}x). \tag{1.6}$$

The domain $\mathbb{D}(\mathcal{E}^{\oplus n})$ is the set of $g \in L^2(\mu^{\otimes n})$ for which $g(\cdot, x_{-i}) \in \mathbb{D}(\mathcal{E})$ for $\mu^{\otimes n}$-a.e. $x \in E^n$ and the integral on the right-hand side of the above equation is finite. The Fisher information $I(\nu \,|\, \mu^{\otimes n})$ for $\nu \in \mathcal{P}(E^n)$ is defined analogously to (1.3):

$$I(\nu \,|\, \mu^{\otimes n}) = \begin{cases} \mathcal{E}^{\oplus n}(\sqrt{f}, \sqrt{f}) & \text{if } \nu \ll \mu^{\otimes n},\ \frac{\mathrm{d}\nu}{\mathrm{d}\mu^{\otimes n}} = f, \text{ and } \sqrt{f} \in \mathbb{D}(\mathcal{E}^{\oplus n}) \\ +\infty & \text{otherwise.} \end{cases} \tag{1.7}$$

For $f \in B(E^n)$, let $(\mathrm{P}_{n,t}^f)_{t \geq 0}$ be the n-dimensional Feynman-Kac semigroup, given by

$$\mathrm{P}_{n,t}^f\, g(x) := \mathbb{E}_x^n \left[g(X_t^1, \ldots, X_t^n) \exp\left(\int_0^t f(X_s^1, \ldots, X_s^n)\, \mathrm{d}s \right) \right], \quad x \in E^n. \tag{1.8}$$

Its operator norm is defined in the usual way, by

$$\left\|\mathrm{P}^f_{n,t}\right\|_{L^2(\mu^{\otimes n})} \coloneqq \sup\left\{\left\|\mathrm{P}^f_{n,t}\, g\right\|_{L^2(\mu^{\otimes n})} : g \geq 0, \int_{E^n} g^2 \,\mathrm{d}\mu^{\otimes n} \leq 1\right\}. \tag{1.9}$$

This product Markov process is ubiquitous in probability because of its central role in the *tensorization* of functional inequalities. Tensorization properties of functional inequalities are critical to their usefulness in the study of concentration of measure in high dimension. The popular $\mathcal{W}_2H$, Poincaré, and log-Sobolev inequalities are all well known to be dimension-free [43, 36, 44], in the sense that if they hold for a measure $\mu \in \mathcal{P}(E)$ with some constant C then they also hold for $\mu^{\otimes n} \in \mathcal{P}(E^n)$ with the same constant C for every $n \in \mathbb{N}$. The same is true for the more recently introduced $\mathcal{W}_2I$ inequality, by [5, Corollary 2.13]. These statements, and more fundamentally the very definitions of the Poincaré, log-Sobolev, and $\mathcal{W}_2I$ inequalities, depend on a choice of Markov process (i.e., an infinitesimal generator). On $(E,\mu)^{\otimes n}$ the canonical choice is precisely this product Markov process.

Our main result, Theorem 1.1.4, identifies the precise high-dimensional concentration property encoded by the inequality $\mathcal{W}_2I$. Unless stated otherwise, the product space E^n is equipped with the ℓ^2-metric

$$((x_1,\dots,x_n),(y_1,\dots,y_n)) \mapsto \sqrt{\sum_{i=1}^n d^2(x_i,y_i)}. \tag{1.10}$$

The Wasserstein distance $\mathcal{W}_p$ on $\mathcal{P}(E^n)$ is defined relative to this metric, as is the $\mathcal{W}_pI(C)$ inequality for $\mu^{\otimes n}$.

Theorem 1.1.4. *Assume there exists $x_0 \in E$ such that $\int_E d^2(x,x_0)\,\mu(\mathrm{d}x) < \infty$. Let $C > 0$. The following are equivalent:*

(1) μ satisfies the $\mathcal{W}_2I(C)$ inequality.

(2) For each $n \in \mathbb{N}$, $\mu^{\otimes n}$ satisfies the $\mathcal{W}_1I(C)$ inequality.

(3) For each $n \in \mathbb{N}$, $\lambda \in \mathbb{R}$, $t > 0$, and 1-Lipschitz function $f : E^n \to \mathbb{R}$,

$$\frac{1}{t} \log \left\| \mathrm{P}_{n,t}^{\lambda f} \right\|_{L^2(\mu^{\otimes n})} \le \lambda \int_{E^n} f \, \mathrm{d}\mu^{\otimes n} + \frac{C\lambda^2}{4}. \tag{1.11}$$

(4) For each $n \in \mathbb{N}$, $r, t > 0$, 1-Lipschitz function $f : E^n \to \mathbb{R}$, and $\nu \in \mathcal{P}(E^n)$ such that $\mathrm{d}\nu/\mathrm{d}\mu^{\otimes n} \in L^2(\mu^{\otimes n})$, we have

$$\mathbb{P}_\nu^n \left(\frac{1}{t} \int_0^t f\left(X_s^1, \dots, X_s^n\right) \mathrm{d}s - \int_{E^n} f \, \mathrm{d}\mu^{\otimes n} \ge r \right) \le \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu^{\otimes n}} \right\|_{L^2(\mu^{\otimes n})} \exp\left(-\frac{tr^2}{C} \right), \tag{1.12}$$

where $\mathbb{P}_\nu^n(\cdot) := \int_{E^n} \mathbb{P}_x^n(\cdot)\, \nu(\mathrm{d}x)$.

The implication (1) $\Rightarrow$ (2) follows immediately from Jensen's inequality and [5, Corollary 2.13], which shows that if μ satisfies $\mathcal{W}_2 I(C)$ then so does $\mu^{\otimes n}$. The equivalence (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4) is simply Theorem 1.1.3 applied to $\mu^{\otimes n}$ for each n. Hence, our contribution is to complete the equivalence by showing that (3) $\Rightarrow$ (1).

Remark 1.1.5. It is worth pointing out that the $\mathcal{W}_1 I$ inequality itself implies concentration inequalities for the product measure $\mu^{\otimes n}$ which are similar to (1.11) and (1.12), but with a worse dependence on the dimension n in comparison with the $\mathcal{W}_2 I$ inequality. For instance, suppose μ satisfies the $\mathcal{W}_1 I(C)$ inequality. Then the tensorization argument of [5, Corollary 2.13] shows that $\mu^{\otimes n}$ satisfies the $\mathcal{W}_1 I(nC)$ inequality, with E^n equipped with the ℓ^1-metric instead of the ℓ^2-metric. On the one hand, a transport inequality $\mathcal{W}_p I$ in the ℓ^1-norm is stronger than its ℓ^2-norm counterpart with the same constant, because the ℓ^1-norm dominates the ℓ^2-norm. On the other hand, the dimension-free nature of $\mathcal{W}_2 I$ (in the ℓ^2-norm) leads to much stronger concentration properties than $\mathcal{W}_1 I$. To see this concretely, consider sample averages: The $\mathcal{W}_1 I(nC)$ inequality along with the implication (1) $\Rightarrow$ (3) from Theorem 1.1.3 imply that, for any 1-Lipschitz function $f : E \to \mathbb{R}$, $n \in \mathbb{N}$, and $\nu \in \mathcal{P}(E)$ such that $\mathrm{d}\nu/\mathrm{d}\mu \in L^2(\mu)$,

$$\mathbb{P}_\nu^n \left(\frac{1}{nt} \sum_{i=1}^n \int_0^t f(X_s^i)\, \mathrm{d}s \ge \int_E f \, \mathrm{d}\mu + r \right) \le \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu^{\otimes n}} \right\|_{L^2(\mu^{\otimes n})} \exp\left(-\frac{tr^2}{C} \right), \quad r, t > 0.$$

On the other hand, if μ satisfies the $\mathcal{W}_2 I(C)$ inequality, then the exponent on the right-hand side improves to $-ntr^2/C$.

1.1.3 A Laplace-type principle for Feynman-Kac semigroups

The proof of Theorem 1.1.4, given in Chapter 1.2, makes use of a new Laplace-type principle for operator norms of Feynman-Kac semigroups, which is interesting in its own right. In the following, let $L_n : E^n \to \mathcal{P}(E)$ denote the empirical measure map, defined by

$$L_n(x_1, \ldots, x_n) := \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}. \tag{1.13}$$

Theorem 1.1.6. *Let $t > 0$. Then, for any bounded lower semicontinuous function $F : \mathcal{P}(E) \to \mathbb{R}$,*

$$\liminf_{n\to\infty} \frac{1}{nt} \log \left\| \mathrm{P}_{n,t}^{nF\circ L_n} \right\|_{L^2(\mu^{\otimes n})} \geq \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - I(\nu \,|\, \mu) \right). \tag{1.14}$$

Suppose in addition that the sub-level sets of $I(\cdot \,|\, \mu)$ are compact. Then, for any bounded upper semicontinuous function $F : \mathcal{P}(E) \to \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{nt} \log \left\| \mathrm{P}_{n,t}^{nF\circ L_n} \right\|_{L^2(\mu^{\otimes n})} \leq \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - I(\nu \,|\, \mu) \right). \tag{1.15}$$

Only the lower bound (1.14) is needed for the proof of Theorem 1.1.4. The proof of (1.14) is based on the fact that $I(\cdot \,|\, \mu^{\otimes n})$ and $f \mapsto \frac{1}{t} \log \left\| \mathrm{P}_{n,t}^{f} \right\|$ are convex conjugates (see Lemma 1.2.2 below), along with a *chain rule* formula relating $I(\cdot \,|\, \mu^{\otimes n})$ with $I(\cdot \,|\, \mu)$ (see Lemma 1.2.3 below). The main idea of our proof of (3) $\Rightarrow$ (1) in Theorem 1.1.4 is to apply (1.14) with $F = \mathcal{W}_2(\cdot, \mu) \wedge M$, for $M > 0$ which we later send to infinity. Essentially, the inequality (1.14) plays the same role for us that the lower bound of Sanov's theorem plays in the proof of Gozlan's theorem (recalled in Theorem 1.1.10 below).

The matching upper bound (1.15) is of independent interest but is not needed for the proof of Theorem 1.1.4. We derive it in Chapter 1.3 from a general Sanov-type theorem involving a tensorization $\alpha_n : \mathcal{P}(E^n) \to (-\infty, \infty]$ of an abstract functional $\alpha : \mathcal{P}(E) \to (-\infty, \infty]$ of a probability measure, inspired by recent work of the first author [45]. In this framework, we show in Chapter 1.4 that $\mathcal{W}_2^2(\mu, \cdot) \le \alpha$ if and only if $\mathcal{W}_1^2(\mu^{\otimes n}, \cdot) \le \alpha_n$ for every $n \in \mathbb{N}$. Combined with a dual form of the latter inequality, this generalizes the implications (1) ⇔ (2) ⇔ (3) in both Theorem 1.1.4 and Gozlan's Theorem 1.1.10 below. See Chapters 1.3 and 1.4 for full details.

Remark 1.1.7. Theorem 1.1.6 is very different from the usual large deviation principle for the occupation measure of the Markov process (see [46, 47, 40]), despite sharing the same “rate function" $I(\cdot \,|\, \mu)$. The usual large deviation principle, combined with Varadhan's lemma, takes the form

$$\lim_{T \to \infty} \frac{1}{T} \log \mathbb{E}_x \left[\exp\left(TF\left(\frac{1}{T} \int_0^T \delta_{X_t} \, \mathrm{d}t \right) \right) \right] = \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - I(\nu \,|\, \mu) \right),$$

for bounded continuous $F : \mathcal{P}(E) \to \mathbb{R}$. It is not clear if there is a deeper connection between Theorem 1.1.6 and this large deviation principle, but we will make no use of the latter.

Remark 1.1.8. Note that the upper bound of Theorem 1.1.6 requires the additional assumption of compactness of the sub-level sets of $I(\cdot \,|\, \mu)$. They are always closed, because $I(\cdot \,|\, \mu)$ is well known to be lower semicontinuous. Hence, the additional assumption holds automatically if E is compact. In the non-compact case, there are tractable sufficient conditions, such the hypotheses of [47, Lemma 7.1], or the uniform integrability of the semigroup as in [48].

1.1.4 Related literature and $\mathcal{W}_p H$ inequalities

Transportation-information inequalities were introduced in the papers [5, 6], which developed several necessary and sufficient conditions as well as connections with other functional inequalities. The most satisfying results are in the context of Example 1.1.2: $\mathcal{W}_2 I$ is weaker than a log-Sobolev inequality but stronger than a Poincaré inequality [5, Proposition 2.9]. In addition, $\mathcal{W}_p I$ implies the corresponding transportation-entropy inequality $\mathcal{W}_p H$ (defined below), for $p = 1, 2$ [6, Theorems 2.1 and 2.4]. More recently, $\mathcal{W}_2 I$ was characterized in terms of a Lyapunov condition [10, Theorem 1.3], though again only in the context of Example 1.1.2, and without explicit constants. In full generality, a characterization of the $\mathcal{W}_2 I$ inequality in terms of inf-convolution inequalities was given in [5, Corollary 2.5].

Our main result, Theorem 1.1.4, is best understood in comparison with Gozlan's characterization of Talagrand's inequality in terms of dimension-free concentration [12]. To explain this, we first recall the basics of transportation-entropy inequalities, referring to the survey [36] for a more comprehensive overview. The relative entropy between probability measures ν and μ is defined as usual by

$$H(\nu \,|\, \mu) := \begin{cases} \int_E \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} \, d\mu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.16}$$

For $C > 0$ and $p \geq 1$, we say that μ *satisfies the $\mathcal{W}_p H(C)$ inequality* if

$$\mathcal{W}_p^2(\mu, \nu) \leq C H(\nu \,|\, \mu), \quad \text{for all } \nu \in \mathcal{P}(E). \tag{1.17}$$

Inequalities of this form gained prominence from the work of Marton [1] and Talagrand [2], with a number of subsequent contributions further clarifying their precise role in characterizing concentration properties. We first mention a famous dual characterization due to Bobkov and Götze:

Theorem 1.1.9 ([3, Theorem 1.3])**.** *Let $C > 0$. The following are equivalent:*

(1) μ satisfies the $\mathcal{W}_1 H(C)$ inequality.

(2) For every 1-Lipschitz function $f : E \to \mathbb{R}$,

$$\log \int_E e^{\lambda f} d\mu \leq \lambda \int_E f d\mu + \frac{C\lambda^2}{4}, \quad \textit{for all } \lambda \in \mathbb{R}.$$

This is the $\mathcal{W}_1 H$ analogue of the $\mathcal{W}_1 I$ characterization (1) $\Leftrightarrow$ (2) stated in Theorem 1.1.3. There are several other equivalent formulations possible in Theorem 1.1.9, at least if one is willing to change the constant (by a universal factor). Notably, if μ satisfies the $\mathcal{W}_1 H(C)$ inequality, then

(3) For every 1*-Lipschitz function* $f : E \to \mathbb{R}$*,*

$$\mu\left(f - \int_E f \, \mathrm{d}\mu > r\right) \le e^{-r^2/C}, \quad \textit{for all } r > 0.$$

Conversely, if μ satisfies (3), then it satisfies the $\mathcal{W}_1 H(C')$ inequality for some C'; see [49, Theorem 2.3] for details on how to bound C'. This is analogous to (3) in Theorem 1.1.3. The $\mathcal{W}_1 H$ inequality thus encodes concentration properties of the underlying measure, whereas the $\mathcal{W}_1 I$ inequality encodes concentration properties for time-averages of a Markov process around its equilibrium.

Turning now to the quadratic inequality, it has been known since the work of Marton [1] and Talagrand [2] that $\mathcal{W}_2 H$ tensorizes: If μ satisfies $\mathcal{W}_2 H(C)$ then so does the product measure $\mu^{\otimes n}$ for any n. Since $\mathcal{W}_2 H(C)$ implies $\mathcal{W}_1 H(C)$, this yields any of the above expressions of concentration for $\mu^{\otimes n}$, with a dimension-free constant C. Gozlan proved a remarkable converse to this statement in [12, Theorem 1.3], though below we quote a somewhat different formulation. Recall that we equip E^n with the ℓ^2-metric defined in (1.10).

Theorem 1.1.10 ([44, Theorem 9.6.4], [50, Theorem 4.31])**.** *Let* $C > 0$*. The following are equivalent:*

(1) μ *satisfies the* $\mathcal{W}_2 H(C)$ *inequality.*

(2) For each $n \in \mathbb{N}$*,* $\mu^{\otimes n}$ *satisfies the* $\mathcal{W}_1 H(C)$ *inequality.*

(3) For each $n \in \mathbb{N}$ *and 1-Lipschitz function* $f : E^n \to \mathbb{R}$*,*

$$\log \int_{E^n} e^{\lambda f} \mathrm{d}\mu^{\otimes n} \le \lambda \int_{E^n} f \, \mathrm{d}\mu^{\otimes n} + \frac{C\lambda^2}{4}, \quad \textit{for all } \lambda \in \mathbb{R}.$$

(4) There exists $K > 0$ *such that, for every* $n \in \mathbb{N}$ *and 1-Lipschitz function* $f : E^n \to \mathbb{R}$*,*

$$\mu^{\otimes n}\left(f - \int_{E^n} f \, \mathrm{d}\mu^{\otimes n} > r\right) \le K \exp\left(-\frac{r^2}{C}\right), \quad \textit{for all } r > 0.$$

The parallels with our Theorem 1.1.4 should be clear. The implications (1) $\Rightarrow$ (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4) in Theorem 1.1.10 were known (up to a universal change in the constant for (4)), with Gozlan's result completing the equivalence. Our Theorem 1.1.4 fills a gap in the literature by completing this analogy between $\mathcal{W}_2H$ and $\mathcal{W}_2I$.

Remark 1.1.11. Gozlan's proof of Theorem 1.1.10 in [12] relies on Sanov's theorem, but alternative proofs have since been found based on a characterization of the $\mathcal{W}_2H$ inequality in terms of inf-convolution inequalities; see [51, Theorem 5.1], [38, Proposition 3.4], or [44, Theorem 9.6.4]. There is an analogous inf-convolution inequality given in [5, Corollary 2.5] which characterizes the $\mathcal{W}_2I$ inequality. This raises the natural question, which we have not resolved, of whether an alternative proof of Theorem 1.1.4 is possible based on this inf-convolution inequality.

1.1.5 Organization of the chapter

The rest of the chapter is organized as follows. In Chapter 1.2 we prove Theorem 1.1.4 after first proving the Laplace-type lower bound of Theorem 1.1.6. In Chapter 1.3 we develop an abstract Sanov-type theorem and use it to prove the upper bound in Theorem 1.1.6. Finally, in the abstract setting, Chapter 1.4 generalizes (some of) the characterizations of $\mathcal{W}_2I$ and $\mathcal{W}_2H$ given in Theorems 1.1.4 and 1.1.10.

1.2 The characterization of $\mathcal{W}_2I$

This section is devoted to the proof of Theorem 1.1.4. We first collect a few well-known properties of the Fisher information and Feynman-Kac semigroups. Recall the definitions of $\mathcal{E}^{\oplus n}$ and $I(\cdot \mid \mu^{\otimes n})$ from (1.6) and (1.7).

Lemma 1.2.1. *For every $t > 0$, $n \in \mathbb{N}$, and $f \in B(E^n)$, the Feynman-Kac semigroup defined in* (1.8) *satisfies*

$$\frac{1}{t} \log \left\| \mathrm{P}^f_{n,t} \right\|_{L^2(\mu^{\otimes n})} = \sup \left\{ \int_{E^n} f g^2 \, \mathrm{d}\mu^{\otimes n} - \mathcal{E}^{\oplus n}(g,g) : g \in \mathbb{D}(\mathcal{E}^{\oplus n}), \int_{E^n} g^2 \mathrm{d}\mu^{\otimes n} = 1 \right\}.$$

Proof. This result can be found in [5, Lemma 6.1] and seems to be folklore. We include a proof here for the sake of completeness. The inequality ($\leq$) was proved in [40, Proof of Theorem 1, Case 1] by applying the Lumer-Philips theorem.

For the opposite inequality, note that since f is bounded, the generator $\mathcal{L}_f \coloneqq \mathcal{L} + f$ is self-adjoint and its domain coincides with $\mathbb{D}(\mathcal{L})$. Denote by $\mathcal{B}(\mathbb{R})$ the Borel sets of $\mathbb{R}$ and by $\mathscr{B}(L^2(\mu))$ the set of bounded linear operators on $L^2(\mu)$. By the spectral theorem [52, Theorem 13.30], there is a unique resolution of the identity (a.k.a. spectral family) $\mathbf{E} : \mathcal{B}(\mathbb{R}) \to \mathscr{B}(L^2(\mu))$, such that

$$\int_E h\mathcal{L}_f g \,\mathrm{d}\mu = \int_{\mathbb{R}} \lambda \,\mathbf{E}_{g,h}(\mathrm{d}\lambda), \qquad \text{for all } g \in \mathbb{D}(\mathcal{L}), h \in L^2(\mu).$$

Here, for any $f, g \in L^2(\mu)$, $\mathbf{E}_{g,h}$ is a finite Borel measure on $\mathbb{R}$, defined by

$$\mathcal{B}(\mathbb{R}) \ni A \mapsto \mathbf{E}_{g,h}(A) \coloneqq \int_E h\,\mathbf{E}(A) g \,\mathrm{d}\mu \in \mathbb{R}.$$

Moreover, $\mathbf{E}$ is concentrated on the spectrum of $\mathcal{L}_f$, which we denote by $\sigma(\mathcal{L}_f) \subset \mathbb{R}$. In other words, $\mathbf{E}(\sigma(\mathcal{L}_f)) = I$, the identity operator on $L^2(\mu)$. In particular,

$$\int_E g\mathcal{L}_f g \,\mathrm{d}\mu = \int_{\sigma(\mathcal{L}_f)} \lambda \,\mathbf{E}_{g,g}(\mathrm{d}\lambda) = \int_{\sigma(\mathcal{L}_f)} \lambda \left\|\mathbf{E}(\mathrm{d}\lambda) g\right\|^2_{L^2(\mu)}, \qquad \text{for all } g \in \mathbb{D}(\mathcal{L}), \int_E g^2 \,\mathrm{d}\mu = 1, \tag{1.18}$$

where the last equality follows from the fact that each $\mathbf{E}(\mathrm{d}\lambda)$ is a self-adjoint projection on $L^2(\mu)$; see [52, Theorm 12.14 and Equation 1 on pp. 317]. Note that in the above integral, $\left\|\mathbf{E}(\mathrm{d}\lambda) g\right\|^2_{L^2(\mu)}$ is a probability measure on $\sigma(\mathcal{L}_f)$, since it is a positive measure with total mass $\left\|\mathbf{E}(\sigma(\mathcal{L}_f)) g\right\|^2_{L^2(\mu)} = \|g\|^2_{L^2(\mu)} = 1$.

On the other hand, the symbolic calculus (a.k.a. functional calculus) [52, Theorem 12.21] and similar reasoning as before implies that

$$\left\|\mathrm{P}^f_t g\right\|^2_{L^2(\mu)} = \int_{\sigma(\mathcal{L}_f)} e^{2t\lambda} \left\|\mathbf{E}(\mathrm{d}\lambda) g\right\|^2_{L^2(\mu)}, \qquad \text{for all } g \in L^2(\mu), \int_E g^2 \mathrm{d}\mu = 1, \tag{1.19}$$

where we make use of the fact that the spectrum $\sigma(\mathcal{L}_f) \subset \mathbb{R}$ is bounded from above; see [52, Theorem 13.38]. Hence, the integrand $\lambda \mapsto e^{2t\lambda}$ is a bounded function on the domain $\sigma(\mathcal{L}_f)$. By Jensen's inequality, it follows from (1.18) and (1.19) that

$$\left\|\mathrm{P}_t^f g\right\|_{L^2(\mu)} \geq \exp\left(t \int_E g\mathcal{L}_f g \,\mathrm{d}\mu\right), \quad \text{for all } g \in \mathbb{D}(\mathcal{L}), \int_E g^2 \,\mathrm{d}\mu = 1. \tag{1.20}$$

Recalling the definition of the operator norm in (1.5), it follows that

$$\left\|\mathrm{P}_t^f\right\|_{L^2(\mu)} \geq \exp\left(t \int_E g\mathcal{L}_f g \,\mathrm{d}\mu\right), \quad \text{for all } g \in \mathbb{D}(\mathcal{L}), \int_E g^2 \mathrm{d}\mu = 1. \tag{1.21}$$

Taking the supremum over $g \in \mathbb{D}(\mathcal{L})$, recalling the definition of $\mathcal{L}_f$ above and also the definition of $\mathcal{E}$ in (1.2),

$$\left\|\mathrm{P}_t^f\right\|_{L^2(\mu)} \geq \exp\left(t \sup\left\{\int_E g\mathcal{L}_f g : g \in \mathbb{D}_2(\mathcal{L}), \int_E g^2 \,\mathrm{d}\mu = 1\right\}\right) \tag{1.22}$$

$$= \exp\left(t \sup\left\{\int_E f g^2 \mathrm{d}\mu - \mathcal{E}(g,g) : g \in \mathbb{D}(\mathcal{L}), \int_E g^2 \,\mathrm{d}\mu = 1\right\}\right) \tag{1.23}$$

$$= \exp\left(t \sup\left\{\int_E f g^2 \mathrm{d}\mu - \mathcal{E}(g,g) : g \in \mathbb{D}(\mathcal{E}), \int_E g^2 \,\mathrm{d}\mu = 1\right\}\right), \tag{1.24}$$

where the last equality follows from the fact that $\mathbb{D}(\mathcal{E})$ is the completion of $\mathbb{D}(\mathcal{L})$, with respect to the Dirichlet norm given by $\|g\|_{\mathcal{E}} := \sqrt{\mathcal{E}(g,g) + \|g\|^2_{L^2(\mu)}}$; see [42, Theorem 3.1.1]. This concludes the proof of the inequality ($\geq$). $\square$

Lemma 1.2.2. *For every $t > 0$, $n \in \mathbb{N}$, and $f \in B(E^n)$, the following variational formula holds:*

$$\frac{1}{t} \log \left\|\mathrm{P}_{n,t}^f\right\|_{L^2(\mu^{\otimes n})} = \sup_{\nu \in \mathcal{P}(E^n)} \left(\int_{E^n} f \,\mathrm{d}\nu - I(\nu \,|\, \mu^{\otimes n})\right). \tag{1.25}$$

Proof. This is a straightforward consequence of Lemma 1.2.1, similar to the argument in [5, Proof of Theorem 2.4]. The contraction property of the Dirichlet form [42, Theorem 1.4.1] ensures that $\mathcal{E}^{\oplus n}(|g|, |g|) \leq \mathcal{E}^{\oplus n}(g,g)$ for any $g \in \mathbb{D}(\mathcal{E}^{\oplus n})$. This allows us to restrict the supremum in the formula of Lemma 1.2.1 to nonnegative functions g, which we may then identify with a probability measure via $g = \sqrt{\mathrm{d}\nu/\mathrm{d}\mu^{\otimes n}}$. That is, for $f \in B(E^n)$,

$$
\begin{aligned}
\frac{1}{t}\log\left\|\mathrm{P}^f_{n,t}\right\|_{L^2(\mu^{\otimes n})} &= \sup\left\{\int_{E^n} fg^2\,\mathrm{d}\mu^{\otimes n} - \mathcal{E}^{\oplus n}(g,g) : g \in \mathbb{D}(\mathcal{E}^{\oplus n}), g \geq 0, \int_{E^n} g^2\,\mathrm{d}\mu^{\otimes n} = 1\right\} \\
&= \sup_{\nu\in\mathcal{P}(E^n)}\left(\int_{E^n} f\,\mathrm{d}\nu - I(\nu\,|\,\mu^{\otimes n})\right).
\end{aligned}
$$

□

Our third lemma is an important *chain rule* for the Fisher information $I(\cdot\,|\,\mu^{\otimes n})$, borrowed from [5]. For integers $n \geq k \geq 1$ and $x = (x_1, \dots, x_n) \in E^n$, we denote by $x_{-k} \coloneqq (x_i)_{i\in\{1,\dots,n\}\setminus k}$ the vector consisting of all but the k^{th} coordinate. Let $\pi_{-k} : E^n \to E^{n-1}$ be the natural projection, i.e., $\pi_{-k}(x) = x_{-k}$. For $\nu \in \mathcal{P}(E^n)$, we define the measurable map $\nu_{-k} : E^{n-1} \to \mathcal{P}(E)$ via disintegration

$$
\nu(\mathrm{d}x_1, \dots, \mathrm{d}x_n) = \nu_{-k}(x_{-k})(\mathrm{d}x_k)\,\nu \circ \pi^{-1}_{-k}(\mathrm{d}x_{-k}). \tag{1.26}
$$

In probabilistic terms, if $X = (X_1, \dots, X_n)$ has joint law ν, then $\nu_{-k}(X_{-k})$ is a version of the conditional law of X_k given X_{-k}. Note that ν_{-k} is uniquely determined up to ν-a.s. equality.

Lemma 1.2.3 ([5, Lemma 2.12])**.** *For each $n \in \mathbb{N}$ and $\nu \in \mathcal{P}(E^n)$, it holds that*

$$
I(\nu\,|\,\mu^{\otimes n}) = \int_{E^n}\sum_{k=1}^{n} I(\nu_{-k}(x_{-k})\,|\,\mu)\,\nu(\mathrm{d}x).
$$

We are now ready to give the proof of the lower bound (1.14) of Theorem 1.1.6. Again, only the lower bound is needed for the proof of Theorem 1.1.4, given just below. The upper bound (1.15) of Theorem 1.1.6 requires some additional assumptions and machinery and is less self-contained, so we defer its proof to the very end of Chapter 1.3.

Proof of the lower bound (1.14) **of Theorem 1.1.6.** According to Lemma 1.2.2, we have

$$
\frac{1}{nt}\log\left\|\mathrm{P}^{nF\circ L_n}_{n,t}\right\|_{L^2(\mu^{\otimes n})} = \sup_{\nu\in\mathcal{P}(E^n)}\left(\int_{E^n} F\circ L_n\,\mathrm{d}\nu - \frac{1}{n}I(\nu\,|\,\mu^{\otimes n})\right).
$$

Choose $\nu \in \mathcal{P}(E)$ arbitrarily. From the formula of Lemma 1.2.3 it follows immediately that $I(\cdot \,|\, \mu^{\otimes n})$ simplifies for product measures, in the sense that $I(\nu^{\otimes n} \,|\, \mu^{\otimes n}) = nI(\nu \,|\, \mu)$. Hence,

$$\frac{1}{nt} \log \left\| \mathrm{P}_{n,t}^{nF \circ L_n} \right\|_{L^2(\mu^{\otimes n})} \geq \int_{E^n} F \circ L_n \, \mathrm{d}\nu^{\otimes n} - I(\nu \,|\, \mu).$$

By the law of large numbers for empirical measures [53, Theorem 11.4.1], we have $\nu^{\otimes n} \circ L_n^{-1} \to \delta_\nu$ in $\mathcal{P}(\mathcal{P}(E))$, in the sense that $\int_{E^n} \phi \circ L_n \, \mathrm{d}\nu^{\otimes n} \to \phi(\nu)$ for any bounded and weakly continuous function $\phi : \mathcal{P}(E) \to \mathbb{R}$. Together with the lower semicontinuity and boundedness of F and a version of the Portmanteau theorem [54, Theorem A.3.12], this yields

$$\liminf_{n \to \infty} \frac{1}{nt} \log \left\| \mathrm{P}_{n,t}^{nF \circ L_n} \right\|_{L^2(\mu^{\otimes n})} \geq F(\nu) - I(\nu \,|\, \mu).$$

The lower bound (1.14) follows by taking the supremum over $\nu \in \mathcal{P}(E)$. □

Proof of Theorem 1.1.4.

- (1) ⇒ (2): Let $n \in \mathbb{N}$. Since μ satisfies $\mathcal{W}_2 I(C)$, the tensorization property of transportation-information inequalities [5, Corollary 2.13] ensures that the product measure $\mu^{\otimes n}$ also satisfies $\mathcal{W}_2 I(C)$. Hence, by Jensen's inequality, $\mu^{\otimes n}$ satisfies $\mathcal{W}_1 I(C)$.
- (2) ⇔ (3) ⇔ (4): This follows by applying Theorem 1.1.3 to $(E^n, \mu^{\otimes n})$ for each n.
- (3) ⇒ (1): Let $M > 0$. Define $F : \mathcal{P}(E) \to \mathbb{R}$ by

  $$F(\nu) := \mathcal{W}_2(\mu, \nu) \wedge M.$$

  For each $n \in \mathbb{N}$ and $x, y \in E^n$, note that

  $$\mathcal{W}_2^2(L_n(x), L_n(y)) = \mathcal{W}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{i=1}^n \delta_{y_i}\right) \leq \frac{1}{n}\sum_{i=1}^n d^2(x_i, y_i),$$

  where the last inequality follows by bounding the infimum in the definition (1.1) of the

Wasserstein distance using the trivial coupling $\frac{1}{n}\sum_{i=1}^{n}\delta_{(x_i,y_i)}$ of $L_n(x)$ and $L_n(y)$. Recalling that E^n is equipped with the ℓ^2-metric, this implies by the triangle inequality that $\sqrt{n}F\circ L_n$ is 1-Lipschitz on E^n. Therefore, by (1.11), for all $\lambda \geq 0$,

$$\frac{1}{nt}\log\left\|\mathrm{P}_{n,t}^{\lambda n F\circ L_n}\right\|_{L^2(\mu^{\otimes n})} \leq \lambda\int_{E^n} F\circ L_n \,\mathrm{d}\mu^{\otimes n} + \frac{C\lambda^2}{4}. \tag{1.27}$$

Since μ has finite second moment by assumption, the law of large numbers in Wasserstein distance implies

$$\lim_{n\to\infty}\int_{E^n} F\circ L_n\,\mathrm{d}\mu^{\otimes n} = 0. \tag{1.28}$$

Indeed, if $(X_i)_{i\in\mathbb{N}}$ are i.i.d. E-valued random variables with law μ, then $L_n(X_1,\dots,X_n)\to\mu$ weakly a.s. by the law of large numbers [53, Theorem 11.4.1], and $\frac{1}{n}\sum_{i=1}^{n}d^2(X_i,x_0)\to\int_E d^2(x,x_0)\,\mu(\mathrm{d}x)$ for any $x_0\in E$ by the law of large numbers and square-integrability of μ. Use [55, Theorem 7.12, (iii) $\Rightarrow$ (i)] to deduce $\mathcal{W}_2(L_n(X_1,\dots,X_n),\mu)\to 0$ a.s., and then the bounded convergence theorem yields (1.28). Note also that $\mathcal{W}_2(\mu,\cdot)$ is lower semicontinuous (which follows from Kantorovich duality [55, Theorem 1.3], for instance). Since F is thus lower semicontinuous and also bounded, we may apply the lower bound of Theorem 1.1.6, followed by (1.27) and (1.28), to get

$$\sup_{\nu\in\mathcal{P}(E)}\Big(\lambda\mathcal{W}_2(\mu,\nu)\wedge M - I(\nu\,|\,\mu)\Big) \leq \liminf_{n\to\infty}\frac{1}{nt}\log\left\|\mathrm{P}_{n,t}^{\lambda n F\circ L_n}\right\|_{L^2(\mu^{\otimes n})} \leq \frac{C\lambda^2}{4},$$

for $\lambda\geq 0$. Consequently, for all $\nu\in\mathcal{P}(E)$ and $\lambda\geq 0$,

$$\lambda\mathcal{W}_2(\mu,\nu)\wedge M - \frac{C\lambda^2}{4} \leq I(\nu\,|\,\mu).$$

Since $M>0$ was arbitrary, letting $M\to\infty$ gives

$$\lambda\mathcal{W}_2(\mu,\nu) - \frac{C\lambda^2}{4} \leq I(\nu\,|\,\mu).$$

for all $\nu \in \mathcal{P}(E)$ and $\lambda \geq 0$. Optimize over $\lambda \geq 0$ to get $C^{-1}\mathcal{W}_2^2(\mu, \nu) \leq I(\nu \,|\, \mu)$ for all $\nu \in \mathcal{P}(E)$, so that μ satisfies $\mathcal{W}_2 I(C)$. $\square$

1.3 A limit theorem of Sanov type

In this section, we prove an abstract version of Theorem 1.1.6, inspired by recent work of the first author [45]. Fix throughout this section a measurable functional $\alpha : \mathcal{P}(E) \to (-\infty, \infty]$ which is bounded from below and not identically $+\infty$. At the end of this section, we will specialize to $\alpha = I(\cdot \,|\, \mu)$ in order to prove the upper bound of Theorem 1.1.6.

To define a tensorized functional $\alpha_n : \mathcal{P}(E^n) \to (-\infty, \infty]$ for each $n \in \mathbb{N}$, recall the notation for the conditional measures ν_{-k} for $\nu \in \mathcal{P}(E^n)$, defined in (1.26). Define

$$\alpha_n(\nu) := \int_{E^n} \sum_{k=1}^{n} \alpha\left(\nu_{-k}\left(x_{-k}\right)\right) \, \nu\left(\mathrm{d}x_1, \ldots, \mathrm{d}x_n\right), \qquad \nu \in \mathcal{P}(E^n). \tag{1.29}$$

Note that α_n is well defined and bounded from below because α was assumed to be measurable and bounded from below. We define the convex conjugate $\rho_n : B(E^n) \to \mathbb{R}$ by

$$\rho_n(f) := \sup_{\nu \in \mathcal{P}(E^n)} \left(\int_{E^n} f \, \mathrm{d}\nu - \alpha_n\left(\nu\right) \right), \qquad f \in B(E^n). \tag{1.30}$$

Note that ρ_n is indeed real-valued because α_n is bounded from below and not identically $+\infty$.

In the case $\alpha = I(\cdot \,|\, \mu)$, it holds that $\alpha_n = I(\cdot \,|\, \mu^{\otimes n})$ by Lemma 1.2.3, where the tensorized Fisher information was defined in Chapter 1.1.2. Moreover, in this case, $\rho_n(f) = \frac{1}{t} \log \|\mathrm{P}^f_{n,t}\|_{L^2(\mu^{\otimes n})}$ for any $t > 0$, by Lemma 1.2.2.

Recall in the following the definition of the empirical measure map from (1.13).

Theorem 1.3.1. *For any bounded lower semicontinuous function $F : \mathcal{P}(E) \to \mathbb{R}$,*

$$\liminf_{n \to \infty} \frac{1}{n} \rho_n\left(nF \circ L_n\right) \geq \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - \alpha(\nu)\right). \tag{1.31}$$

Suppose in addition that α is convex and has compact sub-level sets. Then, for any bounded upper *semicontinuous function* $F : \mathcal{P}(E) \to \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n}\rho_n\left(nF\circ L_n\right) \le \sup_{\nu\in\mathcal{P}(E)} \left(F(\nu) - \alpha(\nu)\right). \tag{1.32}$$

Proof of the lower bound (1.31)**.** This is essentially identical to the proof of the lower bound in Theorem 1.1.6. Let $\nu \in \mathcal{P}(E)$, and note for product measures that we have the simplification $\alpha_n(\nu^{\otimes n}) = n\alpha\left(\nu\right)$. Bound the supremum in the definition of ρ_n from below using the measure $\nu^{\otimes n}$ to get

$$\frac{1}{n}\rho_n\left(nF\circ L_n\right) \ge \int_{E^n} F\circ L_n \,\mathrm{d}\nu^{\otimes n} - \frac{1}{n}\alpha_n\left(\nu^{\otimes n}\right) = \int_{E^n} F\circ L_n \,\mathrm{d}\nu^{\otimes n} - \alpha(\nu).$$

Use the law of large numbers along with lower semicontinuity and boundedness of F to get

$$\liminf_{n\to\infty} \frac{1}{n}\rho_n\left(nF\circ L_n\right) \ge F\left(\nu\right) - \alpha\left(\nu\right).$$

The lower bound (1.31) now follows by taking the supremum over $\nu \in \mathcal{P}(E)$. $\square$

To prove the upper bound, we next develop an alternative tensorization $\widehat{\alpha}_n$ which, unlike α_n, takes into account an order of the coordinates. For $n \in \mathbb{N}$ and $\nu \in \mathcal{P}(E^n)$, we define $\nu_{0,1} \in \mathcal{P}(E)$ and the measurable maps $\nu_{k-1,k} : E^{k-1} \to \mathcal{P}(E)$ for $k = 2, \ldots, n$ via the disintegration

$$\nu(\mathrm{d}x_1, \ldots, \mathrm{d}x_n) = \nu_{0,1}(\mathrm{d}x_1)\prod_{k=2}^{n} \nu_{k-1,k}(x_1, \ldots, x_{k-1})(\mathrm{d}x_k).$$

In other words, if $X = (X_1, \ldots, X_n)$ has joint law ν, then $\nu_{0,1}$ is the marginal law of X_1, and $\nu_{k-1,k}(X_1, \ldots, X_{k-1})$ is a version of the conditional law of X_k given $(X_1, \ldots, X_{k-1})$. Next, define

$\widehat{\alpha}_n : \mathcal{P}(E^n) \to (-\infty, \infty]$ and its conjugate $\widehat{\rho}_n : B(E^n) \to \mathbb{R}$ by

$$\widehat{\alpha}_n(\nu) := \int_{E^n} \sum_{k=1}^{n} \alpha\left(\nu_{k-1,k}\left(x_1, \ldots, x_{k-1}\right)\right) \nu\left(\mathrm{d}x_1, \ldots, \mathrm{d}x_n\right). \tag{1.33}$$

$$\widehat{\rho}_n(f) := \sup_{\nu \in \mathcal{P}(E^n)} \left(\int_{E^n} f \, \mathrm{d}\nu - \widehat{\alpha}_n(\nu) \right). \tag{1.34}$$

The analogue of Theorem 1.3.1 for this form of tensorization is known:[2]

Theorem 1.3.2. *[45, Theorem 1.1] For any bounded lower semicontinuous function $F : \mathcal{P}(E) \to \mathbb{R}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \widehat{\rho}_n\left(nF \circ L_n\right) \geq \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - \alpha(\nu)\right).$$

Suppose in addition that α is convex and has compact sub-level sets. Then, for any bounded upper semicontinuous function $F : \mathcal{P}(E) \to \mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \widehat{\rho}_n\left(nF \circ L_n\right) \leq \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - \alpha(\nu)\right).$$

Remark 1.3.3. As is explained in [45], Theorem 1.3.2 can be seen as a generalization of Sanov's theorem. Indeed, if $\alpha = H(\cdot \,|\, \mu)$, then the chain rule for relative entropy [54, Theorem B.2.1] yields $\widehat{\alpha}_n = H(\cdot \,|\, \mu^{\otimes n})$, and the Gibbs variational formula [54, Proposition 1.4.2] yields $\widehat{\rho}_n(f) = \log \int_{E^n} e^f \, \mathrm{d}\mu^{\otimes n}$. For any bounded continuous $F : \mathcal{P}(E) \to \mathbb{R}$, Theorem 1.3.2 then states that

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{E^n} e^{nF \circ L_n} \, \mathrm{d}\mu^{\otimes n} = \sup_{\nu \in \mathcal{P}(E)} \left(F(\nu) - H(\nu \,|\, \mu)\right).$$

This is precisely Sanov's theorem, in Laplace principle form [54, Theorem 2.2.1]. This explains why we describe Theorems 1.3.1 and 1.1.6 also as Sanov-type theorems.

Note that Theorems 1.3.1 and 1.3.2 both give identical upper and lower bounds, despite dealing with different tensorizations α_n and $\widehat{\alpha}_n$. These two tensorizations reflect the two different kinds of

[2] Strictly speaking, [45, Theorem 1.1] assumes convexity of α and compactness of its sub-level sets, but these assumptions are not needed for the easy proof of the lower bound, which is identical to that of Theorem 1.3.1.

“chain rules" satisfied by Fisher information and relative entropy, respectively. They are related by the following:

Lemma 1.3.4. *Assume α is convex and lower semicontinuous. For every $n \in \mathbb{N}$, we have $\widehat{\alpha}_n(\nu) \le \alpha_n(\nu)$ for all $\nu \in \mathcal{P}(E^n)$, and $\widehat{\rho}_n(f) \ge \rho_n(f)$ for all $f \in B(E^n)$.*

Proof. The second claim clearly follows from the first. For the first claim, let $\nu \in \mathcal{P}(E^n)$, and let $X = (X_1, \ldots, X_n)$ have law ν. The claim follows from Jensen's inequality after noting that $\nu_{k-1,k}(X_1, \ldots, X_{k-1}) = \mathbb{E}[\nu_{-k}(X_{-k}) \,|\, X_1, \ldots, X_{k-1}]$. That is, for $f \in B(E)$,

$$\begin{aligned}\int_E f \, \mathrm{d}\nu_{k-1,k}(X_1, \ldots, X_{k-1}) &= \mathbb{E}[f(X_k) \,|\, X_1, \ldots, X_{k-1}] = \mathbb{E}\big[\mathbb{E}[f(X_k) \,|\, X_{-k}] \,\big|\, X_1, \ldots, X_{k-1}\big] \\ &= \mathbb{E}\left[\int_E f \, \mathrm{d}\nu_{-k}(X_{-k}) \,\middle|\, X_1, \ldots, X_{k-1}\right], \quad \text{a.s.}\end{aligned}$$

Convexity and lower semicontinuity of α imply, by a form of Jensen's inequality [45, Proposition B.2],

$$\alpha(\nu_{k-1,k}(X_1, \ldots, X_{k-1}) \le \mathbb{E}[\, \alpha(\nu_{-k}(X_{-k})) \,|\, X_1, \ldots, X_{k-1}], \quad \text{a.s.}$$

Thus

$$\widehat{\alpha}_n(\nu) = \mathbb{E}\left[\sum_{k=1}^n \alpha(\nu_{k-1,k}(X_1, \ldots, X_{k-1})\right] \le \mathbb{E}\left[\sum_{k=1}^n \alpha(\nu_{-k}(X_{-k}))\right] = \alpha_n(\nu).$$

□

Proof of the upper bound (1.32) **of Theorem 1.3.1.** This now follows easily by applying Lemma 1.3.4 along with the upper bound of Theorem 1.3.2:

$$\limsup_{n\to\infty} \frac{1}{n} \rho_n \left(nF \circ L_n\right) \le \limsup_{n\to\infty} \frac{1}{n} \widehat{\rho}_n \left(nF \circ L_n\right) \le \sup_{\nu \in \mathcal{P}(E)} \left(F\left(\nu\right) - \alpha\left(\nu\right)\right).$$

□

Proof of the upper bound (1.15) **of Theorem 1.1.6.** We apply Theorem 1.3.1 to $\alpha = I(\cdot \,|\, \mu)$. The tensorized form is then $\alpha_n = I(\cdot \,|\, \mu^{\otimes n})$, as is easily seen by comparing the definition (1.29) with the formula from Lemma 1.2.3. By Lemma 1.2.2, the convex conjugate defined by (1.30) takes the form $\rho_n(f) = \frac{1}{t} \log \left\| \mathrm{P}^f_{n,t} \right\|_{L^2(\mu^{\otimes n})}$, which we note does not actually depend on t. The claimed upper bound is then immediate from Theorem 1.3.1, once we note that $I(\cdot \,|\, \mu)$ is well known to be convex. See [48, Corollary B.11], for instance, which shows that $I(\cdot \,|\, \mu)$ coincides with the functional J_μ defined in [48, Equation (5.2b)], which is clearly convex. $\square$

See also [56] for an extension of Theorem 1.3.2 to different forms of tensorization oriented toward Markov chains, which however is quite different from our Theorems 1.1.6 or 1.3.1.

1.4 On the Sanov-type theorem and a generalization of Theorem 1.1.4

Continuing in the abstract setting of Chapter 1.3, we next give characterizations of what one might call "$\mathcal{W}_p\,\alpha$ inequalities," for $p = 1, 2$. Assume throughout this section that $\alpha : \mathcal{P}(E) \to (-\infty, \infty]$ is measurable and bounded from below. Define $\alpha_n, \widehat{\alpha}_n : \mathcal{P}(E^n) \to (-\infty, \infty]$ as in (1.29) and (1.33), and define $\rho_n, \widehat{\rho}_n : B(E^n) \to \mathbb{R}$ as in (1.30) and (1.34).

We begin with a simple dual characterization of the $\mathcal{W}_1\,\alpha$ inequality, which generalizes both Theorem 1.1.3 and the equivalence (1) ⇔ (2) of Theorem 1.1.9.

Theorem 1.4.1 ([57, Corollary 3], [36, Theorem 3.5])**.** *Let $C > 0$. The following are equivalent:*

(1) $\mathcal{W}_1^2(\mu, \nu) \le C\alpha(\nu)$ for all $\nu \in \mathcal{P}(E)$.

(2) For each $\lambda \in \mathbb{R}$ and bounded 1-Lipschitz function $f : E \to \mathbb{R}$,

$$\rho_1(\lambda f) \le \lambda \int_E f \, \mathrm{d}\mu + \frac{C\lambda^2}{4}.$$

Proof. This is known from the above references, but we include the straightforward proof for the sake of completeness: Since $C^{-1}x^2 = \sup_{\lambda \geq 0}[\lambda x - (C\lambda^2/4)]$ for $x \geq 0$, (1) is equivalent to

$$\lambda \mathcal{W}_1(\mu, \nu) \leq \alpha(\nu) + \frac{C\lambda^2}{4}, \quad \forall \nu \in \mathcal{P}(E),\ \lambda \geq 0.$$

By Kantorovich duality, this is in turn equivalent to

$$\lambda \int_E f \,\mathrm{d}(\nu - \mu) \leq \alpha(\nu) + \frac{C\lambda^2}{4}, \quad \forall \nu \in \mathcal{P}(E),\ \lambda \geq 0,\ \forall f,$$

where the functions f are understood to be 1-Lipschitz. Using the definition of ρ_1, this is equivalent to

$$\rho_1(\lambda f) = \sup_{\nu \in \mathcal{P}(E)} \left(\lambda \int_E f \,\mathrm{d}\nu - \alpha(\nu) \right) \leq \lambda \int_E f \,\mathrm{d}\mu + \frac{C\lambda^2}{4}, \quad \forall \lambda \geq 0,\ \forall f.$$

□

There is an analogue for $\mathcal{W}_2$ α, which we state next, which generalizes the equivalence of (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) in both Theorems 1.1.4 and 1.1.10. It works for either of the tensorized forms, α_n or $\widehat{\alpha}_n$. Recall in the following that we always equip E^n with the ℓ^2-metric, defined in (1.10).

Theorem 1.4.2. *Assume there exists $x_0 \in E$ such that $\int_E d^2(x, x_0)\, \mu(\mathrm{d}x) < \infty$. Let $C > 0$. The following are equivalent:*

(1) $\mathcal{W}_2^2(\mu, \nu) \leq C\alpha(\nu)$ for all $\nu \in \mathcal{P}(E)$.

(2) For each $n \in \mathbb{N}$, $\mathcal{W}_1^2(\mu, \nu) \leq C\alpha_n(\nu)$ for all $\nu \in \mathcal{P}(E^n)$.

(2') For each $n \in \mathbb{N}$, $\mathcal{W}_1^2(\mu, \nu) \leq C\widehat{\alpha}_n(\nu)$ for all $\nu \in \mathcal{P}(E^n)$.

(3) For each $n \in \mathbb{N}$, $\lambda \in \mathbb{R}$, and bounded 1-Lipschitz function $f : E^n \to \mathbb{R}$, we have

$$\rho_n(\lambda f) \leq \lambda \int_{E^n} f \,\mathrm{d}\mu^{\otimes n} + \frac{C\lambda^2}{4}.$$

(3') Property (3) holds with $\widehat{\rho}_n$ in place of ρ_n.

Proof.

- (1) $\Rightarrow$ (2): Let $n \in \mathbb{N}$ and $\nu \in \mathcal{P}(E^n)$. Apply Jensen's inequality, followed by a known tensorization inequality for $\mathcal{W}_2$ given in [5, Lemma 2.11], and then (1):

$$\begin{aligned}\mathcal{W}_1^2(\mu^{\otimes n}, \nu) \le \mathcal{W}_2^2(\mu^{\otimes n}, \nu) &\le \int_{E^n} \sum_{k=1}^n \mathcal{W}_2^2\big(\mu, \nu_{-k}(x_{-k})\big)\, \nu(\mathrm{d}x) \\ &\le C \int_{E^n} \sum_{k=1}^n \alpha\big(\nu_{-k}(x_{-k})\big)\, \nu(\mathrm{d}x) = C\alpha_n(\nu).\end{aligned}$$

- (1) $\Rightarrow$ (2'): Let $n \in \mathbb{N}$ and $\nu \in \mathcal{P}(E^n)$. Apply a (different) known tensorization inequality for $\mathcal{W}_2$ given in [36, Proposition A.1], and then (1):

$$\begin{aligned}\mathcal{W}_1^2(\mu^{\otimes n}, \nu) \le \mathcal{W}_2^2(\mu^{\otimes n}, \nu) &\le \int_{E^n} \sum_{k=1}^n \mathcal{W}_2^2\big(\mu, \nu_{k-1,k}(x_1, \ldots, x_{k-1})\big)\, \nu(\mathrm{d}x) \\ &\le C \int_{E^n} \sum_{k=1}^n \alpha\big(\nu_{k-1,k}(x_1, \ldots, x_{k-1})\big)\, \nu(\mathrm{d}x) = C\widehat{\alpha}_n(\nu).\end{aligned}$$

- (2) $\Leftrightarrow$ (3) and (2') $\Leftrightarrow$ (3') follow by applying Theorem 1.4.1 to the conjugate pairs (ρ_n, α_n) and $(\widehat{\rho}_n, \widehat{\alpha}_n)$, respectively.

- (3) $\Rightarrow$ (1): Let $M > 0$ and $\lambda \ge 0$. As in the proof of Theorem 1.1.4, define $F : \mathcal{P}(E) \to \mathbb{R}$ by $F := \mathcal{W}_2(\mu, \cdot) \wedge M$, and note that $\sqrt{n}F \circ L_n$ is 1-Lipschitz on E^n for each n. Thus (3) yields

$$\frac{1}{n}\rho_n(\lambda n F \circ L_n) \le \lambda \int_{E^n} F \circ L_n \, \mathrm{d}\mu^{\otimes n} + \frac{C\lambda^2}{4}.$$

The right-hand side converges as $n \to \infty$ to $C\lambda^2/4$ by the law of large numbers in Wasserstein distance. Since F is bounded and lower-semicontinuous, we may apply the lower bound of Theorem 1.3.1 to get

$$\sup_{\nu \in \mathcal{P}(E)} \Big(\lambda \mathcal{W}_2(\mu, \nu) \wedge M - \alpha(\nu)\Big) \le \liminf_{n \to \infty} \frac{1}{n}\rho_n(\lambda n F \circ L_n) \le \frac{C\lambda^2}{4}.$$

Consequently, for all $\nu \in \mathcal{P}(E)$, $\lambda \geq 0$, and $M > 0$, we have

$$\alpha(\nu) \geq \lambda \mathcal{W}_2(\mu,\nu) \wedge M - \frac{C\lambda^2}{4}$$

Send $M \to \infty$ and optimize over λ to get $\alpha(\nu) \geq C^{-1}\mathcal{W}_2^2(\mu,\nu)$.

- (3') $\Rightarrow$ (1): This is proved exactly as (3) $\Rightarrow$ (1), simply replacing ρ_n by $\widehat{\rho}_n$ and applying Theorem 1.3.2 instead of Theorem 1.3.1. $\square$

Chapter 2: A trajectorial approach to the relative entropy dissipation of McKean-Vlasov diffusions

In this chapter, we formulate a trajectorial version of the relative entropy dissipation identity for McKean–Vlasov diffusions, extending recent results which apply to non-interacting diffusions. Our stochastic analysis approach is based on time-reversal of diffusions and Lions' differential calculus over Wasserstein space. It allows us to compute explicitly the rate of relative entropy dissipation along every trajectory of the underlying diffusion via the semimartingale decomposition of the corresponding relative entropy process. As a first application, we obtain a new interpretation of the gradient flow structure for the granular media equation, generalizing a formulation developed recently for the linear Fokker–Planck equation. Secondly, we show how the trajectorial approach leads to a new derivation of the HWBI inequality, which relates relative entropy (H), Wasserstein distance (W), barycenter (B) and Fisher information (I). This chapter is based on the paper [22] joint with Bertram Tschiderer.

2.1 Introduction

We are interested in the relative entropy dissipation of McKean–Vlasov stochastic differential equations of the form

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla (W * \mathrm{P}_t)(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \leq t \leq T, \tag{2.1}$$

where X_0 has some given initial distribution P_0 on $\mathbb{R}^d$. Here, the functions $V, W \colon \mathbb{R}^d \to [0,\infty)$ play the roles of confinement and interaction potentials and are assumed to be suitably regular, $\mathrm{P}_t := \mathrm{Law}(X_t)$ denotes the distribution of the random vector X_t, the symbol $*$ stands for the standard convolution operator, and $(B_t)_{0 \leq t \leq T}$ is a standard n-dimensional Brownian motion. In particular, this SDE is non-local (or non-linear) in the sense that the drift term depends on the distribution of the state variable. Non-local equations of this form arise in the modeling of weakly interacting diffusion equations, after the seminal work of McKean [16].

Since the work of Carrillo–McCann–Villani [17, 18], relative entropy dissipation has been known to be an effective method for studying convergence rates to equilibrium and propagation of chaos of McKean–Vlasov equations. Some notable examples include the works [58, 59, 60, 61, 62, 63]. In a broader context, [64, 65] recently applied entropy methods to the mean-field theory of neural networks.

We denote by $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ the set of absolutely continuous probability measures on $\mathbb{R}^d$, which we will often identify with their corresponding probability density functions with respect to Lebesgue measure. The *free energy functional*

$$\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d) \ni p \longmapsto \mathscr{F}(p) := \mathscr{U}(p) + \mathscr{V}(p) + \mathscr{W}(p) \tag{2.2}$$

is defined as the sum of the energy functionals

$$\mathscr{U}(p) := \int_{\mathbb{R}^d} p(x) \log p(x) \, \mathrm{d}x, \quad \mathscr{V}(p) := \int_{\mathbb{R}^d} V(x) \, p(x) \, \mathrm{d}x, \quad \mathscr{W}(p) := \tfrac{1}{2} \int_{\mathbb{R}^d} (W * p)(x) \, p(x) \, \mathrm{d}x \tag{2.3}$$

corresponding to internal ($\mathscr{U}$), potential ($\mathscr{V}$) and interaction ($\mathscr{W}$) energy, respectively. Defining the *relative entropy dissipation functional*

$$\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d) \ni p \longmapsto \mathscr{D}(p) := \int_{\mathbb{R}^d} |\nabla \log p(x) + \nabla V(x) + \nabla (W * p)(x)|^2 \, p(x) \, \mathrm{d}x, \tag{2.4}$$

the well-known relative entropy dissipation identity takes the form

$$\mathscr{F}(p_t) - \mathscr{F}(p_{t_0}) = - \int_{t_0}^{t} \mathscr{D}(p_u) \, \mathrm{d}u. \tag{2.5}$$

This identity is of a deterministic nature: it only depends on the curve of probability density functions $(p_t)_{0 \le t \le T}$, but not on the trajectories of the underlying process $(X_t)_{0 \le t \le T}$ itself. It is then natural to ask whether there is a process-level analogue of the relative entropy dissipation identity (2.5), depending directly on the trajectories of the McKean–Vlasov process $(X_t)_{0 \le t \le T}$. The main contribution of this paper is to give an affirmative answer to this question, by formulating a trajectorial version of the relative entropy dissipation identity via a stochastic analysis approach.

Before going into details, let us briefly describe the main ideas. We draw inspiration from prior literature [66, 15] based on a simpler (linear) setting without interaction, i.e., $W \equiv 0$. In this case, the McKean–Vlasov SDE (2.1) reduces to a Langevin–Smoluchowski diffusion equation of the form

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \le t \le T. \tag{2.6}$$

In particular, the drift term does not depend on the distribution of X_t. Moreover, there is an explicit stationary distribution (also known as the Gibbs distribution [67, 68, 69]) with density proportional to $\mathbb{R}^d \ni x \mapsto q(x) := \mathrm{e}^{-V(x)}$. Defining the likelihood ratio function (or Radon–Nikodym derivative) $\ell_t(x) := p_t(x)/q(x)$, the free energy at time t can be expressed as $\mathscr{F}(p_t) = \mathbb{E}[\log \ell_t(X_t)]$, and the resulting stochastic process

$$\log \ell_t(X_t) = \log p_t(X_t) + V(X_t)\,, \qquad 0 \le t \le T \tag{2.7}$$

is called *free energy* or *relative entropy process*. As shown in [66, 15], the time-reversal

$$\big(\log \ell_{T-s}(X_{T-s})\big)_{0 \le s \le T}$$

of this process is a submartingale, and Itô calculus can be used to obtain its Doob–Meyer decomposition

$$\log \ell_{T-s}(X_{T-s}) - \log \ell_T(X_T) = M_{T-s} + F_{T-s}. \tag{2.8}$$

Here, $(M_{T-s})_{0 \le s \le T}$ is a martingale and $(F_{T-s})_{0 \le s \le T}$ is an increasing process of finite first variation, both with explicit expressions. This decomposition describes exactly the rate of relative entropy dissipation along every trajectory of the Langevin–Smoluchowski diffusion. Therefore, it can be viewed as a trajectorial analogue of the (deterministic) relative entropy dissipation identity (2.5).

Let us now return to our McKean–Vlasov setting. In order to take into account the interaction potential W, it is natural to consider a generalized relative entropy process of the form

$$\log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t)\,, \qquad 0 \leq t \leq T. \tag{2.9}$$

The task is now to compute the semimartingale decomposition of this process. We will provide a detailed analysis of this extension, which is subtler than might appear at first sight. The main difficulty is that, even when it exists, the stationary distribution of the McKean–Vlasov diffusion does not have a closed-form expression and is not even unique in general; see the works [70, 71, 72, 61, 73, 74, 75, 76]. This prevents us from defining the likelihood ratio function in a straightforward manner as in the setting of Langevin–Smoluchowski diffusions, where one can rely on the invariant Gibbs distribution. An appropriate definition of the generalized likelihood ratio function turns out to be that (2.9) should be viewed as a function of the form $\log \ell_t(X_t, \mathrm{P}_t)$, depending explicitly on the distribution P_t of X_t itself, in addition to the state X_t. This form of generalized likelihood ratio function allows us to take the *L-derivative* with respect to the probability distribution P_t. The notion of L-differentiation for functions of probability measures was introduced by Lions [77]. We refer to the monograph [78, Chapter 5] for a detailed discussion of differential calculus and stochastic analysis over spaces of probability measures. In particular, we will use a generalized form of Itô's formula for functions of curves of measures, to derive the dynamics of the time-reversal of the relative entropy process (2.9), in terms of the semimartingale decomposition

$$\log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}) - \log \ell_T(X_T, \mathrm{P}_T) = M_{T-s} + F_{T-s}\,, \qquad 0 \leq s \leq T, \tag{2.10}$$

where $(M_{T-s})_{0 \leq s \leq T}$ is a martingale and $(F_{T-s})_{0 \leq s \leq T}$ is a process of finite first variation, both of which will be explicitly computed. Similar to the case of Langevin–Smoluchowski dynamics, this decomposition can be viewed as the trajectorial rate of relative entropy dissipation. The classical (deterministic) identity (2.5) can then be recovered by taking expectations.

2.1.1 Gradient flow structure of the granular media equation

As a first application of our trajectorial approach we obtain a new interpretation of the gradient flow structure of the *granular media equation*

$$\partial_t p_t(x) = \operatorname{div}\Big(\nabla p_t(x) + p_t(x)\nabla V(x) + p_t(x)\nabla(W * p_t)(x)\Big), \qquad (t,x) \in (0,T)\times\mathbb{R}^d, \tag{2.11}$$

which describes the evolution of the curve of probability density functions $(p_t)_{0\le t\le T}$ corresponding to the McKean–Vlasov diffusion $(X_t)_{0\le t\le T}$ of (2.1). When $n = 1$, this PDE appears in the modeling of the time evolution of granular media [79, 80, 81]; in that context, the granular medium is modeled as system of particles performing inelastic collisions, and $p_t(x)$ is regarded as the velocity of a representative particle in the system at time t and position x, while V and W represent the friction and the inelastic collision forces, respectively. Note that in the interaction-free case $W \equiv 0$, the equation (2.11) reduces to a linear *Fokker–Planck equation*. As is well known from [17, 18], this curve of probability densities can be characterized as a *gradient flow* in $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$, the space of absolutely continuous probability measures with finite second moments. Roughly speaking, this is an optimality property stating that the curve $(p_t)_{0\le t\le T}$ evolves in the direction of steepest possible descent for the free energy functional (2.2) with respect to the quadratic Wasserstein distance.

The Wasserstein gradient flow structure of the linear Fokker–Planck equation was first discovered by Jordan, Kinderlehrer and Otto in the seminal work [14]. In the paper [21], Otto and Villani developed a formal Riemannian structure on the space of probability measures with finite second moments, leading to heuristic proofs of gradient flow properties as in [82], where the porous medium equation was studied. This pioneering approach is often referred to as “Otto calculus”. Later, a rigorous framework based on minimizing movement schemes and curves of maximal slope was introduced in [19]. Recently, a trajectorial approach to the gradient flow properties of

Langevin–Smoluchowski diffusions [15] and Markov chains [83] was established. We will follow this approach and adapt it to our McKean–Vlasov setting. For gradient flows of McKean–Vlasov equations on discrete spaces we refer to [84].

Returning to the setting of this paper, our main result leads to a new formulation of the gradient flow property of the granular media equation. To show this steepest descent property, the main idea is to consider a *perturbed* McKean–Vlasov diffusion of the form

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla\beta(X_t) + \nabla(W * \mathrm{P}_t^{\beta})(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\beta}\,, \qquad t_0 \leq t \leq T \tag{2.12}$$

which is constructed by adding a perturbation $\beta\colon \mathbb{R}^d \to \mathbb{R}$ to the confinement potential[1] of the original McKean–Vlasov SDE (2.1). In other words, from time t_0 onward, the perturbed diffusion drifts in a direction different from that of the original diffusion, hence the perturbed curve of time-marginal distributions $(\mathrm{P}_t^{\beta})_{t_0 \leq t \leq T}$ also evolves differently from the unperturbed curve $(\mathrm{P}_t)_{t_0 \leq t \leq T}$. In parallel with the unperturbed case, we may compute the dynamics of the perturbed relative entropy process associated with (2.12). As a consequence, we derive the rate of relative entropy dissipation for the perturbed McKean–Vlasov diffusion. On the other hand, the rate of change of the Wasserstein distance along the perturbed curve $(\mathrm{P}_t^{\beta})_{t_0 \leq t \leq T}$ can be computed based on the general theory of metric derivative of absolutely continuous curves, see [19]. Finally, comparing these two rates in both the perturbed and unperturbed settings, allows us to establish the gradient flow property.

2.1.2 The HWBI inequality

The second application of our trajectorial approach deals with the HWBI inequality [20, Theorem 4.2], which is an extension of the HWI inequality [21]. It relates not only relative entropy (H), Wasserstein distance (W), and relative Fisher information (I), but also barycenter (B). These quantities are defined as follows: for two probability measures $\nu, \mu \in \mathcal{P}(\mathbb{R}^d)$, the *relative entropy*

[1] As we will see, the steepest descent property is already visible by perturbing the confinement potential from V to $V + \beta$, thus we avoid complicating the setup further by adding another perturbation to the interaction potential W.

of ν with respect to μ is defined by

$$H(\nu \mid \mu) \coloneqq \begin{cases} \int_{\mathbb{R}^d} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}) \, \mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.13}$$

the *relative Fisher information* of ν with respect to μ is given by

$$I(\nu \mid \mu) \coloneqq \begin{cases} \int_{\mathbb{R}^d} |\nabla \log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu})|^2 \, \mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.14}$$

and the *barycenter* of a probability measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as $b(\nu) \coloneqq \int_{\mathbb{R}^d} x \, \mathrm{d}\nu(x) \in \mathbb{R}^d$, where the integral is understood as a Bochner integral. Informally, the HWBI inequality then states that any two probability measures $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy

$$H(\nu_0 \mid \mu_0) - H(\nu_1 \mid \mu_1) \leq \sqrt{I(\nu_0 \mid \mu_0^{\uparrow})} \, \mathcal{W}_2(\nu_0, \nu_1) - \tfrac{\kappa_V + \kappa_W}{2} \, \mathcal{W}_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2} \, |b(\nu_0) - b(\nu_1)|^2, \tag{2.15}$$

where $\mu_0, \mu_1, \mu_0^{\uparrow}$ are some appropriate σ-finite reference measures depending on the potentials V, W (see Chapter 2.4.3 for the details), and $\kappa_V, \kappa_W \in \mathbb{R}$ are the moduli of uniform convexity for V, W. This inequality describes the evolution of the relative entropy along the *displacement interpolation* $(\nu_t)_{0 \leq t \leq 1}$ between ν_0 and ν_1. Compared with the HWI inequality, there are two additional terms on the right-hand side of (2.15) contributed by the interaction energy functional $\mathcal{W}$ of (2.3). Intuitively, the κ_W-uniform convexity of W leads to the first additional term $-\frac{\kappa_W}{2} \mathcal{W}_2^2(\nu_0, \nu_1)$, which alone would correspond to the κ_W-uniform displacement convexity of $\mathscr{W}$ along $(\nu_t)_{0 \leq t \leq 1}$. But since $\mathscr{W}(p)$ is invariant under any translation of p, the functional $\mathscr{W}$ might fail to be uniformly displacement convex when the barycenter shifts. This suggests that the barycentric shift along $(\nu_t)_{0 \leq t \leq 1}$ should be factored out of the consideration of the displacement convexity of $\mathscr{W}$, which is intuitively why the second additional term $\frac{\kappa_W}{2} |b(\nu_0) - b(\nu_1)|^2$ in (2.15) appears.

Coming back to our second application, we illustrate how our approach yields a trajectorial proof of the inequality (2.15), in the slightly strengthened form of [85, Theorem 4.1] and [86, Theorem D.50]. Much of this consists of arguments similar in spirit to our main result (2.10), but with one key difference: instead of the time-marginals of the McKean–Vlasov diffusion, we apply the trajectorial approach to the displacement interpolation $(\nu_t)_{0 \leq t \leq 1}$. In this regard, our derivation can be seen as a generalization of the trajectorial proof of the HWI inequality in [15, Section 4.2]; see also [83, Section 9.4], where the same idea was used to derive a discrete version of the HWI inequality in a Riemannian-geometric framework. Let us also point out that for the proof of the HWBI inequality we shall impose convexity assumptions (see Assumptions 2.4.18) on the potentials V, W. We do not require these assumptions in the rest of the paper.

In the literature, similar trajectorial approaches have also been applied in the context of martingale inequalities [87, 88], functional inequalities [89, 90, 91, 92], and their stability estimates [93, 94]. In particular, we refer to [91, Corollary 1.4] for a related HWI inequality derived from the entropic interpolation of the mean-field Schrödinger problem.

2.2 Organization of the chapter

We set up the probabilistic framework and discuss some regularity assumptions in Chapter 2.3. In Chapter 2.4, we state our main trajectorial results, Theorem 2.4.1 and Theorem 2.4.9, and develop two explicit examples for illustration. As immediate consequences, we derive the classical relative entropy dissipation identities in Corollary 2.4.4 and Corollary 2.4.10. Building on these results, we formulate the gradient flow property of the granular media equation in Theorem 2.4.15. The HWBI inequality is then stated in Theorem 2.4.19. The proofs of the trajectorial results and of the HWBI inequality are developed in Chapter 2.5. Some proofs of auxiliary results postponed in previous parts are contained in Chapter 2.6.

2.3 The probabilistic framework

2.3.1 The setting

We fix a terminal time $T \in (0, \infty)$ and let $\Omega \coloneqq C([0,T];\mathbb{R}^d)$ be the path space of $\mathbb{R}^d$-valued continuous functions defined on $[0,T]$. We denote by $(X_t)_{0\le t\le T}$ the canonical process defined by $X_t(\omega) \coloneqq \omega(t)$ for $\omega \in \Omega$, and fix a probability distribution $\mathrm{P}_0 \in \mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$.

As will be shown in Lemma 2.3.2, under the Assumptions 2.3.1 below, the SDE (2.1) with initial distribution P_0 has a unique strong solution, when it is posed on an arbitrary filtered probability space. This implies that there exists a probability measure $\mathbb{P}$ on Ω and a $\mathbb{P}$-Brownian motion $(B_t)_{0\le t\le T}$ such that the SDE (2.1) holds. We write $\mathcal{F} = (\mathcal{F}_t)_{0\le t\le T}$ for the right-continuous augmentation of the canonical filtration.

For each time $t \in [0,T]$, we denote by $\mathrm{P}_t \coloneqq \mathbb{P} \circ X_t^{-1}$ the distribution of X_t under $\mathbb{P}$, and by p_t the corresponding probability density function on $\mathbb{R}^d$. The density functions $(p_t)_{0\le t\le T}$ then solve the granular media equation (2.11).

2.3.2 Regularity assumptions

Assumptions 2.3.1. The following regularity assumptions will be used frequently.

(i) The functions $V, W\colon \mathbb{R}^d \to [0,\infty)$ are smooth and have Lipschitz continuous gradients with Lipschitz constants $\|\nabla V\|_{\mathrm{Lip}}, \|\nabla W\|_{\mathrm{Lip}}$. All derivatives of V and W grow at most exponentially as $|x|$ tends to infinity, and the first derivatives are of linear growth. The latter condition means that there exists a constant $C > 0$ such that

$$\forall\, x \in \mathbb{R}^d\colon \qquad |\nabla V(x)| \le C(1+|x|)\,, \qquad |\nabla W(x)| \le C(1+|x|). \tag{2.16}$$

Furthermore, the function W is even (in other words, symmetric), i.e., $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$.

(ii) The probability distribution P_0 is an element of the space $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$ and the corresponding probability density function $\mathbb{R}^d \ni x \mapsto p_0(x)$ is strictly positive. Moreover, the initial free energy $\mathscr{F}(p_0)$ is finite.

These assumptions ensure that the equation (2.1) belongs to a broad class of strongly solvable McKean–Vlasov SDEs. We relegate the proof of the following result to Chapter 2.6.1.

Lemma 2.3.2. *Suppose Assumptions 2.3.1 hold. Then on an arbitrary filtered probability space, the McKean–Vlasov SDE* (2.1) *has a pathwise unique, strong solution* $(X_t)_{0\leq t\leq T}$ *satisfying*

$$\mathbb{E}\bigg[\sup_{0\leq t\leq T}|X_t|^2\bigg] < \infty. \tag{2.17}$$

Moreover, its marginal distributions $(\mathrm{P}_t)_{0\leq t\leq T}$ *belong to* $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$*, and the corresponding curve of probability density functions* $(p_t)_{0\leq t\leq T}$ *is a classical solution of the granular media equation* (2.11).

2.3.3 Probabilistic representations of gradient flow functionals

To set up our framework, the first step is to express the free energy as well as the relative entropy dissipation functional in probabilistic terms. To this end, we introduce the generalized potential $\Psi\colon \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to [0,\infty)$ and its close relative $\Psi^{\uparrow}$ given by

$$\Psi(x,\mu) := V(x) + \tfrac{1}{2}(W * \mu)(x)\,, \qquad \Psi^{\uparrow}(x,\mu) := V(x) + (W * \mu)(x) \tag{2.18}$$

for $(x,\mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. Furthermore, we define the density functions

$$q(x,\mu) := e^{-\Psi(x,\mu)}\,, \qquad q^{\uparrow}(x,\mu) := e^{-\Psi^{\uparrow}(x,\mu)}\,, \qquad q^{\downarrow}(x) := e^{-V(x)} \tag{2.19}$$

and the corresponding generalized likelihood ratio functions

$$\ell_t(x,\mu) := \frac{p_t(x)}{q(x,\mu)}\,, \qquad \ell_t^{\uparrow}(x,\mu) := \frac{p_t(x)}{q^{\uparrow}(x,\mu)}\,, \qquad \ell_t^{\downarrow}(x) := \frac{p_t(x)}{q^{\downarrow}(x)} \tag{2.20}$$

for $t \in [0, T]$. Note that if $W \equiv 0$, these three likelihood ratio functions coincide.

For each time $t \in [0, T]$, we introduce σ-finite measures on the Borel sets of $\mathbb{R}^d$, given by

$$\mathrm{Q}_t(A) := \int_A q(x, \mathrm{P}_t)\, \mathrm{d}x\,, \qquad \mathrm{Q}_t^{\uparrow}(A) := \int_A q^{\uparrow}(x, \mathrm{P}_t)\, \mathrm{d}x\,, \qquad A \in \mathcal{B}(\mathbb{R}^d). \tag{2.21}$$

Intuitively, these measures are (unnormalized) time-dependent Gibbs distributions. If $W \equiv 0$, they coincide with the true Gibbs distribution of the Langevin–Smoluchowski equation (2.6), which is also its stationary distribution (when normalized to a probability measure).

With these definitions, we can now write the gradient flow functionals $\mathscr{F}$ and $\mathscr{D}$, introduced in (2.2) and (2.4), in probabilistic terms: the relative entropy (defined in (2.13)) of P_t with respect to Q_t and the relative Fisher information (defined in (2.14)) of P_t with respect to $\mathrm{Q}_t^{\uparrow}$ can be expressed respectively as

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = \mathbb{E}_{\mathbb{P}}\big[\log \ell_t(X_t, \mathrm{P}_t)\big]\,, \qquad I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}) = \mathbb{E}_{\mathbb{P}}\big[|\nabla \log \ell_t^{\uparrow}(X_t, \mathrm{P}_t)|^2\big]\,; \tag{2.22}$$

and we have the relations $H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = \mathscr{F}(\mathrm{P}_t)$ as well as $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}) = \mathscr{D}(\mathrm{P}_t)$. In particular, the relative entropy $H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ can be written as the $\mathbb{P}$-expectation of the *relative entropy process*

$$\log \ell_t(X_t, \mathrm{P}_t) = \log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t)\,, \qquad 0 \le t \le T. \tag{2.23}$$

The dynamics of this stochastic process, together with its perturbed counterpart to be introduced in Chapter 3.2.3 below, will be our main objects of interest.

Remark 2.3.3. If the reference measure Q_t in (2.21) is a *probability* measure, then the expression (2.22) matches the classical definition of relative entropy given in (2.13). In the general case when Q_t is a σ-finite measure, the definition (2.13) is also valid under the condition that P_t has finite second moment, with the only difference that the range of the function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is extended from $[0, \infty]$ to $(-\infty, \infty]$; we refer to [95, Appendix C] or [96, Section 3] for the details.

2.4 Main results

2.4.1 Trajectorial dissipation of relative entropy for McKean–Vlasov diffusions

Our first main result is the semimartingale decomposition of the relative entropy process (2.23). It describes the dissipation of relative entropy along every trajectory of a particle undergoing the McKean–Vlasov dynamics (2.1). In the same spirit as the trajectorial approaches of [66] and [15], we shall study the dynamics of the relative entropy process in the backward direction of time. Concretely, we consider for arbitrary, fixed $T \in (0, \infty)$ the time-reversed canonical process

$$\overline{X}_s := X_{T-s}\,, \qquad 0 \leq s \leq T \tag{2.24}$$

on the filtered probability space $(\Omega, \mathbb{G}, \mathbb{P})$, where $\mathbb{G} = (\mathcal{G}_s)_{0 \leq s \leq T}$ is the $\mathbb{P}$-augmented filtration generated by $(\overline{X}_s)_{0 \leq s \leq T}$.

In order to formulate Theorem 2.4.1 below, we introduce the time-reversed *Fisher information process*

$$\begin{aligned} \overline{I}_s := \Big(\big|\nabla \log \overline{\ell}^{\downarrow}_s\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}_s)\big|^2 + \Big\langle \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_s)\,,\, 2\nabla \log \overline{\ell}^{\downarrow}_s + \nabla V \Big\rangle \Big)(\overline{X}_s) \\ - \mathbb{E}_{\tilde{\mathbb{P}}}\Big[\Big\langle \tfrac{1}{2}\nabla W(\overline{X}_s - \overline{Y}_s)\,,\, \Big(2\nabla \log \overline{\ell}^{\downarrow}_s - \nabla V + \nabla(W * \overline{\mathrm{P}}_s)\Big)(\overline{Y}_s) \Big\rangle \Big] \end{aligned} \tag{2.25}$$

for $0 \leq s \leq T$. Here, the process $(\overline{Y}_s)_{0 \leq s \leq T}$ is defined on another probability space $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}})$ such that the tuple $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}, (\overline{Y}_s)_{0 \leq s \leq T})$ is an exact copy of $(\Omega, \mathbb{G}, \mathbb{P}, (\overline{X}_s)_{0 \leq s \leq T})$. A bar over a letter means that time is reversed as in (2.24).

We also define the time-reversed *cumulative Fisher information process* as the time integral

$$\overline{F}_s := \int_0^s \overline{I}_u \, \mathrm{d}u\,, \qquad 0 \leq s \leq T. \tag{2.26}$$

This process will act as the compensator in the semimartingale decomposition of the relative entropy process (2.23). Its relation with the relative Fisher information (2.22) will be given in (3.17) below.

Theorem 2.4.1. *Suppose Assumptions 2.3.1 hold. On* $(\Omega, \mathbb{G}, \mathbb{P})$*, the time-reversed relative entropy process*

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) = \log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}), \qquad 0 \leq s \leq T \tag{2.27}$$

admits the semimartingale decomposition

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) - \log \overline{\ell}_0(\overline{X}_0, \overline{\mathrm{P}}_0) = \overline{M}_s + \overline{F}_s. \tag{2.28}$$

Here $(\overline{M}_s)_{0 \leq s \leq T}$ *is the* $L^2(\mathbb{P})$*-bounded martingale*

$$\overline{M}_s := \int_0^s \Big\langle \nabla \log \overline{\ell}_u(\overline{X}_u, \overline{\mathrm{P}}_u) \, , \, \sqrt{2}\, \mathrm{d}\overline{B}_u \Big\rangle, \tag{2.29}$$

with $(\overline{B}_s)_{0 \leq s \leq T}$ *a* $\mathbb{P}$*-Brownian motion of the backward filtration* $\mathbb{G}$*, and the compensator* $(\overline{F}_s)_{0 \leq s \leq T}$ *satisfies*

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s I\big(\overline{\mathrm{P}}_u \,\big|\, \overline{\mathrm{Q}}_u^{\uparrow}\big)\, \mathrm{d}u = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^s \big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2 \, \mathrm{d}u\bigg] < \infty. \tag{2.30}$$

Examples

We give two concrete examples to illustrate Theorem 2.4.1.

Example 2.4.2. We set $n = 1$ and specialize Theorem 2.4.1 to the case of quadratic confinement potential $V(x) = \frac{x^2}{2}$ and no interaction potential $W \equiv 0$. The initial position X_0 is chosen to be independent of $(B_t)_{0 \leq t \leq T}$ and to be normally distributed with mean 0 and variance $\sigma_0^2 > 0$. In this case, the SDE of (2.1) becomes

$$\mathrm{d}X_t = -X_t \, \mathrm{d}t + \sqrt{2}\, \mathrm{d}B_t \, , \qquad 0 \leq t \leq T \tag{2.31}$$

and its solution is given by the Ornstein–Uhlenbeck process

$$X_t = \mathrm{e}^{-t} X_0 + \sqrt{2} \int_0^t \mathrm{e}^{u-t} \, \mathrm{d}B_u \,, \qquad 0 \le t \le T \tag{2.32}$$

with probability density function

$$p_t(x) = \tfrac{1}{\sqrt{2\pi\sigma_t^2}} \exp\big(- \tfrac{x^2}{2\sigma_t^2} \big) \,, \qquad \sigma_t^2 \coloneqq 1 + \mathrm{e}^{-2t}(\sigma_0^2 - 1). \tag{2.33}$$

Recalling (2.18) – (2.20) and using (2.33), we see that in this setting the cumulative Fisher information process of (2.26) is explicitly given by

$$\overline{F}_s^{\mathrm{OU}} = \int_0^s \Big(\nabla \log \overline{\ell}_u^{\downarrow}(\overline{X}_u) \Big)^2 \mathrm{d}u = \int_0^s \Big(\nabla \log \overline{p}_u(\overline{X}_u) + \overline{X}_u \Big)^2 \mathrm{d}u = \int_0^s \big(1 - \tfrac{1}{\overline{\sigma}_u^2}\big)^2 \, \overline{X}_u^2 \, \mathrm{d}u \tag{2.34}$$

for $0 \le s \le T$. In particular, the non-negativity of the integrand in (2.34) implies that the relative entropy decreases along almost every trajectory.

Now we set $T = 1$ and $\sigma_0^2 = 0.1$. The blue lines in Figure 2.1 below are ten simulated trajectories $s \mapsto \overline{F}_s^{\mathrm{OU}}(\omega_i)$, for $i = 1, \ldots, 10$. The thick black line plots the expected path $s \mapsto \mathbb{E}_{\mathbb{P}}[\overline{F}_s^{\mathrm{OU}}]$ of all possible trajectories.

Example 2.4.3. We set again $n = 1$ and now consider the case of no confinement potential $V \equiv 0$, quadratic interaction potential $W(x) = \frac{x^2}{2}$, and a centered Gaussian initial position X_0 with variance $\sigma_0^2 > 0$, which is independent of the Brownian motion $(B_t)_{t \ge 0}$. In this case, for any $t \ge 0$, the drift term of the SDE of (2.1) is

$$- \nabla(W * \mathrm{P}_t)(X_t) = - \int_{\mathbb{R}^d} \nabla W(X_t - y) p_t(y) \, \mathrm{d}y = -\big(X_t - \mathbb{E}[X_t]\big). \tag{2.35}$$

In particular, the drift term depends on the distribution P_t only through its mean. Substituting it

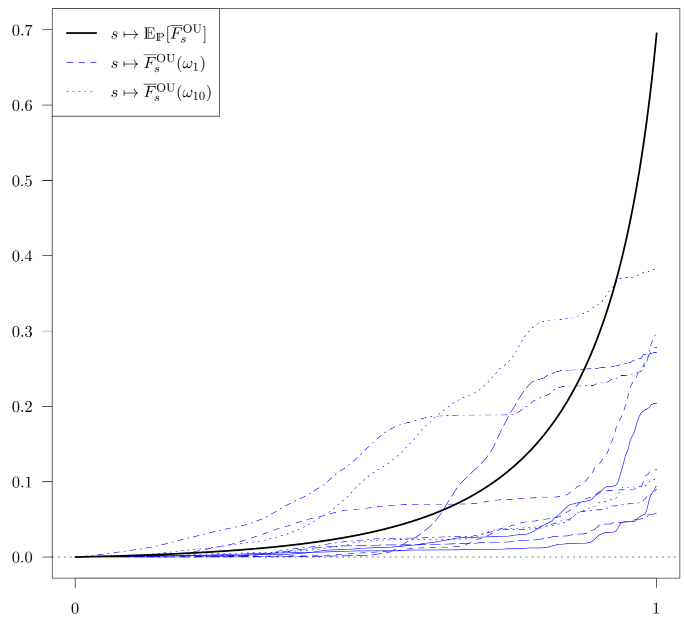

Figure 2.1: Simulations of the cumulative Fisher information process (2.34) for the Ornstein–Uhlenbeck diffusion (2.31).

into (2.1) , this SDE reduces to

$$\mathrm{d}X_t = -\big(X_t - \mathbb{E}[X_t]\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t\,, \qquad 0 \leq t \leq T. \tag{2.36}$$

This type of nonlinear, self-interacting SDE has been studied since [70], where it was shown that its solution is also given by the Ornstein–Uhlenbeck process of (2.32). Therefore, similar computations as in Example 2.4.2 show that in this setting the cumulative Fisher information process is given by

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\overline{X}_u^2 + \mathbb{E}_{\tilde{\mathbb{P}}}\Big[\tfrac{1}{2}(\overline{X}_u - \overline{Y}_u)\big(\tfrac{2}{\overline{\sigma}_u^2}\overline{Y}_u - \overline{Y}_u\big)\Big]\Big)\,\mathrm{d}u \tag{2.37}$$

for $0 \leq s \leq T$. Using the fact that $(\overline{X}_u)_{\#}(\tilde{\mathbb{P}}) = (\overline{Y}_u)_{\#}(\tilde{\mathbb{P}}) = \mathcal{N}(0, \overline{\sigma}_u^2)$, which we have from (2.33), we can compute the expectation appearing in (2.37) and obtain

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\overline{X}_u^2 + \tfrac{\overline{\sigma}_u^2}{2} - 1\Big)\,\mathrm{d}u. \tag{2.38}$$

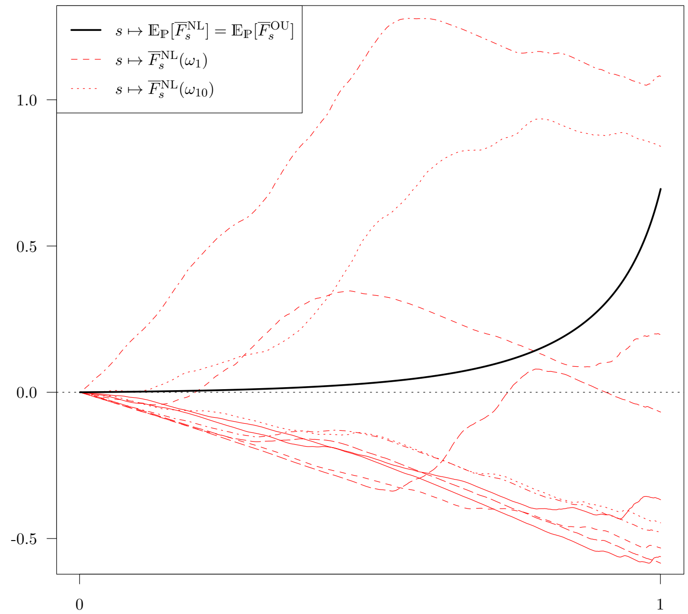

Figure 2.2: Simulations of the cumulative Fisher information process (2.38) for the nonlinear, self-interacting diffusion (2.36).

Clearly, $\overline{F}_s^{\mathrm{OU}} \neq \overline{F}_s^{\mathrm{NL}}$. In particular, the integrand in (2.38) is non-negative if and only if $\overline{X}_u^2 \geq \big(\frac{1}{\overline{\sigma}_u^4} + \frac{1}{2} - \frac{1}{\overline{\sigma}_u^2}\big)^{-1}\big(1 - \frac{\overline{\sigma}_u^2}{2}\big)$. In other words, as opposed to (2.4.2), relative entropy only decreases along a trajectory if $\overline{X}_u$ is far from its mean. However, after taking expectations in (2.34) and (2.38), we see that the expected rate of relative entropy dissipation in both cases is equal to

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{OU}}\big] = \mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{NL}}\big] = \int_0^s \big(\overline{\sigma}_u - \tfrac{1}{\overline{\sigma}_u}\big)^2 \,\mathrm{d}u\,, \qquad 0 \leq s \leq T. \tag{2.39}$$

Now we set again $T = 1$ and $\sigma_0^2 = 0.1$. In the same vein as in Figure 2.1, we plot in Figure 2.2 below the paths of ten simulated trajectories $s \mapsto \overline{F}_s^{\mathrm{NL}}(\omega_i)$, for $i = 1, \ldots, 10$. We observe that some of the red lines describing the paths of these trajectories indeed take negative values. In other words, the cumulative Fisher information process of (2.38), and hence its integrand, can both be negative. Finally, the thick black line in Figure 2.2 follows the expected path $s \mapsto \mathbb{E}_{\mathbb{P}}[\overline{F}_s^{\mathrm{NL}}]$ of all possible trajectories. According to (2.39), this is the same black line as in (2.1).

Consequences of Theorem 2.4.1

We now return to the general statement of Theorem 2.4.1 and deduce some direct consequences. By averaging the trajectorial result of Theorem 2.4.1 according to the path measure $\mathbb{P}$, we derive the well-known *relative entropy identity* (2.40) and the *dissipation of relative entropy* (2.41) below. A sketch of proof for the latter result was first given in [17, Proposition 2.1].

Corollary 2.4.4. *Suppose Assumptions 2.3.1 hold. For all $t, t_0 \in [0,T]$, we have the relative entropy identity*

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0}) = -\int_{t_0}^{t} I(\mathrm{P}_u \,|\, \mathrm{Q}_u^{\uparrow})\, \mathrm{d}u. \tag{2.40}$$

In particular, the relative entropy function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is monotonically decreasing. Furthermore, for Lebesgue-a.e. $t \in [0,T]$, the relative Fisher information $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow})$ is finite, and the rate of relative entropy dissipation is given by

$$\tfrac{\mathrm{d}}{\mathrm{d}t} H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = -I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}). \tag{2.41}$$

Proof. The identity (2.40) follows by taking expectations with respect to the probability measure $\mathbb{P}$ in (2.28), recalling the definitions of (2.22), using (3.17), and invoking the fact that the $\mathbb{P}$-expectation of the martingale (2.29) vanishes. Finally, applying the Lebesgue differentiation theorem to the monotone function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ gives (2.41). $\square$

Remark 2.4.5. The relation (2.41) describes the temporal dissipation of relative entropy at the *ensemble* level. It asserts that the rate of decay of the relative entropy $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ is given by the relative Fisher information $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow})$.

Finally, let us place ourselves again on the filtered probability space $(\Omega, \mathbb{G}, \mathbb{P})$ as in Theorem 2.4.1 and emphasize that this trajectorial result is valid along almost every trajectory $s \mapsto \overline{X}_s(\omega)$ of the underlying McKean–Vlasov process. As a consequence, by taking conditional expectations, we can generalize (2.41) and deduce the following trajectorial rate of relative entropy dissipation.

Corollary 2.4.6. *Suppose* Assumptions 2.3.1 *hold and* $\int_0^T \mathbb{E}_{\mathbb{P}}[|\overline{I}_u|]\,\mathrm{d}u < \infty$*. For* $\mathbb{P}$*-a.e.* $\omega \in \Omega$ *there exists a Lebesgue null set* $N_\omega \subseteq [0,T]$ *such that for any* $s_0 \in [0,T] \setminus N_\omega$ *we have*

$$\lim_{s \downarrow s_0} \frac{\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big)}{s - s_0} = \overline{I}_{s_0}(\omega). \tag{2.42}$$

Remark 2.4.7. Recalling (3.17), we observe that $\mathbb{E}_{\mathbb{P}}[\overline{I}_{s_0}] = I(\overline{\mathrm{P}}_{s_0} \,|\, \overline{\mathrm{Q}}^{\uparrow}_{s_0})$. Therefore the limiting assertion (2.42) can indeed be viewed as a trajectorial version of the deterministic relative entropy dissipation identity (2.41).

Proof. We let $0 \leq s_0 \leq s \leq T$. By (2.28), (2.26) and Fubini's theorem, we have for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big) = \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\,\mathrm{d}u. \tag{2.43}$$

Furthermore, Jensen's inequality gives

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_{s_0}^{s} \big|\mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big]\big|\,\mathrm{d}u\bigg] \leq \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[|\overline{I}_u|\big]\,\mathrm{d}u < \infty, \tag{2.44}$$

which implies

$$\int_{s_0}^{s} \big|\mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\big|\,\mathrm{d}u < \infty \qquad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{2.45}$$

By the Lebesgue differentiation theorem, for every such ω there exists a Lebesgue null set $N_\omega \subseteq [0,T]$ so that the limiting assertion

$$\lim_{s \downarrow s_0} \frac{\int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\,\mathrm{d}u}{s - s_0} = \mathbb{E}_{\mathbb{P}}\big[\overline{I}_{s_0} \,|\, \mathcal{G}_{s_0}\big](\omega) = \overline{I}_{s_0}(\omega) \tag{2.46}$$

holds for every $s_0 \in [0,T] \setminus N_\omega$. Finally, combining (2.43) and (2.46) proves (2.42). $\square$

2.4.2 Gradient flow structure of the granular media equation

In this section, we apply the trajectorial approach of Chapter 2.4.1 in order to formulate the gradient flow property of the granular media equation (2.11). To this end, we consider a function

$\beta\colon \mathbb{R}^d \to \mathbb{R}$, which will be treated as a perturbation potential. We denote by $V^\beta := V + \beta$ the perturbed confinement potential and invoke the following regularity assumptions.

Assumptions 2.4.8. The function $\beta\colon \mathbb{R}^d \to \mathbb{R}$ is smooth and compactly supported, and we require that Assumptions 2.3.1 are still satisfied if we replace V by V^β.

Note that Assumptions 2.4.8 allow us to apply Lemma 2.3.2 to the "perturbed" McKean–Vlasov SDE

$$\mathrm{d}X_t = -\big(\nabla V^\beta(X_t) + \nabla(W * \mathrm{P}_t^\beta)(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^\beta\,, \qquad t_0 \le t \le T \tag{2.47}$$

starting at time $t_0 \ge 0$, with X_{t_0} having initial distribution $\mathrm{P}_{t_0}^\beta = \mathrm{P}_{t_0}$. Therefore, by analogy with Chapter 2.3.1, we can construct a probability measure $\mathbb{P}^\beta$ on $\Omega := C([t_0,T];\mathbb{R}^d)$, under which the canonical process $(X_t)_{t_0\le t\le T}$ satisfies the SDE (2.47), with $(B_t^\beta)_{t_0\le t\le T}$ being a $\mathbb{P}^\beta$-Brownian motion.

For each time $t \in [t_0,T]$, we denote by $\mathrm{P}_t^\beta := \mathbb{P}^\beta \circ X_t^{-1}$ the probability distribution and by p_t^β the probability density function of X_t under $\mathbb{P}^\beta$. The "perturbed" curve of density functions $(p_t^\beta)_{t_0\le t\le T}$ then satisfies the *perturbed granular media equation*

$$\begin{cases} \partial_t p_t^\beta(x) = \operatorname{div}\Big(\nabla p_t^\beta(x) + p_t^\beta(x)\nabla V^\beta(x) + p_t^\beta(x)\nabla(W * p_t^\beta)(x)\Big)\,, & (t,x) \in (t_0,T)\times\mathbb{R}^d, \\ p_{t_0}^\beta(x) = p_{t_0}(x)\,, & x \in \mathbb{R}^d. \end{cases} \tag{2.48}$$

By analogy with (2.18), we define the perturbed potentials

$$\Psi^\beta(x,\mu) := V^\beta(x) + \tfrac{1}{2}(W*\mu)(x)\,, \qquad \Psi^{\beta\uparrow}(x,\mu) := V^\beta(x) + (W*\mu)(x)\,, \qquad \Psi^{\beta\downarrow} := V^\beta \tag{2.49}$$

for $(x,\mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. In parallel to (2.20), we introduce the *perturbed likelihood ratio* *functions*

$$\ell_t^\beta(x,\mu) \coloneqq \frac{p_t^\beta(x)}{q(x,\mu)}\,, \qquad \ell_t^{\beta\uparrow}(x,\mu) \coloneqq \frac{p_t^\beta(x)}{q^\uparrow(x,\mu)}\,, \qquad \ell_t^{\beta\downarrow}(x) \coloneqq \frac{p_t^\beta(x)}{q^\downarrow(x)} \tag{2.50}$$

for $t \in [t_0, T]$. Finally, we define the σ-finite measures

$$\mathrm{Q}_t^\beta(A) \coloneqq \int_A q(x, \mathrm{P}_t^\beta)\,\mathrm{d}x\,, \qquad \mathrm{Q}_t^{\beta\uparrow}(A) \coloneqq \int_A q^\uparrow(x, \mathrm{P}_t^\beta)\,\mathrm{d}x\,, \qquad A \in \mathcal{B}(\mathbb{R}^d). \tag{2.51}$$

They are the perturbed versions of the measures Q_t and $\mathrm{Q}_t^\uparrow$ defined in (2.21). The relative entropy of P_t^β with respect to Q_t^β is then given by

$$H\big(\mathrm{P}_t^\beta \,\big|\, \mathrm{Q}_t^\beta\big) = \mathbb{E}_{\mathbb{P}^\beta}\big[\log \ell_t^\beta(X_t, \mathrm{P}_t^\beta)\big] = \mathscr{F}(\mathrm{P}_t^\beta) \tag{2.52}$$

and the relative Fisher information of P_t^β with respect to $\mathrm{Q}_t^{\beta\uparrow}$ equals

$$I\big(\mathrm{P}_t^\beta \,\big|\, \mathrm{Q}_t^{\beta\uparrow}\big) = \mathbb{E}_{\mathbb{P}^\beta}\Big[\big|\nabla \log \ell_t^{\beta\uparrow}(X_t, \mathrm{P}_t^\beta)\big|^2\Big] = \mathscr{D}(\mathrm{P}_t^\beta). \tag{2.53}$$

The following trajectorial result, Theorem 2.4.9 below, provides the semimartingale decomposition of the *perturbed relative entropy process*

$$\log \ell_t^\beta(X_t, \mathrm{P}_t^\beta) = \log p_t^\beta(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t^\beta)(X_t)\,, \qquad t_0 \le t \le T. \tag{2.54}$$

In line with its unperturbed counterpart,Theorem 2.4.1, we shall formulate this result in the reverse direction of time. We first introduce the perturbed analogues of (2.25) and (2.26): the *time-reversed* *perturbed Fisher information process* is defined as

$$\overline{I}_s^\beta := \Big(\big| \nabla \log \overline{\ell}_s^{\beta\downarrow} \big|^2 + \tfrac{1}{2} \big| \nabla (W * \overline{\mathrm{P}}_s^\beta) \big|^2 + \Big\langle \tfrac{1}{2} \nabla (W * \overline{\mathrm{P}}_s^\beta) \,,\, 2 \nabla \log \overline{\ell}_s^{\beta\downarrow} + \nabla V^\beta \Big\rangle \Big) (\overline{X}_s) \tag{2.55}$$

$$- \mathbb{E}_{\tilde{\mathbb{P}}^\beta} \Big[\Big\langle \tfrac{1}{2} \nabla W (\overline{X}_s - \overline{Y}_s) \,,\, \Big(2 \nabla \log \overline{\ell}_s^{\beta\downarrow} - \nabla V + \nabla (W * \overline{\mathrm{P}}_s^\beta) + \nabla \beta \Big) (\overline{Y}_s) \Big\rangle \Big] \tag{2.56}$$

$$+ \Big(\langle \nabla V \,,\, \nabla \beta \rangle - \Delta \beta \Big) (\overline{X}_s) \tag{2.57}$$

for all $0 \leq s \leq T - t_0$, where $(\overline{Y}_s)_{0 \leq s \leq T - t_0}$ is a copy of the process $(\overline{X}_s)_{0 \leq s \leq T - t_0}$ on a copy $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}^\beta)$ of the original probability space $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$; the *perturbed cumulative Fisher information process* is defined as

$$\overline{F}_s^\beta := \int_0^s \overline{I}_u^\beta \, \mathrm{d}u \,, \qquad 0 \leq s \leq T - t_0. \tag{2.58}$$

Theorem 2.4.9. *Suppose* Assumptions 2.4.8 *hold. On* $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$*, the time-reversed perturbed relative entropy process*

$$\log \overline{\ell}_s^\beta (\overline{X}_s, \overline{\mathrm{P}}_s^\beta) = \log \ell_{T-s}^\beta (X_{T-s}, \mathrm{P}_{T-s}^\beta) \,, \qquad 0 \leq s \leq T - t_0 \tag{2.59}$$

admits the semimartingale decomposition

$$\log \overline{\ell}_s^\beta (\overline{X}_s, \overline{\mathrm{P}}_s^\beta) - \log \overline{\ell}_0^\beta (\overline{X}_0, \overline{\mathrm{P}}_0^\beta) = \overline{M}_s^\beta + \overline{F}_s^\beta. \tag{2.60}$$

Here $(\overline{M}_s^\beta)_{0 \leq s \leq T - t_0}$ *is the* $L^2(\mathbb{P}^\beta)$*-bounded martingale*

$$\overline{M}_s^\beta := \int_0^s \Big\langle \nabla \log \overline{\ell}_u^\beta (\overline{X}_u, \overline{\mathrm{P}}_u^\beta) \,,\, \sqrt{2} \, \mathrm{d}\overline{B}_u^\beta \Big\rangle, \tag{2.61}$$

with $(\overline{B}_s^\beta)_{0 \leq s \leq T - t_0}$ *a* $\mathbb{P}^\beta$*-Brownian motion of the backward filtration* $\mathbb{G}$*, and the compensator* (2.58) *satisfies*

$$\mathbb{E}_{\mathbb{P}^\beta}\big[\overline{F}^\beta_s\big] = \int_0^s \bigg(I\big(\overline{\mathrm{P}}^\beta_u \,\big|\, \overline{\mathrm{Q}}^{\beta\uparrow}_u\big) + \mathbb{E}_{\mathbb{P}^\beta}\Big[\Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}^\beta_u)\,,\nabla\beta\Big\rangle - \Delta\beta\Big)(\overline{X}_u)\Big]\bigg)\,\mathrm{d}u < \infty. \tag{2.62}$$

With the dynamics of the time-reversed perturbed relative entropy process at hand, we repeat the same procedure which was carried out for the unperturbed case. Taking expectations with respect to the probability measure $\mathbb{P}^\beta$, we arrive at the perturbed relative entropy identity (2.63), and applying the Lebesgue differentiation theorem gives the perturbed relative entropy production identity (2.64).

Corollary 2.4.10. *Suppose* Assumptions 2.4.8 *hold. For all* $0 \leq t_0 \leq t \leq T$*, we have the perturbed relative entropy identity*

$$H\big(\mathrm{P}^\beta_t \,\big|\, \mathrm{Q}^\beta_t\big) - H\big(\mathrm{P}^\beta_{t_0} \,\big|\, \mathrm{Q}^\beta_{t_0}\big) = -\int_{t_0}^t \bigg(I\big(\mathrm{P}^\beta_u \,\big|\, \mathrm{Q}^{\beta\uparrow}_u\big) + \mathbb{E}_{\mathbb{P}^\beta}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}^\beta_u)\,,\nabla\beta\Big\rangle - \Delta\beta\Big)(X_u)\Big]\bigg)\,\mathrm{d}u. \tag{2.63}$$

For Lebesgue-a.e. $t_0 \in [0,T]$*, the perturbed rate of relative entropy dissipation is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|^+_{t=t_0} H\big(\mathrm{P}^\beta_t \,\big|\, \mathrm{Q}^\beta_t\big) = -\bigg(I\big(\mathrm{P}_{t_0} \,\big|\, \mathrm{Q}_{t_0}\big) + \mathbb{E}_{\mathbb{P}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0})\,,\nabla\beta\Big\rangle - \Delta\beta\Big)(X_{t_0})\Big]\bigg). \tag{2.64}$$

Similarly, we have the following trajectorial rate of relative entropy dissipation for the perturbed diffusion.

Corollary 2.4.11. *Suppose* Assumptions 2.4.8 *hold and* $\int_0^{T-t_0} \mathbb{E}_{\mathbb{P}^\beta}[|\overline{I}^\beta_u|]\,\mathrm{d}u < \infty$*. For* $\mathbb{P}^\beta$*-a.e.* $\omega \in \Omega$ *there exists a Lebesgue null set* $N^\beta_\omega \subseteq [0, T-t_0]$ *such that for any* $s_0 \in [0, T-t_0] \setminus N^\beta_\omega$ *we have*

$$\lim_{s \downarrow s_0} \frac{\mathbb{E}_{\mathbb{P}^\beta}\big[\log \overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}^\beta_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}^\beta_{s_0}\big)}{s - s_0} = \overline{I}^\beta_{s_0}(\omega). \tag{2.65}$$

Proof. The proof proceeds almost verbatim as the proof of Corollary 2.4.6. The only difference is that we now use the semimartingale decomposition (2.60) and the $\mathbb{P}^\beta$-martingale property of the process (2.61) in Theorem 2.4.9. □

We now turn to the computation of the rate of change of the Wasserstein distance along the curve of probability distributions $(\mathrm{P}_t^\beta)_{t_0 \le t \le T}$. To this end, we set

$$v_t^\beta(x) \coloneqq -\big(\nabla \log p_t^\beta + \nabla V^\beta + \nabla(W * p_t^\beta)\big)(x)\,, \qquad (t,x) \in [t_0,T] \times \mathbb{R}^d, \tag{2.66}$$

so that the perturbed granular media equation (2.48) can be viewed as a *continuity equation*

$$\partial_t p_t^\beta(x) + \operatorname{div}\big(v_t^\beta(x)\, p_t^\beta(x)\big) = 0\,, \qquad (t,x) \in (t_0,T) \times \mathbb{R}^d, \tag{2.67}$$

with $v_t^\beta(\,\cdot\,)$ as the corresponding *velocity field*. We recall the definition of the tangent space (see Definition 8.4.1 in [19])

$$\operatorname{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) \coloneqq \overline{\{\nabla\varphi \colon \varphi \in C_c^\infty(\mathbb{R}^d;\mathbb{R})\}}^{L^2(\mu)} \tag{2.68}$$

of $\mathcal{P}_2(\mathbb{R}^d)$ at the point $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, and impose the following additional assumptions.

Assumptions 2.4.12. In addition to Assumptions 2.4.8, we suppose that

$$v_t(\,\cdot\,) \in \operatorname{Tan}_{\mathrm{P}_t} \mathcal{P}_2(\mathbb{R}^d) \qquad \text{for Lebesgue-a.e. } t \in [0,T], \tag{2.69}$$

where $v_t(\,\cdot\,)$ is obtained by taking $\beta \equiv 0$ and $t_0 = 0$ in (2.66).

Remark 2.4.13. For example, we know from [19, Theorem 10.4.13] that the condition (2.69) is satisfied if, in addition to Assumptions 2.4.8, V is uniformly convex, i.e., $\operatorname{Hess}(V) \ge \kappa_V I_n$ for some real constant κ_V, and W is a convex function satisfying the *doubling condition*

$$\exists\, C_W > 0 \ \text{ such that } \forall\, x,y \in \mathbb{R}^d\colon \qquad W(x+y) \le C_W\big(1 + W(x) + W(y)\big). \tag{2.70}$$

The proof of the following result is based on the general theory of Wasserstein metric derivatives of absolutely continuous curves in $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$; for a thorough discussion, we refer to Chapter

8 in [19].

Lemma 2.4.14. *Suppose* Assumptions 2.4.12 *hold. For Lebesgue-a.e.* $t_0 \in [0,T]$*, the Wasserstein metric derivative of the perturbed curve* $(\mathrm{P}_t^\beta)_{t_0 \le t \le T}$ *is equal to*

$$\lim_{t \downarrow t_0} \frac{W_2\big(\mathrm{P}_t^\beta, \mathrm{P}_{t_0}^\beta\big)}{t - t_0} = \big\| v_{t_0}^\beta(X_{t_0}) \big\|_{L^2(\mathbb{P})} = \big\| \nabla \log \ell_{t_0}^{\uparrow}(X_{t_0}, \mathrm{P}_{t_0}) + \nabla \beta(X_{t_0}) \big\|_{L^2(\mathbb{P})}. \tag{2.71}$$

Proof. Without loss of generality we can set $\beta \equiv 0$. Note that from (3.17) we have

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^2 \,\mathrm{d}p_t(x)\,\mathrm{d}t = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla \log \ell_t^{\uparrow}(X_t, \mathrm{P}_t)\big|^2 \,\mathrm{d}t \bigg] < \infty, \tag{2.72}$$

which implies that $v_t(\,\cdot\,) \in L^2(\mathrm{P}_t)$ for Lebesgue-a.e. $t \in [0,T]$. Therefore we can apply Theorem 8.3.1 and Proposition 8.4.5 of [19] to the absolutely continuous curve $(\mathrm{P}_t)_{0 \le t \le T}$, which yields (2.71). $\square$

We now have all the ingredients to formulate the gradient flow property of the granular media equation. The Wasserstein metric slope of the free energy functional $\mathscr{F}$ along the McKean–Vlasov curve $(\mathrm{P}_t)_{t_0 \le t \le T}$ is defined as

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}) := \lim_{t \downarrow t_0} \frac{H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0})}{\mathcal{W}_2(\mathrm{P}_t, \mathrm{P}_{t_0})}. \tag{2.73}$$

In order to show that this is the slope of steepest descent, we will compare it with the slope

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}\big(\mathrm{P}_{t_0}^\beta\big) := \lim_{t \downarrow t_0} \frac{H\big(\mathrm{P}_t^\beta \,|\, \mathrm{Q}_t^\beta\big) - H\big(\mathrm{P}_{t_0}^\beta \,|\, \mathrm{Q}_{t_0}^\beta\big)}{\mathcal{W}_2\big(\mathrm{P}_t^\beta, \mathrm{P}_{t_0}^\beta\big)} \tag{2.74}$$

along the perturbed curve $(\mathrm{P}_t^\beta)_{t_0 \le t \le T}$.

Theorem 2.4.15. *Suppose* Assumptions 2.4.12 *hold. The following assertions hold for Lebesgue-a.e.* $t_0 \in [0,T]$*: The random variables*

$$\mathfrak{a} := \nabla \log \ell_{t_0}^{\uparrow}(X_{t_0}, \mathrm{P}_{t_0}) \qquad \text{and} \qquad \mathfrak{b} := \nabla \beta(X_{t_0}) \tag{2.75}$$

are elements of $L^2(\mathbb{P})$*, and the Wasserstein metric slope of the free energy functional* $\mathscr{F}$ *along the McKean–Vlasov curve* $(\mathrm{P}_t)_{t_0 \leq t \leq T}$ *is given by*

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}) = -\|\mathfrak{a}\|_{L^2(\mathbb{P})}. \tag{2.76}$$

If $\mathfrak{a} + \mathfrak{b} \neq 0$*, the metric slope along the perturbed curve* $(\mathrm{P}_t^{\beta})_{t_0 \leq t \leq T}$ *is equal to*

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}^{\beta}) = -\left\langle \mathfrak{a}\,,\, \frac{\mathfrak{a}+\mathfrak{b}}{\|\mathfrak{a}+\mathfrak{b}\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \tag{2.77}$$

In particular,

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}) \leq \big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}^{\beta}) \tag{2.78}$$

with equality if and only if $\mathfrak{a} + \mathfrak{b}$ *is a positive multiple of* $\mathfrak{a}$*.*

Proof. The equality (2.76) follows from (2.41) and by taking $\beta \equiv 0$ in (2.71). For the proof of (2.77), we first observe that from (2.64) and (2.71) we obtain the equality

$$\big|\partial \mathscr{F}\big|_{\mathcal{W}_2}(\mathrm{P}_{t_0}^{\beta}) = -\frac{\|\mathfrak{a}\|^2_{L^2(\mathbb{P})} + \mathbb{E}_{\mathbb{P}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0})\,,\, \nabla\beta \Big\rangle - \Delta\beta\Big)(X_{t_0})\Big]}{\|\mathfrak{a}+\mathfrak{b}\|_{L^2(\mathbb{P})}} \tag{2.79}$$

for Lebesgue-a.e. $t_0 \in [0,T]$. Integrating by parts and recalling the notations in (2.18) – (2.20), we find that the expectation in the numerator of (2.79) is equal to

$$\int_{\mathbb{R}^d} \Big\langle \log \nabla \ell_{t_0}^{\uparrow}(x)\,,\, \nabla\beta(x) \Big\rangle\, p_{t_0}(x)\, \mathrm{d}x = \big\langle \mathfrak{a}\,,\, \mathfrak{b} \big\rangle_{L^2(\mathbb{P})}. \tag{2.80}$$

Now (2.78) follows by the Cauchy–Schwarz inequality. □

2.4.3 A trajectorial proof of the HWBI inequality

In this section, we show how our trajectorial approach can be adapted to give a simple proof of the HWBI inequality. While the techniques that will be used are similar, the setting of this section is independent from the rest of the paper. In particular, we shall impose convexity assumptions on the potentials V, W.

We fix two probability measures ν_0 and ν_1 in $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$. By Brenier's theorem [97], there exists a convex function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ such that

$$\mathcal{W}_2^2(\nu_0, \nu_1) = \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 \,\mathrm{d}\nu_0(x). \tag{2.81}$$

The displacement interpolation of McCann [98] between ν_0 and ν_1 is given by

$$\nu_t \coloneqq (T_t)_{\#}\nu_0\,, \qquad T_t(x) \coloneqq (1-t)x + t\nabla\varphi(x)\,, \qquad 0 \le t \le 1. \tag{2.82}$$

In particular, since the endpoints ν_0 and ν_1 belong to $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$, each ν_t has a probability density function ρ_t; see, e.g., [99, Remarks 5.13 (i)].

As before, we consider a confinement potential V and an interaction potential W. For each $t \in [0,1]$, we then define by analogy with (2.21), the σ-finite measures

$$\mu_t(A) \coloneqq \int_A q(x, \nu_t) \,\mathrm{d}x\,, \qquad \mu_t^{\uparrow}(A) \coloneqq \int_A q^{\uparrow}(x, \nu_t) \,\mathrm{d}x\,, \qquad A \in \mathcal{B}(\mathbb{R}^d), \tag{2.83}$$

where we recall the definitions of the density functions q and $q^{\uparrow}$ in (2.19). In parallel to the likelihood ratio functions in (2.19), we define

$$r_t(x,\nu) \coloneqq \frac{\rho_t(x)}{q(x,\nu)}\,, \qquad r_t^{\uparrow}(x,\nu) \coloneqq \frac{\rho_t(x)}{q^{\uparrow}(x,\nu)}\,, \qquad (t,x,\nu) \in [0,1] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d). \tag{2.84}$$

Then the relative entropy of ν_t with respect to μ_t is given by

$$H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^d} \rho_t(x) \log r_t(x, \nu_t) \,\mathrm{d}x \tag{2.85}$$

and the relative Fisher information of ν_t with respect to $\mu_t^{\uparrow}$ is equal to

$$I\big(\nu_t \,|\, \mu_t^{\uparrow}\big) = \int_{\mathbb{R}^d} |\nabla \log r_t^{\uparrow}(x, \nu_t)|^2 \, \rho_t(x) \, \mathrm{d}x. \tag{2.86}$$

We impose the following regularity conditions for Proposition 2.4.17, noting that the strong assumptions placed on ρ_0 and ρ_1 are only temporary and will be removed in Assumptions 2.4.18 of Theorem 2.4.19.

Assumptions 2.4.16. The functions $V, W \colon \mathbb{R}^d \to [0, \infty)$ are smooth and W is symmetric. The probability density functions ρ_0 and ρ_1 are smooth, compactly supported and strictly positive in the interior of their respective supports.

Proposition 2.4.17. *Suppose* Assumptions 2.4.16 *hold. Along the displacement interpolation* $(\nu_t)_{0 \le t \le 1}$*, the rate of relative entropy dissipation at time* $t = 0$*, with respect to the "reference curve of probability measures"* $(\mu_t)_{0 \le t \le 1}$*, is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}^{+} H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^d} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \,, \nabla \varphi(x) - x \Big\rangle \, \rho_0(x) \, \mathrm{d}x. \tag{2.87}$$

Combining Proposition 2.4.17 with the displacement convexity results of McCann [98], we obtain the following generalization of the HWBI inequality. Equivalent versions of this inequality can be found in [85, Theorem 4.1] and [86, Theorem D.50].

Assumptions 2.4.18. The functions $V, W \colon \mathbb{R}^d \to [0, \infty)$ are smooth and W is symmetric. Furthermore, V and W are uniformly convex, i.e., there exist real constants κ_V and κ_W such that

$$\mathrm{Hess}(V) \geq \kappa_V I_n \,, \qquad \mathrm{Hess}(W) \geq \kappa_W I_n. \tag{2.88}$$

Theorem 2.4.19. *Suppose* Assumptions 2.4.18 *hold and the relative entropy* $H(\nu_1 \,|\, \mu_1)$ *is finite.*

Then

$$H(\nu_0 \,|\, \mu_0) - H(\nu_1 \,|\, \mu_1) \leq - \int_{\mathbb{R}^d} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \,, \nabla\varphi(x) - x \Big\rangle \, \rho_0(x) \, \mathrm{d}x \tag{2.89}$$

$$- \tfrac{\kappa_V + \kappa_W}{2} \, \mathcal{W}_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2} \, |b(\nu_0) - b(\nu_1)|^2. \tag{2.90}$$

Remark 2.4.20. By the Cauchy–Schwarz inequality, the right-hand side of (2.89) can be bounded from above by

$$\sqrt{\int_{\mathbb{R}^d} \big|\nabla \log r_0^{\uparrow}(x, \nu_0)\big|^2 \, \rho_0(x) \, \mathrm{d}x} \, \sqrt{\int_{\mathbb{R}^d} |\nabla\varphi(x) - x|^2 \, \rho_0(x) \, \mathrm{d}x} = \sqrt{I(\nu_0 \,|\, \mu_0^{\uparrow})} \, \mathcal{W}_2(\nu_0, \nu_1), \tag{2.91}$$

and we obtain the usual form of the HWBI inequality (2.15); see also [20, Theorem 4.2].

2.5 Proofs of the main results

This section is devoted to the proofs of the results stated in Chapter 2.4. We shall first prove the main trajectorial results: Theorem 2.4.1 and its "perturbed" counterpart, Theorem 2.4.9.

2.5.1 The proofs of Theorems 2.4.1 and 2.4.9

Since Theorem 2.4.1 follows immediately from Theorem 2.4.9 by setting the perturbation $\beta \colon \mathbb{R}^d \to \mathbb{R}$ to be the zero function, we start with the general setting of Theorem 2.4.9. We first recall a classical result concerning the time reversal of diffusions.

Lemma 2.5.1 ([100, Theorem 2.1], [95, Theorems G.2, G.5])**.** *Suppose Assumptions 2.4.8 hold. On* $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$*, the process*

$$\overline{B}_s^\beta \coloneqq B_{T-s}^\beta - B_T^\beta - \sqrt{2} \int_0^s \nabla \log \overline{p}_u^\beta(\overline{X}_u^\beta) \, \mathrm{d}u \,, \qquad 0 \leq s \leq T - t_0 \tag{2.92}$$

is a Brownian motion. Moreover, the time-reversed canonical process $(\overline{X}_s)_{0\leq s\leq T-t_0}$ *satisfies*

$$\mathrm{d}\overline{X}_s = \Big(2\nabla \log \overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W * \overline{\mathrm{P}}^{\beta}_s) + \nabla\beta\Big)(\overline{X}_s)\,\mathrm{d}s + \sqrt{2}\,\mathrm{d}\overline{B}^{\beta}_s. \tag{2.93}$$

By means of Lemma 2.5.1, the first step in the proof of Theorem 2.4.9 is to compute the dynamics of the time-reversed perturbed relative entropy process (2.59). For the reader's convenience, we recall the following characterization of the L-derivative in [78, pp. 383].

Definition 2.5.2. Let $f\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. On a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, let X_0 be a random variable with distribution μ_0. We define $\partial_\mu f(\mu_0)\colon \mathbb{R}^d \to \mathbb{R}^d$ as the L-*derivative* of f at μ_0, if for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and any random variable X with distribution μ,

$$f(\mu) = f(\mu_0) + \mathbb{E}_{\mathbb{P}}\Big[\big\langle \partial_\mu f(\mu_0)(X_0)\,,\, X - X_0\big\rangle\Big] + o\big(\|X - X_0\|_{L^2(\mathbb{P})}\big).$$

Remark 2.5.3. The above characterization of the L-derivative depends neither on the choice of the probability space $(\Omega, \mathbb{F}, \mathbb{P})$, nor of the random variables X and X_0 used to represent μ and μ_0, respectively. Moreover, if the L-derivative exists, it is uniquely defined up to μ_0-equivalence. We refer to Proposition 5.25 and Remark 5.26 in [78] for the details.

Proposition 2.5.4. *Suppose Assumptions 2.4.8 hold. On* $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$*, the time-reversed perturbed relative entropy process* (2.59) *satisfies*

$$\mathrm{d}\log \overline{\ell}^{\beta}_s(\overline{X}_s, \overline{\mathrm{P}}^{\beta}_s) = \Big\langle \nabla \log \overline{\ell}^{\beta}_s(\overline{X}_s, \overline{\mathrm{P}}^{\beta}_s)\,,\, \sqrt{2}\,\mathrm{d}\overline{B}^{\beta}_s\Big\rangle \tag{2.94}$$

$$+ \Big(\big|\nabla \log \overline{\ell}^{\beta\downarrow}_s\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}^{\beta}_s)\big|^2\Big)(\overline{X}_s)\,\mathrm{d}s \tag{2.95}$$

$$+ \Big(\Big\langle \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}^{\beta}_s)\,,\, 2\nabla \log \overline{\ell}^{\beta\downarrow}_s + \nabla V^{\beta}\Big\rangle + \langle \nabla V\,,\, \nabla\beta\rangle - \Delta\beta\Big)(\overline{X}_s)\,\mathrm{d}s \tag{2.96}$$

$$- \mathbb{E}_{\tilde{\mathbb{P}}^{\beta}}\Big[\Big\langle \tfrac{1}{2}\nabla W(\overline{X}_s - \overline{Y}_s)\,,\, \Big(2\nabla \log \overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W * \overline{\mathrm{P}}^{\beta}_s) + \nabla\beta\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s, \tag{2.97}$$

where $(\overline{Y}_s)_{0\leq s\leq T-t_0}$ *is a copy of the process* $(\overline{X}_s)_{0\leq s\leq T-t_0}$ *on a copy* $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}^{\beta})$ *of the original* *probability space* $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$.

Proof. Applying a generalized version of Itô's formula for McKean–Vlasov diffusions [78, Proposition 5.102] and using the backward dynamics in (2.93), we obtain

$$\mathrm{d}\log\overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s) = \Big\langle \nabla\log\overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s)\,,\, \sqrt{2}\,\mathrm{d}\overline{B}^\beta_s \Big\rangle + \Big(\partial_s \log\overline{\ell}^\beta_s + \Delta\log\overline{\ell}^\beta_s\Big)(\overline{X}_s, \overline{\mathrm{P}}^\beta_s)\,\mathrm{d}s \tag{2.98}$$

$$+ \Big\langle \nabla\log\overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s)\,, \Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W * \overline{\mathrm{P}}^\beta_s) + \nabla\beta\Big)(\overline{X}_s)\Big\rangle\,\mathrm{d}s \tag{2.99}$$

$$+ \mathbb{E}_{\tilde{\mathbb{P}}^\beta}\Big[\Big\langle\Big(\partial_\mu \log\overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s)\Big)\,, \Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W * \overline{\mathrm{P}}^\beta_s) + \nabla\beta\Big)\Big\rangle(\overline{Y}_s)\Big]\,\mathrm{d}s \tag{2.100}$$

$$+ \mathbb{E}_{\tilde{\mathbb{P}}^\beta}\Big[\mathrm{trace}\Big(\partial_y\partial_\mu \log\overline{\ell}^\beta_s(\overline{X}_s, \overline{\mathrm{P}}^\beta_s)(\overline{Y}_s)\Big)\Big]\,\mathrm{d}s, \tag{2.101}$$

where $(\overline{Y}_s)_{0\le s\le T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T-t_0}$ on a copy $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}^\beta)$ of the original probability space $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$. The L-derivative appearing in (2.100) and (2.101) is calculated to be

$$\Big(\partial_\mu \log\overline{\ell}^\beta_s(x,\mu)\Big)(y) = \tfrac{1}{2}\big(\partial_\mu\left(W*\mu\right)(x)\big)(y) = -\tfrac{1}{2}\nabla W(x-y) \tag{2.102}$$

for $(x,\mu,y) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$, see [78, Section 5.2.2, Example 1] for the computation of the L-derivative of a function which is linear in the distribution variable. Consequently, we have

$$\mathrm{trace}\Big(\partial_y\partial_\mu \log\overline{\ell}^\beta_s(x,\mu)(y)\Big) = -\tfrac{1}{2}\mathrm{trace}\big(\partial_y\nabla W(x-y)\big) = \tfrac{1}{2}\Delta W(x-y). \tag{2.103}$$

Putting (2.102) and (2.103) into (2.100) and (2.101), respectively, as well as using the identities

$$\partial_s \log\overline{\ell}^\beta_s(x,\mu) = \partial_s \log\overline{\ell}^{\beta\downarrow}_s(x), \tag{2.104}$$

$$\nabla\log\overline{\ell}^\beta_s(x,\mu) = \nabla\log\overline{\ell}^{\beta\downarrow}_s(x) + \tfrac{1}{2}\nabla(W*\mu)(x), \tag{2.105}$$

$$\Delta\log\overline{\ell}^\beta_s(x,\mu) = \Delta\log\overline{\ell}^{\beta\downarrow}_s(x) + \tfrac{1}{2}\Delta(W*\mu)(x), \tag{2.106}$$

we obtain

$$\mathrm{d}\log\overline{\ell}^{\beta}_s(\overline{X}_s,\overline{\mathrm{P}}^{\beta}_s) = \Big\langle \nabla\log\overline{\ell}^{\beta}_s(\overline{X}_s,\overline{\mathrm{P}}^{\beta}_s)\,,\,\sqrt{2}\,\mathrm{d}\overline{B}^{\beta}_s\Big\rangle + \Big(\partial_s\log\overline{\ell}^{\beta\downarrow}_s + \Delta\log\overline{\ell}^{\beta\downarrow}_s\Big)(\overline{X}_s)\,\mathrm{d}s \tag{2.107}$$

$$+\Big\langle\Big(\nabla\log\overline{\ell}^{\beta\downarrow}_s + \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}^{\beta}_s)\Big)\,,\Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}^{\beta}_s) + \nabla\beta\Big)\Big\rangle(\overline{X}_s)\,\mathrm{d}s \tag{2.108}$$

$$-\,\mathbb{E}_{\tilde{\mathbb{P}}^{\beta}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s)\,,\Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}^{\beta}_s) + \nabla\beta\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s \tag{2.109}$$

$$+\,\tfrac{1}{2}\Big(\Delta(W*\overline{\mathrm{P}}^{\beta}_s)(\overline{X}_s) + \mathbb{E}_{\tilde{\mathbb{P}}^{\beta}}\Big[\Delta W(\overline{X}_s-\overline{Y}_s)\Big]\Big)\,\mathrm{d}s. \tag{2.110}$$

Regarding the expression of (2.110), we observe that $\Delta(W*\overline{\mathrm{P}}^{\beta}_s)(\overline{X}_s) = \mathbb{E}_{\tilde{\mathbb{P}}^{\beta}}[\Delta W(\overline{X}_s-\overline{Y}_s)]$. Finally, elementary computations based on (2.19), (2.48) and (2.50) show that the perturbed log-likelihood ratio function $(s,x)\mapsto\log\overline{\ell}^{\beta\downarrow}_s(x)$ of (2.50) satisfies

$$\begin{aligned}\partial_s\log\overline{\ell}^{\beta\downarrow}_s = \Big\langle\nabla\log\overline{\ell}^{\beta\downarrow}_s\,,\nabla V - \nabla(W*\overline{\mathrm{P}}^{\beta}_s) - \nabla\beta\Big\rangle - \big|\nabla\log\overline{\ell}^{\beta\downarrow}_s\big|^2 - \Delta\log\overline{\ell}^{\beta\downarrow}_s \\ +\Big\langle\nabla V\,,\nabla(W*\overline{\mathrm{P}}^{\beta}_s) + \nabla\beta\Big\rangle - \Delta(W*\overline{\mathrm{P}}^{\beta}_s) - \Delta\beta\end{aligned} \tag{2.111}$$

on $(0,T-t_0)\times\mathbb{R}^d$, with terminal condition $\log\overline{\ell}^{\beta\downarrow}_{T-t_0} = \log\overline{\ell}^{\downarrow}_{T-t_0}$. Inserting (2.111) into (2.107), we obtain (2.94) – (2.97). □

Setting the perturbation β to be the zero function, we obtain the following result.

Corollary 2.5.5. *Suppose Assumptions 2.3.1 hold. On* $(\Omega,\mathbb{G},\mathbb{P})$*, the time-reversed relative entropy process* (2.27) *satisfies*

$$\mathrm{d}\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s) = \Big\langle\nabla\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s)\,,\,\sqrt{2}\,\mathrm{d}\overline{B}_s\Big\rangle \tag{2.112}$$

$$+\Big(\big|\nabla\log\overline{\ell}^{\downarrow}_s\big|^2 + \tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_s)\big|^2 + \Big\langle\tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_s)\,,\,2\nabla\log\overline{\ell}^{\downarrow}_s + \nabla V\Big\rangle\Big)(\overline{X}_s)\,\mathrm{d}s \tag{2.113}$$

$$-\,\mathbb{E}_{\tilde{\mathbb{P}}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s)\,,\Big(2\nabla\log\overline{\ell}^{\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}_s)\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s. \tag{2.114}$$

Here, the process

$$\overline{B}_s \coloneqq B_{T-s} - B_T - \sqrt{2}\int_0^s \nabla \log \overline{p}_u(\overline{X}_u)\,\mathrm{d}u\,, \qquad 0 \le s \le T \tag{2.115}$$

is a $\mathbb{P}$-Brownian motion with respect to the backward filtration $\mathbb{G}$, and $(\overline{Y}_s)_{0\le s\le T}$ is a copy of the process $(\overline{X}_s)_{0\le s\le T}$ on a copy $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}})$ of the original probability space $(\Omega, \mathbb{G}, \mathbb{P})$.

Before turning to the final part of the proof of Theorem 2.4.1, we state a classical result based on the general theory of the Cameron–Martin–Maruyama–Girsanov transformation [101]. The connection between relative entropy (the left-hand side of (2.117) below) and energy (the right-hand side of (2.117)) is the foundation of Föllmers entropy approach to the time reversal of diffusion processes on Wiener space [102, 103, 104]. We denote by $\mathbb{W}_x$ the Wiener measure on $\Omega = C([0,T];\mathbb{R}^d)$ with starting point $x \in \mathbb{R}^d$, and define by

$$\mathbb{W}_{x,2}(A) \coloneqq \mathbb{W}_x\Big(\omega \in \Omega\colon (\sqrt{2}X)(\omega) \in A\Big)\,, \qquad A \in \mathcal{B}(\Omega) \tag{2.116}$$

the Wiener measure with starting point x and variance 2.

Lemma 2.5.6. *The relative entropy of $\mathbb{P}$ with respect to $\mathbb{W}_{\mathrm{P}_0,2} \coloneqq \int_{\mathbb{R}^d} \mathbb{W}_{x,2}\,\mathrm{P}_0(\mathrm{d}x)$ is given by*

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = \frac{1}{4}\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla V(X_t) + \nabla(W * \mathrm{P}_t)(X_t)\big|^2\,\mathrm{d}t\bigg] < \infty. \tag{2.117}$$

Proof. Recalling (2.18), the drift of the McKean–Vlasov dynamics (2.1) can be expressed as

$$-\nabla\Psi^{\uparrow}(x, \mathrm{P}_t) = -\big(\nabla V(x) + \nabla(W * \mathrm{P}_t)(x)\big)\,, \qquad (t,x) \in [0,T]\times\mathbb{R}^d. \tag{2.118}$$

For any $t \in [0,T]$, using the elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$, we have

$$\mathbb{E}_{\mathbb{P}}\Big[|\nabla\Psi^{\uparrow}(X_t, \mathrm{P}_t)|^2\Big] \le 2\,\mathbb{E}_{\mathbb{P}}\Big[|\nabla V(X_t)|^2\Big] + 2\,\mathbb{E}_{\mathbb{P}}\Big[|\nabla(W * \mathrm{P}_t)(X_t)|^2\Big]. \tag{2.119}$$

Using the linear growth condition (2.16) from Assumptions 2.3.1 (i), we find

$$\mathbb{E}_{\mathbb{P}}\Big[|\nabla V(X_t)|^2\Big] \leq 2C^2\Big(1+\mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \leq 2C^2\Big(1+\mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leq t\leq T}|X_t|^2\Big]\Big). \tag{2.120}$$

Similarly, by Jensen's inequality and (2.16), we obtain

$$\mathbb{E}_{\mathbb{P}}\Big[|\nabla(W*\mathrm{P}_t)(X_t)|^2\Big] \leq \int_{\mathbb{R}^d\times\mathbb{R}^d}|\nabla W(x-y)|^2\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x \tag{2.121}$$

$$\leq 2C^2\bigg(1+\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) \tag{2.122}$$

$$\leq 2C^2\bigg(1+2\int_{\mathbb{R}^d\times\mathbb{R}^d}(|x|^2+|y|^2)\,p_t(y)\,p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) \tag{2.123}$$

$$= 2C^2\Big(1+4\,\mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \leq 8C^2\Big(1+\mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leq t\leq T}|X_t|^2\Big]\Big). \tag{2.124}$$

Altogether, we get

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T|\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\,\mathrm{d}t\bigg] \leq 20T\Big(1+\mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leq t\leq T}|X_t|^2\Big]\Big) < \infty, \tag{2.125}$$

where the finiteness of this expression follows from the uniform second moment property (2.17) of Lemma 2.3.2. From [101, Section 7.6.4] we now conclude that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{W}_{\mathrm{P}_0,2}$, and the Radon–Nikodym derivatives are given by

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}_{\mathrm{P}_0,2}}\bigg|_{\mathcal{F}_t} = \exp\left(-\frac{1}{2}\int_0^t\big\langle\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u)\,,\,\sqrt{2}\,\mathrm{d}B_u\big\rangle+\frac{1}{4}\int_0^t|\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u)|^2\,\mathrm{d}u\right), \qquad 0\leq t\leq T. \tag{2.126}$$

The integrability property (2.125) implies that the $\mathbb{P}$-expectation of the stochastic integral in (2.126) vanishes, and we obtain

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = \mathbb{E}_{\mathbb{P}}\bigg[\log\bigg(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}_{\mathrm{P}_0,2}}\bigg)\bigg] = \frac{1}{4}\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T|\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\,\mathrm{d}t\bigg] < \infty, \tag{2.127}$$

which shows (2.117). $\square$

Denoting n-dimensional Lebesgue measure by λ, we consider on $\Omega = C([0,T];\mathbb{R}^d)$ the σ-finite measure $\mathbb{W}_{\lambda,2} := \int_{\mathbb{R}^d} \mathbb{W}_{x,2}\, \lambda(\mathrm{d}x)$, which is known as the law of the *reversible Brownian motion* on $\mathbb{R}^d$ with variance 2; see [105, 96]. The fundamental property of reversible Brownian motion is that it is invariant under time reversal. This property can be formalized as follows. Let $R\colon \Omega \to \Omega$ be the pathwise time reversal operator on Ω, given by $X_s \circ R = X_{T-s}$ for $s \in [0,T]$. For any measure μ on Ω, we denote its time reversal by $\overline{\mu} := R_{\#}\mu$. Then we have the invariance property $\overline{\mathbb{W}}_{\lambda,2} = \mathbb{W}_{\lambda,2}$. Let us also consider the probability measure $\mathbb{W}_{\mathrm{P}_T,2} := \int_{\mathbb{R}^d} \mathbb{W}_{x,2}\, \mathrm{P}_T(\mathrm{d}x)$ and its time reversal given by $\overline{\mathbb{W}}_{\mathrm{P}_T,2} = \int_{\mathbb{R}^d} \overline{\mathbb{W}}_{x,2}\, \mathrm{P}_T(\mathrm{d}x)$. Then, as already noted in [102, Remarks 3.7], we have the following result.

Lemma 2.5.7. *We have the relative entropy relations*

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}\big) = H\big(\mathrm{P}_0\,|\,\lambda\big) + H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) \tag{2.128}$$

and

$$H\big(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\lambda,2}\big) = H\big(\mathrm{P}_T\,|\,\lambda\big) + H\big(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\mathrm{P}_T,2}\big). \tag{2.129}$$

Furthermore, all these relative entropies are finite.

Proof. For any $x \in \mathbb{R}^d$, we let $\mathbb{P}_x(\,\cdot\,) := \mathbb{P}(\,\cdot\,|\,X_0 = x)$ denote (a version of) the conditional probability measure $\mathbb{P}$ given $X_0 = x$. By the chain rule for relative entropy [96, Theorem 2] we have

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}\big) = H\big(\mathrm{P}_0\,|\,\lambda\big) + \int_{\mathbb{R}^d} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{d}\mathrm{P}_0(x) \tag{2.130}$$

and at the same time

$$H\big(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}\big) = H\big(\mathrm{P}_0\,|\,\mathrm{P}_0\big) + \int_{\mathbb{R}^d} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{d}\mathrm{P}_0(x) = \int_{\mathbb{R}^d} H\big(\mathbb{P}_x\,|\,\mathbb{W}_{x,2}\big)\,\mathrm{d}\mathrm{P}_0(x), \tag{2.131}$$

implying the first identity (2.128). Regarding the finite entropy assertions, we recall (2.2), (2.3) and observe that

$$H\big(\mathrm{P}_0 \,|\, \lambda\big) = \int_{\mathbb{R}^d} p_0(x) \log p_0(x) \,\mathrm{d}x = \mathscr{U}(\mathrm{P}_0) \leq \mathscr{F}(\mathrm{P}_0) < \infty, \tag{2.132}$$

where the finiteness follows from Assumptions (i) (ii). Furthermore, from Lemma 2.5.6 we know that $H(\mathbb{P} \,|\, \mathbb{W}_{\mathrm{P}_0,2}) < \infty$.

By the same arguments as above, (2.129) follows again by the chain rule for relative entropy. Using the invariance property $\overline{\mathbb{W}}_{\lambda,2} = \mathbb{W}_{\lambda,2}$, and as we already know that $H(\mathbb{P} \,|\, \mathbb{W}_{\lambda,2}) < \infty$, it follows that $H(\mathbb{P} \,|\, \overline{\mathbb{W}}_{\lambda,2}) < \infty$. Let us recall now that $\mathrm{P}_T \in \mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$ by Lemma 2.3.2. On the one hand, since P_T has finite second moment, $H(\mathrm{P}_T \,|\, \lambda)$ cannot take the value $-\infty$ as noted in Remark 2.3.3. On the other hand, the absolute continuity of P_T implies that $H(\mathrm{P}_T \,|\, \lambda)$ cannot take the value $+\infty$. Therefore, we conclude that $H(\mathbb{P} \,|\, \overline{\mathbb{W}}_{\mathrm{P}_T,2}) < \infty$. □

We have assembled now all the ingredients needed for the proof of Theorem 2.4.1.

Proof of Theorem 2.4.1. Recalling the definition of the stochastic integral process $(\overline{M}_s)_{0 \leq s \leq T}$ in (2.29) and of the cumulative Fisher information process $(\overline{F}_s)_{0 \leq s \leq T}$ in (2.26), we see that the stochastic differential of (2.112) – (2.114) can be expressed as claimed in (2.28).

Since $(\overline{M}_s)_{0 \leq s \leq T}$ is a stochastic integral process, it is a continuous local martingale. In order to show that it is an $L^2(\mathbb{P})$-bounded martingale, it suffices to show the integrability condition

$$\mathbb{E}_{\mathbb{P}}\Big[\big\langle \overline{M}, \overline{M} \big\rangle_T\Big] = \mathbb{E}_{\mathbb{P}}\bigg[2 \int_0^T \big|\nabla \log \overline{\ell}_u(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2 \,\mathrm{d}u\bigg] < \infty; \tag{2.133}$$

see, e.g. [106, Corollary IV.1.25]. On $(\Omega, \mathbb{G}, \mathbb{P})$, the time-reversed canonical process $(\overline{X}_s)_{0 \leq s \leq T}$ has backward dynamics

$$\mathrm{d}\overline{X}_s = \overline{\vartheta}_s(\overline{X}_s) \,\mathrm{d}s + \sqrt{2}\,\mathrm{d}\overline{B}_s\,, \qquad 0 \leq s \leq T \tag{2.134}$$

with initial distribution P_T, where the drift term is given by

$$\overline{\vartheta}_s(x) \coloneqq \Big(2\nabla \log \overline{\ell}_s^{\downarrow} - \nabla V + \nabla(W * \overline{\mathrm{P}}_s)\Big)(x) = 2\nabla \log \overline{\ell}_s(x, \overline{\mathrm{P}}_s) - \nabla V(x) \in \mathbb{R}^d \tag{2.135}$$

for $(s,x) \in [0,T] \times \mathbb{R}^d$. Therefore, in order to prove (2.133), it suffices to show the two integrability conditions

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\nabla V(\overline{X}_u)|^2 \,\mathrm{d}u\bigg] < \infty \qquad \text{and} \qquad \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T |\overline{\vartheta}_u(\overline{X}_u)|^2 \,\mathrm{d}u\bigg] < \infty. \tag{2.136}$$

The first condition is a direct consequence of (2.120). From [102, Lemma 2.6] we conclude that the expectation of the second condition is bounded by the relative entropy $H(\mathbb{P} \,|\, \overline{\mathbb{W}}_{\mathrm{P}_T,2})$, which is finite on account of Lemma 2.5.7.

In order to complete the proof of Theorem 2.4.1, it remains to show (3.17). To begin with, we take expectation with respect to $\mathbb{P}$ in (2.26) and invoke Fubini's theorem to interchange the $\mathbb{P}$-expectation and the time integral. Applying once more Fubini's theorem, we swap the $\mathbb{P}$-expectation with the $\tilde{\mathbb{P}}$-expectation appearing in (2.25). Next, we recall Assumptions 2.3.1 (i) and use the symmetry of the interaction potential, which implies that $\nabla W(-x) = -\nabla W(x)$ for all $x \in \mathbb{R}^d$. Furthermore, as the distribution of $\overline{Y}_u$ under $\tilde{\mathbb{P}}$ is the same as the distribution of $\overline{X}_u$ under $\mathbb{P}$, we deduce that

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\Big(\big|\nabla \log \overline{\ell}_u^{\downarrow}\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}_u)\big|^2 + \Big\langle \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_u)\,,\, 2\nabla \log \overline{\ell}_u^{\downarrow} + \nabla V\Big\rangle\Big)(\overline{X}_u)\Big] \,\mathrm{d}u \tag{2.137}$$

$$+ \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_u)\,,\, \Big(2\nabla \log \overline{\ell}_u^{\downarrow} - \nabla V + \nabla(W * \overline{\mathrm{P}}_u)\Big)(\overline{X}_u)\Big\rangle\Big] \,\mathrm{d}u \tag{2.138}$$

for $0 \leq s \leq T$. Recalling the definitions in (2.18) – (2.20), we obtain

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2\Big] \,\mathrm{d}u = \int_0^s I\big(\overline{\mathrm{P}}_u \,\big|\, \overline{\mathrm{Q}}_u^{\uparrow}\big) \,\mathrm{d}u < \infty\,, \qquad 0 \leq s \leq T, \tag{2.139}$$

where the second equality is immediate from (2.22), and the finiteness of the expression in (2.139) is justified as follows. Again, from (2.18) – (2.20) we find

$$\big|\nabla \log \overline{\ell}^{\uparrow}_s(\overline{X}_s, \overline{\mathrm{P}}_s)\big|^2 = \big|\nabla \log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) + \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2 \tag{2.140}$$

$$\leq 2\big|\nabla \log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s)\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2. \tag{2.141}$$

In light of (2.133) and (2.121) – (2.124), we see that the expression in (2.139) is finite, which in turn justifies a posteriori the former applications of Fubini's theorem. □

The proof of Theorem 2.4.9 is now an easy consequence.

Proof of Theorem 2.4.9 . Recalling the definition of the process $(\overline{M}^{\beta}_s)_{0 \leq s \leq T}$ in (2.61) and of the perturbed cumulative Fisher information process $(\overline{F}^{\beta}_s)_{0 \leq s \leq T}$ in (2.58), we see that the stochastic differential of (2.94) – (2.97) can be expressed as claimed in (2.60).

As in the proof of Theorem 2.4.1, we will now argue that

$$\mathbb{E}_{\mathbb{P}^\beta}\Big[\big\langle \overline{M}^{\beta}, \overline{M}^{\beta} \big\rangle_{T-t_0}\Big] = \mathbb{E}_{\mathbb{P}^\beta}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}^{\beta}_u(\overline{X}_u, \overline{\mathrm{P}}^{\beta}_u)\big|^2 \,\mathrm{d}u\bigg] < \infty, \tag{2.142}$$

which will then imply that the stochastic integral process $(\overline{M}^{\beta}_s)_{0 \leq s \leq T-t_0}$ is an $L^2(\mathbb{P}^\beta)$-bounded martingale. To this end, we define the density $q^\beta(x,\mu) \coloneqq \mathrm{e}^{-\Psi^\beta(x,\mu)}$ for $(x,\mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, and consider the "doubly perturbed" likelihood ratio function

$$\ell^{\beta,\beta}_t(x,\mu) \coloneqq \frac{p^{\beta}_t(x)}{q^\beta(x,\mu)}\,, \qquad (t,x) \in [t_0, T] \times \mathbb{R}^d. \tag{2.143}$$

As the Assumptions 2.3.1 are invariant under the passage from the potential V to $V^\beta = V + \beta$, we can apply Theorem 2.4.1 to the potential V^β and obtain

$$\mathbb{E}_{\mathbb{P}^\beta}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}^{\beta,\beta}_u(\overline{X}_u, \overline{\mathrm{P}}^{\beta}_u)\big|^2 \,\mathrm{d}u\bigg] < \infty. \tag{2.144}$$

Now, since $\ell_t^\beta(x,\mu)/\ell_t^{\beta,\beta}(x,\mu) = \mathrm{e}^{\beta(x)}$, we observe that the difference

$$\nabla \log \ell_t^\beta(x,\mu) - \nabla \log \ell_t^{\beta,\beta}(x,\mu) = \nabla \beta(x) \tag{2.145}$$

is a bounded function. Together with (2.144), this implies (2.142).

It remains to check (2.62). A similar calculation as in the proof of Theorem 2.4.1 leads to the identity

$$\mathbb{E}_{\mathbb{P}^\beta}\big[\overline{F}_s^\beta\big] = \int_0^s \mathbb{E}_{\mathbb{P}^\beta}\Big[\big|\nabla \log \overline{\ell}_u^{\beta\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u^\beta)\big|^2 + \Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_u^\beta)\,,\nabla\beta\Big\rangle - \Delta\beta\Big)(\overline{X}_u)\Big]\,\mathrm{d}u \tag{2.146}$$

for $0 \leq s \leq T - t_0$. Repeating the reasoning of the previous paragraph for the function $\ell_t^{\beta\uparrow}$ instead of ℓ_t^β, we find that

$$\mathbb{E}_{\mathbb{P}^\beta}\bigg[\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\beta\uparrow}(\overline{X}_u^\beta, \overline{\mathrm{P}}_u^\beta)\big|^2\,\mathrm{d}u\bigg] < \infty. \tag{2.147}$$

Since the function

$$[0, T - t_0] \times \mathbb{R}^d \ni (t,x) \longmapsto \Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_t^\beta)\,, \nabla\beta\Big\rangle(x) - \Delta\beta(x) \tag{2.148}$$

is bounded, we conclude that the quantity of (2.146) is finite. Finally, recalling the definition (2.53), we arrive at (2.62). $\square$

2.5.2 The proofs of Proposition 2.4.17 and Theorem 2.4.19

Proof of Proposition 2.4.17. The first step is to view the probability density functions $(\rho_t)_{0\leq t\leq 1}$, corresponding to the displacement interpolation $(\nu_t)_{0\leq t\leq 1}$ of (2.82), as a solution to a continuity equation. Recalling the convex function $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ of (2.81), we define a function $\mathsf{u}_0\colon \mathbb{R}^d \to \mathbb{R}$ by $\mathsf{u}_0(x) \coloneqq \varphi(x) - |x|^2/2$; and for each $t \in (0,1]$, we let the function $\mathsf{u}_t\colon \mathbb{R}^d \to \mathbb{R}$ be defined by the Hopf–Lax formula

$$\mathsf{u}_t(x) \coloneqq \inf_{y\in\mathbb{R}^d}\Big(\mathsf{u}_0(y) + \frac{|x-y|^2}{2t}\Big). \tag{2.149}$$

For all $t \in [0, 1)$, we denote the gradient of u_t by $\mathsf{v}_t \coloneqq \nabla \mathsf{u}_t$. For $t = 0$, it is clear that $\mathsf{v}_0 = \nabla \varphi - \mathrm{Id}$ is well-defined. For $t \in (0, 1)$, the gradient v_t is defined Lebesgue-a.e. by [99, Theorem 5.51 (i)], and

$$\mathsf{v}_t(x) = \nabla \mathsf{u}_0 \circ (T_t)^{-1}(x) \,, \qquad \text{for all } x \in T_t(\mathbb{R}^d), \tag{2.150}$$

where T_t is defined in (2.82). Note that the inverse of T_t is well-defined because T_t is injective; see [99, Section 5.4.8]. From (2.150) we see that $(\mathsf{v}_t)_{0 \leq t < 1}$ is the *velocity field* associated with the trajectories $(T_t)_{0 \leq t < 1}$, i.e.,

$$T_t(x) = x + \int_0^t \mathsf{v}_s(T_s(x)) \, \mathrm{d}s \,, \qquad 0 \leq t < 1. \tag{2.151}$$

By [99, Theorem 5.51 (ii)], the curve of probability density functions $(\rho_t)_{0<t<1}$ satisfies the continuity equation

$$\partial_t \rho_t(x) + \operatorname{div}\big(\rho_t(x) \, \mathsf{v}_t(x)\big) = 0 \,, \qquad (t, x) \in (0, 1) \times \mathbb{R}^d. \tag{2.152}$$

On a sufficiently rich probability space $(S, \mathcal{S}, \mathsf{P})$, we let $Z_0 \colon S \to \mathbb{R}^d$ be a random variable with probability distribution ν_0. For each $0 < t \leq 1$, we let $Z_t \coloneqq T_t(Z_0)$. From (2.82) we see that the random variable Z_t has distribution ν_t, and (2.151) yields the representation

$$Z_t = Z_0 + \int_0^t \mathsf{v}_s(Z_s) \, \mathrm{d}s \,, \qquad 0 \leq t < 1. \tag{2.153}$$

In conjunction with (2.152), we deduce

$$\mathrm{d}\rho_t(Z_t) = \partial_t \rho_t(Z_t) + \big\langle \nabla \rho_t(Z_t) \,, \, \mathrm{d}Z_t \big\rangle = -\rho_t(Z_t) \operatorname{div}\big(\mathsf{v}_t(Z_t)\big) \, \mathrm{d}t, \tag{2.154}$$

and thus

$$\mathrm{d} \log \rho_t(Z_t) = -\operatorname{div}\big(\mathsf{v}_t(Z_t)\big) \, \mathrm{d}t. \tag{2.155}$$

Recalling the definition of the density function q in (2.19), a similar argument as in (2.102) shows that

$$\big(\partial_\nu \log q(x,\nu)\big)(y) = \tfrac{1}{2}\big(\partial_\nu \left(W * \nu\right)(x)\big)(y) = -\tfrac{1}{2}\nabla W(x-y) \tag{2.156}$$

for $(x,\nu,y) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$. Applying a generalized version of Itô's formula for McKean–Vlasov diffusions [78, Proposition 5.102], and using the dynamics (2.153) as well as the L-derivative (2.156), we obtain

$$\mathrm{d}\log q(Z_t,\nu_t) = -\Big\langle \nabla V + \tfrac{1}{2}\nabla(W*\nu_t)\,,\,\mathsf{v}_t\Big\rangle(Z_t)\,\mathrm{d}t + \tfrac{1}{2}\mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t)\,,\,\mathsf{v}_t(\tilde{Z}_t)\Big\rangle\Big]\,\mathrm{d}t \tag{2.157}$$

for $0 < t < 1$. Here, the process $(\tilde{Z}_t)_{0<t<1}$ is defined on another probability space $(\tilde{S},\tilde{\mathcal{S}},\tilde{\mathsf{P}})$ such that the tuple $(S,\mathcal{S},\mathsf{P},(Z_t)_{0<t<1})$ is an exact copy of $(\tilde{S},\tilde{\mathcal{S}},\tilde{\mathsf{P}},(\tilde{Z}_t)_{0<t<1})$. Now taking the difference between (2.155) and (2.157) gives the dynamics

$$\log r_t(Z_t,\nu_t) - \log r_0(Z_0,\nu_0) = -\tfrac{1}{2}\int_0^t \mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_s - \tilde{Z}_s)\,,\,\mathsf{v}_s(\tilde{Z}_s)\Big\rangle\Big]\,\mathrm{d}s \tag{2.158}$$

$$+\int_0^t \Big(\Big\langle \nabla V + \tfrac{1}{2}\nabla(W*\nu_s)\,,\,\mathsf{v}_s\Big\rangle(Z_s) - \operatorname{div}\big(\mathsf{v}_s(Z_s)\big)\Big)\,\mathrm{d}s \tag{2.159}$$

of the relative entropy process $(\log r_t(Z_t,\nu_t))_{0<t<1}$. Next, let us make two observations. Firstly, integration by parts yields

$$\mathbb{E}_{\mathsf{P}}\Big[\operatorname{div}\big(\mathsf{v}_t(Z_t)\big)\Big] = -\mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla \log \rho_t(Z_t)\,,\,\mathsf{v}_t(Z_t)\Big\rangle\Big]. \tag{2.160}$$

Secondly, by applying Fubini's theorem, and using that W is an even function as well as $(\tilde{Z}_t)_\#\tilde{\mathsf{P}} = \nu_t$, we obtain the identity

$$\mathbb{E}_{\mathsf{P}}\Big[\mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t)\,,\,\mathsf{v}_t(\tilde{Z}_t)\Big\rangle\Big]\Big] = -\mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla(W*\nu_t)(Z_t)\,,\,\mathsf{v}_t(Z_t)\Big\rangle\Big]. \tag{2.161}$$

Returning to (2.158), (2.159), we take P-expectations and use (2.160), (2.161) to obtain

$$H(\nu_t \,|\, \mu_t) - H(\nu_0 \,|\, \mu_0) = \int_0^t \mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla \log \rho_s + \nabla V + \nabla(W * \nu_s) \,,\, \mathsf{v}_s \Big\rangle (Z_s)\Big] \,\mathrm{d}s \tag{2.162}$$

$$= \int_0^t \mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla \log r_s^{\uparrow}(Z_s, \nu_s) \,,\, \mathsf{v}_s(Z_s) \Big\rangle\Big] \,\mathrm{d}s, \tag{2.163}$$

where for the second equality we recall the notations in (2.84) and (2.19). Finally, letting $t \downarrow 0$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^d} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \,,\, \mathsf{v}_0(x) \Big\rangle \,\rho_0(x) \,\mathrm{d}x; \tag{2.164}$$

and since $\mathsf{v}_0 = \nabla \varphi - \mathrm{Id}$, we arrive at (2.87). ☐

Proof of Theorem 2.4.19 . Without loss of generality, we assume that the probability density functions ρ_0 and ρ_1 satisfy the strong regularity Assumptions 2.4.16. The general case then follows by a density argument. We will not provide the details here, but refer to [99, Chapter 9.4], where this regularization is carried out in the simpler setting of the HWI inequality.

Let us recall the energy functionals $\mathscr{U}$, $\mathscr{V}$, $\mathscr{W}$ defined in (2.3), and introduce the functions

$$f(t) \coloneqq \mathscr{U}(\rho_t) \,, \qquad g(t) \coloneqq \mathscr{V}(\rho_t) \,, \qquad h(t) \coloneqq \mathscr{W}(\rho_t) \,, \qquad 0 \leq t \leq 1, \tag{2.165}$$

where $(\rho_t)_{0 \leq t \leq 1}$ is the curve of probability density functions corresponding to the displacement interpolation $(\nu_t)_{0 \leq t \leq 1}$ of (2.82). Then the sum $F \coloneqq f + g + h$ of these functions satisfies the relation $F(t) = H(\nu_t \,|\, \mu_t)$. In light of [99, Theorem 5.15 (i)], the internal energy functional $\mathscr{U}$ is displacement convex, i.e.,

$$f''(t) \geq 0 \,, \qquad 0 \leq t \leq 1. \tag{2.166}$$

By Assumptions 2.4.18, the confinement potential $V \colon \mathbb{R}^d \to [0, \infty)$ is κ_V-uniformly convex. Therefore, [99, Theorem 5.15 (ii)] implies that the potential energy functional $\mathscr{V}$ is κ_V-uniformly displacement convex. In other words,

$$g''(t) \geq \kappa_V \, \mathcal{W}_2^2(\nu_0, \nu_1) \, , \qquad 0 \leq t \leq 1. \tag{2.167}$$

Again from Assumptions 2.4.18, the interaction potential $W \colon \mathbb{R}^d \to [0, \infty)$ is assumed to be symmetric and κ_W-uniformly convex. Therefore, a similar argument as in the proof of [99, Theorem 5.15 (iii)] leads to the $\kappa_W(\mathcal{W}_2^2(\nu_0, \nu_1) - |b(\nu_0) - b(\nu_1)|^2)$-uniform convexity of h, so

$$h''(t) \geq \kappa_W \Big(\mathcal{W}_2^2(\nu_0, \nu_1) - |b(\nu_0) - b(\nu_1)|^2 \Big) \, , \qquad 0 \leq t \leq 1. \tag{2.168}$$

The details of the proof of (2.168) are postponed to Chapter 2.6.2. By combining the estimates (2.166) – (2.168), we deduce that the relative entropy function $[0,1] \ni t \mapsto F(t) = H(\nu_t \, | \, \mu_t)$ satisfies

$$F''(t) \geq (\kappa_V + \kappa_W) \, \mathcal{W}_2^2(\nu_0, \nu_1) - \kappa_W \, |b(\nu_0) - b(\nu_1)|^2. \tag{2.169}$$

Furthermore, from Proposition 2.4.17 we have

$$F'(0^+) = \int_{\mathbb{R}^d} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \, , \nabla \varphi(x) - x \Big\rangle \, \rho_0(x) \, \mathrm{d}x. \tag{2.170}$$

In conjunction with (2.169) and (2.170), the Taylor formula $F(1) = F(0) + F'(0^+) + \int_0^1 (1 - t) F''(t) \, \mathrm{d}t$ now yields the inequality (2.89) – (2.90). $\square$

2.6 Proofs of auxiliary results

2.6.1 Proof of Lemma 2.3.2

The generalized potential $\Psi^{\uparrow}$ of (2.18) allows us to cast the McKean–Vlasov dynamics of (2.1) in the more compact form

$$\mathrm{d}X_t = -\nabla \Psi^{\uparrow}(X_t, \mathrm{P}_t) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}B_t \, , \qquad 0 \leq t \leq T. \tag{2.171}$$

Then, for any two pairs $(x, \mu), (x', \mu') \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, using the Lipschitz continuity of ∇V in Assumptions 2.3.1 (i) yields

$$|\nabla \Psi^{\uparrow}(x, \mu) - \nabla \Psi^{\uparrow}(x', \mu')| \leq \|\nabla V\|_{\mathrm{Lip}} |x - x'| + |\nabla(W * \mu)(x) - \nabla(W * \mu')(x')|. \tag{2.172}$$

For the convolution term, using Jensen's inequality and the Lipschitz continuity of ∇W in Assumptions 2.3.1 (i) leads to

$$|\nabla(W * \mu)(x) - \nabla(W * \mu')(x')| \leq \|\nabla W\|_{\mathrm{Lip}} |x - x'| + |\nabla(W * \mu)(x') - \nabla(W * \mu')(x')|. \tag{2.173}$$

For the last term above, by the Kantorovich–Rubinstein theorem [99, Theorem 1.14], we have

$$\begin{aligned}\left| \int_{\mathbb{R}^d} \nabla W(x' - \,\cdot\,) \, \mathrm{d}(\mu - \mu') \right| &\leq \|\nabla W\|_{\mathrm{Lip}} \sup \left\{ \int_{\mathbb{R}^d} \varphi \, \mathrm{d}(\mu - \mu') \colon \varphi \in L^1(|\mu - \mu'|), \|\varphi\|_{\mathrm{Lip}} \leq 1 \right\} \\ &= \|\nabla W\|_{\mathrm{Lip}} \, \mathcal{W}_1(\mu, \mu') \leq \|\nabla W\|_{\mathrm{Lip}} \, \mathcal{W}_2(\mu, \mu'), \end{aligned} \tag{2.174}$$

where

$$\mathcal{W}_1(\mu, \mu') = \inf_{Y \sim \mu, Z \sim \mu'} \mathbb{E}|Y - Z| \,, \qquad \mu, \mu' \in \mathcal{P}_1(\mathbb{R}^d) \tag{2.175}$$

denotes the 1-Wasserstein-distance, and the inequality in (2.174) follows from Jensen's inequality. Altogether, we obtain

$$|\nabla \Psi^{\uparrow}(x, \mu) - \nabla \Psi^{\uparrow}(x', \mu')| \leq (\|\nabla V\|_{\mathrm{Lip}} + \|\nabla W\|_{\mathrm{Lip}}) \, |x - x'| + \|\nabla W\|_{\mathrm{Lip}} \, \mathcal{W}_2(\mu, \mu'). \tag{2.176}$$

In particular, this shows that the function $-\nabla \Psi^{\uparrow}$ is Lipschitz continuous on the product metric space $(\mathbb{R}^d, |\cdot|) \times (\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$. In conjunction with Assumptions 2.3.1 (ii) , [78, Theorem 4.21] implies that the McKean–Vlasov SDE (2.171) has a pathwise unique, strong solution satisfying the uniform second moment condition (2.17). Now we can linearize (2.171) by fixing the time-marginals $(\mathrm{P}_t)_{0 \leq t \leq T}$, so that the drift term can be viewed as a function $(t, x) \mapsto \nabla \Psi^{\uparrow}(x, \mathrm{P}_t)$, and

(2.171) becomes an ordinary SDE with a time-inhomogeneous drift coefficient.

The absolute continuity of the time-marginals $(\mathrm{P}_t)_{0\le t\le T}$ is immediate from Lemma 2.5.6. A standard argument using the classical Itô's formula shows that the curve of probability density functions $(p_t)_{0\le t\le T}$ is a weak solution of the granular media equation (2.11). Finally, we turn to the regularity of this solution. From (2.176), we see that the drift $x \mapsto \nabla\Psi^{\uparrow}(x, \mathrm{P}_t)$ is Lipschitz continuous for every $t \in [0, T]$, and Assumptions 2.3.1 (i) implies that the drift is also of linear growth. The desired smoothness of $(p_t)_{0\le t\le T}$ now follows from a straightforward adaptation of the theorem in [107], see also Remarks (i) – (ii) therein. □

2.6.2 Proof of (2.168)

We first rewrite the interaction energy functional $\mathscr{W}$ along the displacement interpolation $(\nu_t)_{0\le t\le 1}$. Using (2.82), for any $t \in [0, 1]$, we have

$$h(t) = \tfrac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\,\nu_t(\mathrm{d}x)\,\nu_t(\mathrm{d}y) \tag{2.177}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W\big(T_t(x)-T_t(y)\big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y) \tag{2.178}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W\Big(x-y-t\big(\theta(x)-\theta(y)\big)\Big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y), \tag{2.179}$$

where $\theta\colon \mathbb{R}^d \to \mathbb{R}^d$ is defined as $\theta(x) := x - \nabla\varphi(x)$. Now, for any $t_1, t_2, \sigma \in [0, 1]$, by the κ_W-uniform convexity of W in Assumptions 2.4.18, we obtain

$$\sigma h(t_1) + (1-\sigma)h(t_2) - h\big(\sigma t_1 + (1-\sigma)t_2\big) \tag{2.180}$$

$$= \tfrac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} \Bigg(\sigma W\Big(x-y-t_1\big(\theta(x)-\theta(y)\big)\Big) + (1-\sigma)W\Big(x-y-t_2\big(\theta(x)-\theta(y)\big)\Big) \tag{2.181}$$

$$- W\Big(x-y-\big(\sigma t_1+(1-\sigma)t_2\big)\big(\theta(x)-\theta(y)\big)\Big)\Bigg)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y) \tag{2.182}$$

$$\ge \tfrac{1}{4}\kappa_W\sigma(1-\sigma)(t_1-t_2)^2\int_{\mathbb{R}^d\times\mathbb{R}^d} |\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y). \tag{2.183}$$

Next, we express the integral in (2.183) as

$$\begin{aligned}\tfrac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}|\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y) &= \int_{\mathbb{R}^d}|\theta(x)|^2\,\nu_0(\mathrm{d}x) - \left|\int_{\mathbb{R}^d}\theta(x)\,\nu_0(\mathrm{d}x)\right|^2 \\ &= \int_{\mathbb{R}^d}|x-\nabla\varphi(x)|^2\,\nu_0(\mathrm{d}x) - \left|\int_{\mathbb{R}^d}x\,\nu_0(\mathrm{d}x) - \int_{\mathbb{R}^n}x\,\nu_1(\mathrm{d}x)\right|^2 \\ &= \mathcal{W}_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2.\end{aligned}$$

Putting this back into (2.183), we deduce that h is uniformly convex, with constant

$$\kappa_W\Big(\mathcal{W}_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2\Big). \tag{2.184}$$

$\square$

Chapter 3: A trajectorial approach to entropy dissipation of degenerate parabolic equations

In this chapter, we consider degenerate diffusion equations of the form $\partial_t p_t = \Delta f(p_t)$ on a bounded domain and subject to no-flux boundary conditions, for a class of nonlinearities f that includes the porous medium equation. We derive for them a trajectorial analogue of the entropy dissipation identity, which describes the rate of entropy dissipation along every path of the diffusion. In line with the recent work [15], our approach is based on applying stochastic analysis to the underlying probabilistic representations, which in our context are stochastic differential equations with normal reflection on the boundary. This trajectorial approach also leads to a new derivation of the Wasserstein gradient flow property for nonlinear diffusions, as well as to a simple proof of the HWI inequality in the present context. This chapter is based on the paper [22] joint with Donghan Kim.

3.1 Introduction

In this chapter, we are interested in a class of quasilinear degenerate parabolic equations with initial and no-flux boundary conditions of the following form:

$$\begin{cases} \partial_t p(t,x) = \Delta\Big(f\big(p(t,x)\big)\Big), & \text{for } (t,x) \in (0,T) \times U, \\ p(0,x) = p_0(x), & \text{for } x \in \overline{U}, \\ \dfrac{\partial p(t,x)}{\partial n(x)} = 0, & \text{for } (t,x) \in (0,T) \times \partial U, \end{cases} \tag{3.1}$$

for a fixed $T \in (0,\infty)$, an open connected bounded domain $U \subset \mathbb{R}^d$, and a given initial probability density function p_0 on $\overline{U}$. Here, $n(x)$ is the outward normal to the boundary ∂U at $x \in \partial U$, and

$f : [0, \infty) \to \mathbb{R}$ is a function representing the nonlinearity. In particular, when $f(u) = u^m$ for some $m > 1$, the partial differential equation of (3.1) becomes the *porous medium equation*.

Under suitable assumptions on f, it is well known from [108] that the solution of (3.1) converges to a unique stationary distribution, i.e., a probability density function p_∞ satisfying $\Delta\big(f(p_\infty(x))\big) = 0$, and that this convergence can be quantified by the rate of entropy dissipation. More precisely, let us define $h : (0, \infty) \to \mathbb{R}$ and $\Phi : [0, \infty) \to \mathbb{R}$ by

$$h(u) := \int_1^u \frac{f'(s)}{s}\,\mathrm{d}s, \quad \Phi(u) := \int_0^u h(s)\,\mathrm{d}s. \tag{3.2}$$

Note that the function h plays the role of the "generalized logarithm" since it becomes the logarithm with the choice $f(u) = u$, i.e., when (3.1) turns into the heat equation. Define also the *entropy functional*

$$\mathscr{F}(p) := \int_U \Phi\big(p(x)\big)\,\mathrm{d}x, \tag{3.3}$$

for any probability density function p on $\overline{U}$ such that the integral is finite. Then it can be shown that the stationary probability density function p_∞ is the minimizer of $\mathscr{F}$. Also, by abbreviating $p_t := p(t, \cdot)$, it is well known (see, e.g. [108, Equation (4)]) that

$$\mathscr{F}(p_t) - \mathscr{F}(p_{t_0}) = -\int_{t_0}^t I(p_u)\mathrm{d}u \tag{3.4}$$

holds for every $0 \le t_0 \le t \le T$, where I is the *entropy dissipation functional*, defined by

$$I(p) := \int_U \big|\Phi''\big(p(x)\big)\nabla p(x)\big|^2\, p(x)\,\mathrm{d}x \tag{3.5}$$

for any differentiable probability density function p such that the integral is finite. This identity measures the rate of entropy dissipation along the flow of the time-marginal probability densities $(p_t)_{0 \le t \le T}$, hence is known as the *entropy dissipation identity*. In particular, the entropy functional

$t \mapsto \mathscr{F}(p_t)$ is decreasing in time. See also [109, Lemma 18.14] and [110, Equation 3.4] for a specific form of this identity for the porous medium equation with drift.

The identity (3.4) describes the rate of entropy dissipation at the *ensemble* level of the diffusion modeled by (3.1), since it is formulated in terms of the probability density functions $(p_t)_{0\le t\le T}$ of the diffusion. The main goal of this paper is to formulate a *trajectorial* analogue to this identity, which describes the rate of entropy dissipation at the level of the individual diffusive particle. To illustrate this, we begin with the following stochastic differential equation (SDE) with values in $\overline{U}$ and normal reflection at the boundary:

$$X_t = X_0 + \int_0^t \sqrt{\frac{2f\big(p(s,X_s)\big)}{p(s,X_s)}}\,\mathrm{d}B_s - \int_0^t n(X_s)\,\mathrm{d}L_s, \qquad X_0 \sim p_0. \tag{3.6}$$

Here, B is a d-dimensional standard Brownian motion and L is a nondecreasing continuous process satisfying

$$L_t = \int_0^t 1_{\{X_s\in\partial\Omega\}}\,\mathrm{d}L_s, \quad L_0 = 0. \tag{3.7}$$

The stochastic process $(X_t)_{0\le t\le T}$ provides the probabilistic representation of (3.1), in the sense that its time-marginal probability density function coincides with the solution $p(t,\cdot)$ of (3.1) (see Lemma 3.2.4 below for details).

Intuitively, the diffusion X_t is reflected on the boundary ∂U in the direction $-n(X_t)$. The reflecting term L is associated with a multi-dimensional analogue of the *local time* on ∂U [111].

With this probabilistic representation, the entropy at time t can then be expressed as an expectation:

$$\mathscr{F}(p_t) = \int_U v(t,x)\,p(t,x)\,\mathrm{d}x = \mathbb{E}\big[v(t,X_t)\big], \quad \text{where}^1\ v(t,x) := \frac{\Phi\big(p(t,x)\big)}{p(t,x)}. \tag{3.8}$$

Using stochastic analysis, we shall derive the dynamics of the *entropy process* $\big(v(t,X_t)\big)_{t\in[0,T]}$, in

[1] In the case of the porous medium equation, v is known as the *pressure function*.

terms of the *semimartingale decomposition*

$$v(t, X_t) - v(0, X_0) = M_t + F_t, \qquad \text{for} \quad 0 \leq t \leq T, \tag{3.9}$$

where M is a martingale and F is a process of finite variation. This decomposition describes the evolution of the entropy process along every trajectory of the diffusive particle, and can thus be seen as a trajectorial analogue of (3.4). In fact, (3.4) can be recovered from (3.9) by averaging over these trajectories; in other words, by taking expectations.

Our work is much inspired from the recent work [15], which provides a trajectorial approach to the relative entropy dissipation for Fokker-Planck equations. This approach has been extended to Markov chains [83] and to McKean-Vlasov equations [22]. It is therefore natural to expect an adaptation of the approach for the porous media type equation (3.1). Compared with prior work, a key difficulty in our setting stems from the degenerate parabolicity of (3.1). More specifically, as f' is not assumed to be bounded from below by some strictly positive constant, the equation (3.1) is not uniformly parabolic. Without uniform parabolicity, equations of this form are only expected to have weak solutions [112], but not classical solutions. However, such regularity is crucial for applying Itô calculus. To this end, we require the initial condition p_0 to be nondegenerate, i.e., to satisfy $\kappa^{-1} \leq p_0(x) \leq \kappa$ for some $\kappa > 1$. This will ensure that (3.1) has a smooth solution. Also, in addition to considering diffusions on a bounded domain, another main difference with the prior work [95, 22] is that our main trajectorial result (Theorem 3.2.5 below) is stated in the forward direction of time.

Along with our trajectorial approach come two applications. The first application is a new derivation of the Wasserstein gradient flow property of (3.1), which states that the curve of time-marginal probability density functions $(p_t)_{t \in [0,T]}$ of (3.1) descends in the steepest possible direction of the entropy functional $\mathscr{F}$ in $\mathcal{P}(\overline{U})$, the space of probability measures on U. Here, $\mathcal{P}(\overline{U})$ is equipped with the quadratic Wasserstein distance $\mathcal{W}_2$, defined by

$$\mathcal{W}_2(\mu,\nu) := \sqrt{\inf_{\pi} \int_{U\times U} |x-y|^2\, \pi(\mathrm{d}x,\mathrm{d}y)}, \quad \text{for any } \mu,\nu \in \mathcal{P}(\overline{U}), \tag{3.10}$$

where the infimum is taken over $\pi \in \mathcal{P}(\overline{U}\times\overline{U})$ with marginals μ and ν.

For the porous medium equation on $\mathbb{R}^d$, this property was discovered by Otto in his seminal paper [82], where he introduced a formal Riemannian structure on $\mathcal{P}(\mathbb{R}^d)$. More recently, Ambrosio, Gigli and Savaré [19] developed a rigorous theory of gradient flows on general metric spaces based on the notion of curves of maximal slopes. Similar results have been established for porous medium equations on discrete spaces [113] and with fractional pressure [114].

To show the entropic steepest descent property, we adopt the methodology in [15] of *perturbing* the SDE (3.6) from some time $t_0 \in [0,T)$ onwards, by adding a gradient drift $\nabla\beta$, namely,

$$X_t^\beta = X_{t_0}^\beta - \int_{t_0}^t \nabla\beta(X_s^\beta)\,\mathrm{d}s + \int_{t_0}^t \sqrt{\frac{2f\big(p^\beta(s,X_s^\beta)\big)}{p^\beta(s,X_s^\beta)}}\,\mathrm{d}B_s - \int_{t_0}^t n(X_s^\beta)\,\mathrm{d}L_s^\beta.$$

Here, $p^\beta(t,\cdot)$ is the time-marginal probability density function for the solution $(X_t^\beta)_{t_0\le t\le T_\beta}$ of this perturbed SDE (see Lemmas 3.20 and 3.2.10 for details). By deriving the dynamics of the associated perturbed entropy process, we obtain an analogous entropy dissipation identity for the perturbed diffusion. On the other hand, we can also explicitly compute the rates of changes of the Wasserstein distances along both the perturbed curve (p_t^β) and the unperturbed curve (p_t). Thus, the entropy dissipation rates can be measured not in terms of time elapsed, but in terms of the Wasserstein distances traveled by the curve of time-marginal probability density functions, both in the perturbed and the unperturbed settings. Comparing these rates allows us to establish the maximal rate of entropy dissipation for the unperturbed diffusion (3.1), by measuring the exact effect of each perturbation.

The second application of the trajectorial approach is a simple proof of the HWI inequality in the context of the nonlinear equation (3.1), which is a special case of [20, Theorem 4.2]. It is an interpolation inequality relating the entropy functional (H)[2], the Wasserstein distance (W) and the entropy dissipation functional (I). More precisely, this inequality states that

$$\mathscr{F}(\rho_0) - \mathscr{F}(\rho_1) \leq \sqrt{I(\rho_0)}\, \mathcal{W}_2(\rho_0, \rho_1) \tag{3.11}$$

holds for any $\rho_0, \rho_1 \in \mathcal{P}(\overline{U})$. We will prove this inequality by applying a trajectorial approach similar to the one just described, but to the *displacement interpolation* between ρ_0 and ρ_1, instead of the time-marginal probability density functions of $(X_t)_{0 \leq t \leq T}$ and $(X_t^{\beta})_{t_0 \leq t \leq T_\beta}$.

The rest of the chapter is organized as follows. In Chapter 3.2.1, we introduce our setup and state some preliminary lemmas. Chapter 3.2.2 states our result on the trajectorial rate of entropy dissipation for the degenerate parabolic equation. Building on it, Chapter 3.2.3 formulates the gradient flow property via a perturbation analysis, while Chapter 3.2.4 develops the HWI inequality. Proofs are provided in Chapter 3.3.

3.2 Setting and main results

3.2.1 Setup

We impose the following assumptions on the initial distribution p_0 and on the nonlinearity f of the degenerate parabolic equation (3.1).

Assumptions 3.2.1.

(a) The domain U is an open connected bounded subset of $\mathbb{R}^d$ for some $d \geq 2$, and the boundary ∂U is smooth.

(b) The initial datum p_0 is a smooth probability density function on $\overline{U}$. Moreover, it is non-degenerate, i.e., there exists a real number $\kappa > 1$ for which $\kappa^{-1} \leq p_0(x) \leq \kappa$ holds for all

[2] The letter “H" comes from the choice $f(u) = u$ in (3.2), in which case the entropy functional defined in (3.3) satisfies $\mathscr{F}(p) = H(p) - 1$, where $H(p) = \int_U p(x) \log p(x)\, \mathrm{d}x$ is the (negative of the) differential entropy.

$x \in \overline{U}$. We assume also

$$\frac{\partial p_0(x)}{\partial n(x)} = 0, \quad \text{for all } x \in \partial U, \tag{3.12}$$

where $n(x)$ is the outward normal to the boundary ∂U.

(c) The function $f : [0, \infty) \to \mathbb{R}$ is smooth and strictly increasing. Moreover, $f(0) = f'(0) = 0$ and $f'(u) > 0$ for all $u > 0$. Its derivative f' is nondecreasing.

(d) The function h, defined in (3.2), belongs to $L^1_{\text{loc}}([0, \infty))$.

Remark 3.2.2. Assumptions 3.2.1(c) - (d) cover the porous medium equations, in which case $f(u) = u^m$ for some $m > 1$. The assumption $f'(u) > 0$ implies that (3.1) is a *parabolic* partial differential equation (PDE), while the assumption $f'(0) = 0$ implies that (3.1) is *degenerate* parabolic.

We collect now some basic properties of the solution to the PDE (3.1) in the following lemma. These properties are classical, and we refer to Chapter 3 of the monograph [109], and to the references therein, for a comprehensive overview.

Lemma 3.2.3. *Under Assumption 3.2.1, there exists a smooth solution* $p \in C^\infty([0, T] \times \overline{U})$ *of* (3.1). *Moreover,* $\int_U p(t, x)\, \mathrm{d}x = 1$ *for all* $t \in [0, T]$ *and* $\kappa^{-1} \le p(t, x) \le \kappa$ *for all* $(t, x) \in [0, T] \times \overline{U}$.

For the rest of the paper, we fix p as the solution given in Lemma 3.2.3. Fix also a filtered probability space $\big(\Omega, \mathbb{F}, \mathcal{F} = (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P}\big)$ supporting a $\mathcal{F}$-Brownian motion B and a $\mathcal{F}_0$-measurable random vector $\xi : \Omega \to \mathbb{R}^d$ with $\mathbb{P} \circ \xi^{-1} = p_0$. We shall denote by $\mathbb{E}$ the expectation taken with respect to $\mathbb{P}$. We will make use of the probabilistic representation of (3.1), which is described in Lemma 3.2.4 below in terms of the solution of the following SDE with normal reflection on the boundary:

$$\begin{cases} X_t = X_0 + \displaystyle\int_0^t \sqrt{\frac{2f\big(p(s,X_s)\big)}{p(s,X_s)}}\,\mathrm{d}B_s - \int_0^t n(X_s)\,\mathrm{d}L_s \in \overline{U}, \qquad t \in [0,T], \\ X_0 = \xi, \\ L_t = \displaystyle\int_0^t 1_{\{X_s \in \partial U\}}\,\mathrm{d}L_s, \quad [0,T] \ni t \to L_t \text{ is nondecreasing, continuous with } L_0 = 0. \end{cases} \tag{3.13}$$

For an introduction to SDEs with reflection and their connections with nonlinear parabolic PDEs, we refer to the lecture notes [115] and [116].

The following result shows that (3.13) is well-posed and provides the probabilistic representation of (3.1). Similar results are known for the porous medium equations [117] as well as for general nonlinear equations of the form (3.1) with discontinuous coefficients [118, 119], or with the half-line as the domain [120]. See also [121] for a martingale method for establishing gradient estimates for the porous medium and fast diffusion equations.

Lemma 3.2.4. *Suppose Assumption 3.2.1 holds. Then the SDE with reflection* (3.13) *has a pathwise unique, strong solution* (X, L)*, for which the probability density functions of* $(X_t)_{t\in[0,T]}$ *are given by the solution* $\big(p(t,\cdot)\big)_{t\in[0,T]}$ *of* (3.1).

3.2.2 Trajectorial entropy dissipation of degenerate parabolic equation

Our first main result describes the dynamics of entropy dissipation along every trajectory of the diffusion (3.13), formulated in terms of the semimartingale decomposition of the entropy process $\big(v(t,X_t)\big)_{t\in[0,T]}$.

To begin, we define the *entropy dissipation function*

$$D(t,x) := \left(\varphi'(p)\Delta f(p) + \frac{f(p)}{p}\Delta v\right)(t,x), \quad (t,x) \in [0,T] \times \overline{U}, \quad \text{where} \quad \varphi(u) := \frac{\Phi(u)}{u}, \tag{3.14}$$

and Φ is defined in (3.2). This function will help provide the exact trajectorial rate of entropy dissipation, as will be explained in Remark 3.2.7 below. We note that the function v of (3.8) can be cast as $v(t,x) = \varphi(p(t,x))$.

Theorem 3.2.5. *Suppose Assumption 3.2.1 holds. Then the entropy process* $\big(v(t,X_t)\big)_{t\in[0,T]}$ *admits the semimartingale decomposition*

$$v(t,X_t) - v(0,X_0) = M_t + F_t, \qquad \text{for} \quad t \in [0,T]. \tag{3.15}$$

where

$$F_t \coloneqq \int_0^t D(s,X_s)\,\mathrm{d}s, \quad \textit{and} \quad M_t \coloneqq \int_0^t \left\langle \sqrt{\frac{2f\big(p(s,X_s)\big)}{p(s,X_s)}}\nabla v(s,X_s),\, \mathrm{d}B_s \right\rangle \tag{3.16}$$

is an L^2*- bounded martingale. Also, we have*

$$\mathbb{E}\big[F_t\big] = -\int_0^t I\big(p_t\big)\,\mathrm{d}t > -\infty\,, \qquad \text{for} \quad t \in [0,T]. \tag{3.17}$$

The proof of Theorem 3.2.5 will be given in Chapter 3.3.3. This result is the analogue of [95, Theorem 4.1] and [22, Theorem 3.1]. In contrast to them, Theorem 3.2.5 here is stated in the forward direction of time.

By aggregating this trajectorial result, i.e., taking expectation, we recover the entropy dissipation identity (3.4) and its differential version (3.18). Furthermore, by taking *conditional* expectation, we obtain below the conditional trajectorial rate of entropy dissipation (3.19).

Corollary 3.2.6. *Suppose Assumption 3.2.1 holds. Then for every* $0 \le t_0 \le t \le T$*, the entropy dissipation identity* (3.4) *holds. The corresponding differential version*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0} \mathscr{F}(p_t) = -I(p_{t_0}) \tag{3.18}$$

also holds for all $t_0 \in [0,T]$*. Moreover, for all* $t_0 \in [0,T]$*, the conditional trajectorial rate of* *entropy dissipation is given by*

$$\lim_{t \downarrow t_0} \frac{\mathbb{E}[v(t, X_t) \,|\, \mathcal{F}_{t_0}] - v(t_0, X_{t_0})}{t - t_0} = D(t_0, X_{t_0}), \tag{3.19}$$

where the limit exists in $L^1(\mathbb{P})$.

Remark 3.2.7. From (3.17), we see that $\mathbb{E}[D(t_0, X_{t_0})] = -I(p_{t_0})$ holds. This explains why (3.19) constitutes a conditional trajectorial version of the entropy dissipation identity (3.18).

3.2.3 Gradient flow property of the degenerate parabolic equation, via perturbation analysis

In this section, we discuss how our trajectorial approach leads to a new interpretation of the Wasserstein gradient flow property of the degenerate parabolic equation (3.1). Following the method of [95, 22], we shall *perturb* the degenerate parabolic equation.

The perturbed degenerate parabolic equation

To this effect, let β be a perturbation potential satisfying the following assumption.

Assumptions 3.2.8. The perturbation potential $\beta : \overline{U} \to \mathbb{R}$ is smooth. Moreover, the gradient of the perturbation potential vanishes on the boundary, i.e., $\nabla \beta(x) = 0$ for $x \in \partial U$.

For the rest of the paper, we fix a $t_0 \in [0, T)$ and a perturbation β satisfying Assumption 3.2.8. Consider the following Neumann problem, which can be viewed as a perturbed version of (3.1):

$$\begin{cases} \partial_t p^\beta(t, x) = \operatorname{div}\Big(\nabla f\big(p^\beta(t, x)\big) + p^\beta(t, x) \nabla \beta(x)\Big), & \text{for } (t, x) \in (t_0, T] \times U, \\ p^\beta(t_0, x) = p(t_0, x), & \text{for } x \in \overline{U}, \\ \dfrac{\partial p^\beta(t, x)}{\partial n(x)} = 0, & \text{for } x \in \partial U. \end{cases} \tag{3.20}$$

The following result, which is the "perturbed analogue" of Lemma 3.2.3, shows the existence of a strictly positive smooth solution to (3.20) in a short time interval.

Lemma 3.2.9. *Under Assumptions 3.2.1 and 3.2.8, there exists $T_\beta \in (t_0, T]$ such that* (3.20) *has a smooth solution in $p^\beta \in C^\infty\big([t_0, T_\beta] \times \overline{U}\big)$. Moreover, $\int_U p^\beta(t,x)\,\mathrm{d}x = 1$ for all $t \in [t_0, T_\beta]$ and $\frac{1}{2\kappa} \le p^\beta(t,x) \le \frac{3}{2\kappa}$ for all $(t,x) \in [t_0, T_\beta] \times \overline{U}$.*

For the rest of the paper, we fix p^β as given by Lemma 3.2.9. The corresponding probabilistic representation of the perturbed PDE (3.20) is the following SDE with reflection:

$$\begin{cases} X_t^\beta = X_{t_0}^\beta - \displaystyle\int_{t_0}^t \nabla\beta(X_s^\beta)\,\mathrm{d}s + \int_{t_0}^t \sqrt{\frac{2f\big(p^\beta(s, X_s^\beta)\big)}{p^\beta(s, X_s^\beta)}}\,\mathrm{d}B_s - \int_{t_0}^t n(X_s^\beta)\,\mathrm{d}L_s^\beta \in \overline{U}, \quad t \in [t_0, T_\beta], \\ X_{t_0}^\beta = X_{t_0}, \\ L_t^\beta = \displaystyle\int_{t_0}^t 1_{\{X_s^\beta \in \partial U\}}\,\mathrm{d}L_s^\beta, \quad [t_0, T_\beta] \ni t \mapsto L_t^\beta \text{ is nondecreasing continuous with } L_{t_0}^\beta = 0. \end{cases} \tag{3.21}$$

By analogy with Lemma 3.2.4, the following result ensures that (3.21) is well-posed and is the stochastic counterpart of (3.20).

Lemma 3.2.10. *Under Assumptions 3.2.1 and 3.2.8, the SDE with reflection* (3.21) *has a pathwise unique, strong solution $(X_t^\beta, L_t^\beta)_{t\in[t_0,T_\beta]}$, for which the probability density functions of $(X_t^\beta)_{t\in[t_0,T_\beta]}$ are given by $\big(p^\beta(t,\cdot)\big)_{t\in[t_0,T_\beta]}$, as in* (3.20)*.*

As before, let us abbreviate $p_t^\beta := p^\beta(t,\cdot)$. In parallel to (3.8), we can express the entropy for the perturbed diffusion at time t as

$$\mathscr{F}\big(p_t^\beta\big) = \int_U v^\beta(t,x)\,p^\beta(t,x)\,\mathrm{d}x = \mathbb{E}\big[v^\beta(t, X_t^\beta)\big], \quad \text{where} \quad v^\beta(t,x) := \frac{\Phi\big(p^\beta(t,x)\big)}{p^\beta(t,x)}. \tag{3.22}$$

Therefore, we similarly call $\big(v^\beta(t, X_t^\beta)\big)_{t\in[t_0,T_\beta]}$ the *perturbed entropy process*. The following result, which is the perturbed counterpart of Theorem 3.2.5, derives the dynamics of this process. By analogy with (3.14), we introduce the *perturbed entropy dissipation function*

$$D^\beta(t,x) := \bigg(\varphi'\big(p^\beta\big)\,\mathrm{div}\Big(\nabla f\big(p^\beta\big) + p^\beta\nabla\beta(x)\Big) + \frac{f\big(p^\beta\big)}{p^\beta}\Delta v^\beta - \langle \nabla v^\beta, \nabla\beta\rangle\bigg)(t,x), \tag{3.23}$$

for $(t,x) \in [t_0, T_\beta] \times \overline{U}$.

Theorem 3.2.11. *Suppose Assumptions 3.2.1 and 3.2.8 hold. Then the perturbed entropy process* $\big(v^\beta(t, X_t^\beta)\big)_{t\in[t_0,T_\beta]}$ *admits the semimartingale decomposition*

$$v^\beta(t, X_t^\beta) - v^\beta(t_0, X_{t_0}^\beta) = M_t^\beta + F_t^\beta, \qquad \textit{for} \quad t \in [t_0, T_\beta], \tag{3.24}$$

where

$$F_t^\beta \coloneqq \int_{t_0}^t D^\beta(s,\, X_s^\beta)\,\mathrm{d}s, \quad \textit{and} \quad M_t^\beta \coloneqq \int_{t_0}^t \left\langle \sqrt{\frac{2f\big(p^\beta(s, X_s^\beta)\big)}{p^\beta(s, X_s^\beta)}}\nabla v^\beta(s, X_s^\beta),\, \mathrm{d}B_s \right\rangle \tag{3.25}$$

is an L^2-bounded martingale. Also, we have for $t \in [t_0, T_\beta]$,

$$\mathbb{E}\big[F_t^\beta\big] = -\int_{t_0}^t I(p_s^\beta)\,\mathrm{d}s - \int_{t_0}^t \mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(s, X_s^\beta)\big),\, \nabla\beta(X_s^\beta)\big\rangle\Big]\,\mathrm{d}s > -\infty. \tag{3.26}$$

Once again, by averaging this trajectorial result, we obtain the following perturbed entropy dissipation identity and its conditional trajectorial version, to which a comment similar to Remark 3.2.7 applies.

Corollary 3.2.12. *Suppose Assumptions 3.2.1 and 3.2.8 hold. For every $t \in [t_0, T_\beta]$, the following perturbed entropy dissipation identity holds:*

$$\mathscr{F}\big(p_t^\beta\big) - \mathscr{F}\big(p_{t_0}^\beta\big) = -\int_{t_0}^t I\big(p_s^\beta\big)\mathrm{d}s - \int_{t_0}^t \mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(s, X_s^\beta)\big),\, \nabla\beta(X_s^\beta)\big\rangle\Big]\,\mathrm{d}s. \tag{3.27}$$

The corresponding differential version also holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0^+} \mathscr{F}(p_t^\beta) = -I(p_{t_0}) - \mathbb{E}\Big[\big\langle \nabla h\big(p(t_0, X_{t_0})\big),\, \nabla\beta(X_{t_0})\big\rangle\Big]. \tag{3.28}$$

Moreover, the conditional trajectorial rate of entropy dissipation for the perturbed diffusion is given by

$$\lim_{t\downarrow t_0} \frac{\mathbb{E}\Big[v^\beta(t, X_t^\beta)\,\big|\,\mathcal{F}_{t_0}\Big] - v^\beta(t_0, X_{t_0})}{t - t_0} = D^\beta(t_0, X_{t_0}), \tag{3.29}$$

where the limit exists in L^1.

Entropic steepest descent property

The last ingredients we need, in order to establish the entropic steepest descent property, are the rates of change of the Wasserstein distances along the curves of the marginal distributions (p_t) and (p_t^β). The following result is a consequence of the general theory of Wasserstein metric derivatives for absolutely continuous curves [19, Chapter 8], but our setting allows for a more direct proof, which we provide in Chapter 3.3.5. We recall the definitions of h in (3.2) and the perturbation potential β described at the beginning of Chapter 3.2.3. Note in the following that the function $(t,x) \mapsto h(p(t,x))$ plays the role of "generalized log-likelihood" function, as it becomes the log-likelihood function $(t,x) \mapsto \log(p(t,x))$ in the case of $f(u) = u$, i.e., when (3.1) turns into the heat equation.

Lemma 3.2.13. *Suppose Assumptions 3.2.1 and 3.2.8 hold. Then the Wasserstein metric slope along the unperturbed curve (p_t) is given by*

$$\lim_{t \downarrow t_0} \frac{\mathcal{W}_2(p_t, p_{t_0})}{t - t_0} = \left\| \nabla h\big(p(t_0, X_{t_0})\big) \right\|_{L^2}. \tag{3.30}$$

Similarly, the Wasserstein metric slope along the perturbed curve $\left(p_t^\beta\right)$ is given by

$$\lim_{t \downarrow t_0} \frac{\mathcal{W}_2\big(p_t^\beta, p_{t_0}^\beta\big)}{t - t_0} = \left\| \nabla h\big(p(t_0, X_{t_0})\big) + \nabla \beta(X_{t_0}) \right\|_{L^2}. \tag{3.31}$$

Combining the entropy dissipation identities (3.18) and (3.28) with the Wasserstein derivatives (3.30)–(3.31) allows us to derive the steepest descent property of the entropy. We define the Wasserstein metric slopes of the entropy functional $\mathscr{F}$ along the unperturbed curve (p_t) and along the perturbed curve (p_t^β), respectively, by

$$|\partial \mathscr{F}|_{\mathcal{W}_2}(p_{t_0}) := \lim_{t \downarrow t_0} \frac{\mathscr{F}(p_t) - \mathscr{F}(p_{t_0})}{\mathcal{W}_2(p_t, p_{t_0})} \quad \text{and} \quad |\partial \mathscr{F}|_{\mathcal{W}_2}(p_{t_0}^\beta) := \lim_{t \downarrow t_0} \frac{\mathscr{F}\big(p_t^\beta\big) - \mathscr{F}\big(p_{t_0}^\beta\big)}{\mathcal{W}_2(p_t^\beta, p_{t_0}^\beta)}. \tag{3.32}$$

The following theorem computes both of these slopes explicitly, which shows in particular that the unperturbed slope $|\partial \mathscr{F}|_{\mathcal{W}_2}(p_{t_0})$ is always steeper than the perturbed slope $|\partial \mathscr{F}|_{\mathcal{W}_2}(p^{\beta}_{t_0})$.

Theorem 3.2.14. *Suppose Assumptions 3.2.1 and 3.2.8 hold. Then the Wasserstein metric slope of the entropy functional along the unperturbed curve (p_t) is given by*

$$|\partial \mathscr{F}|_{\mathcal{W}_2}(p_{t_0}) = -\Big\| \nabla h\big(p(t_0, X_{t_0})\big) \Big\|_{L^2} = -\sqrt{I(p_{t_0})}. \tag{3.33}$$

Similarly, if $||\nabla h\big(p(t_0, X_{t_0})\big) + \nabla \beta(X_{t_0})||_{L^2} > 0$, then the Wasserstein metric slope along the perturbed curve $\big(p^{\beta}_t\big)$ is given by

$$|\partial \mathscr{F}|_{\mathcal{W}_2}\big(p^{\beta}_{t_0}\big) = -\left\langle \nabla h\big(p(t_0, X_{t_0})\big),\ \frac{\nabla h\big(p(t_0, X_{t_0})\big) + \nabla \beta(X_{t_0})}{\Big\| \nabla h\big(p(t_0, X_{t_0})\big) + \nabla \beta(X_{t_0}) \Big\|_{L^2}} \right\rangle_{L^2}. \tag{3.34}$$

In particular,

$$|\partial \mathscr{F}|_{\mathcal{W}_2}(p_{t_0}) \leq |\partial \mathscr{F}|_{\mathcal{W}_2}\big(p^{\beta}_{t_0}\big), \tag{3.35}$$

and equality holds if and only if $\nabla h\big(p(t_0, X_{t_0})\big) + \nabla \beta(X_{t_0})$ is a.s. a scalar multiple of $\nabla h\big(p(t_0, X_{t_0})\big)$.

3.2.4 HWI inequality

In this section, we apply a similar trajectorial approach to derive the HWI inequality (3.11). Let us fix two probability density functions $\rho_0, \rho_1 \in \mathcal{P}(\overline{U})$ and impose the following assumption.

Assumptions 3.2.15.

(a) Both ρ_0, ρ_1 are strictly positive and smooth. Also, $\rho_0(x) = \rho_1(x)$ for all $x \in \partial U$.

(b) The function $f : [0, \infty) \to \mathbb{R}$ is smooth. Moreover, the function h, defined in (3.2), belongs to $L^1_{\text{loc}}([0, \infty))$.

(c) The function $r \mapsto r^d \Phi(r^{-d})$ is convex nonincreasing on $(0, \infty)$, where Φ is defined in (3.2).

Remark 3.2.16. Assumption (b), (c) is satisfied by the porous medium equation, see [99, Examples 5.19].

By Brenier's theorem [99, Theorem 2.12(ii)], there exists a convex function $\psi : \overline{U} \to \mathbb{R}$ such that $\nabla\psi$ is the optimal transport map from ρ_0 to ρ_1, i.e.,

$$\mathcal{W}_2^2(\rho_0, \rho_1) = \int_U \left|x - \nabla\psi(x)\right|^2 \rho_0(\mathrm{d}x). \tag{3.36}$$

Let $(\rho_t)_{t\in(0,1)}$ denote the displacement interpolation between ρ_0 and ρ_1, i.e.,

$$\rho_t = \rho_0 \circ \big((1-t)\mathrm{Id} + t\nabla\psi\big)^{-1}, \quad \text{for } t \in (0,1).$$

It is known that each ρ_t has a probability density function [99, Remarks 5.13(i)]. For the following result, we recall the entropy functional $\mathscr{F}$ defined in (3.3), the nonlinearity f satisfying Assumption 3.2.15(b), and the convex function ψ described just above.

Proposition 3.2.17. *Suppose Assumption 3.2.1(a) and Assumption 3.2.15(a–b) hold. Then the rate of change of $t \mapsto \mathscr{F}(\rho_t)$ at $t = 0$ is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} \mathscr{F}(\rho_t) = \int_U \big\langle \nabla f\big(\rho_0(z)\big), \nabla\psi(z) - z\big\rangle \, \mathrm{d}z. \tag{3.37}$$

Using this proposition and the displacement convexity of the entropy functional $\mathscr{F}$, we obtain the HWI inequality (3.11).

Theorem 3.2.18. *Suppose Assumption 3.2.1(a) and Assumption 3.2.15 hold. Then*

$$\mathscr{F}(\rho_0) - \mathscr{F}(\rho_1) \le -\int_U \big\langle \nabla f\big(\rho_0(z)\big), \nabla\psi(z) - z\big\rangle \, \mathrm{d}z \le \sqrt{I(\rho_0)}\, \mathcal{W}_2(\rho_0, \rho_1), \tag{3.38}$$

where I is the entropy dissipation functional defined in (3.5).

3.3 Proofs

3.3.1 Proofs of Lemmas 3.2.3 and 3.2.9

Proof of Lemma 3.2.3. We adopt the same method as in the proof of [109, Theorem 3.1], which exploits the nondegeneracy of the initial condition of Assumption 3.2.1(b) in a crucial manner . Let $\tilde{f} : [0,\infty) \to \mathbb{R}$ be a smooth function satisfying $\tilde{f}(u) = f(u)$ for $\kappa^{-1} \le u \le \kappa$, $\tilde{f}'(u) > \epsilon^{-1}$ for $0 \le u \le \kappa^{-1}$ and $\tilde{f}'(u) < \epsilon$ for $u \ge \kappa$, where $\epsilon > 1$ is some fixed constant. Consider the PDE

$$\partial_t p(t,x) = \Delta\Big(\tilde{f}\big(p(t,x)\big)\Big), \qquad \text{for } (t,x) \in (0,T] \times U \tag{3.39}$$

subject to the same initial and Neumann boundary conditions as in (3.1). By Assumption 3.2.1(c), f' is increasing, so $\tilde{f}'(u) > \epsilon^{-1}$ for all $u \ge 0$. This implies that (3.39) is uniformly parabolic. We can therefore apply standard quasilinear theory [109, Chapter 3.1] to obtain a smooth solution $p \in C^\infty([0,T] \times \overline{U})$ to (3.39). By the comparison principle, $\kappa^{-1} \le p \le \kappa$, so p also satisfies (3.1).

Finally, the mass conservation law [109, Chapter 3.3.3] implies that the total mass $\int_U p(t,x)\,\mathrm{d}x = 1$ is conserved over time $t \in [0,T]$. □

Proof of Lemma 3.2.9. The proof is similar to that of Lemma 3.2.3. Let $\bar{f} : [0,\infty) \to \mathbb{R}$ be a smooth function satisfying $\bar{f}(u) = f(u)$ for $\frac{1}{2\kappa} \le u \le \kappa + \frac{1}{2\kappa}$, $\bar{f}'(u) > \epsilon^{-1}$ for $0 \le u \le \frac{1}{2\kappa}$ and $\bar{f}'(u) < \epsilon$ for $u \ge \kappa + \frac{1}{2\kappa}$, where $\epsilon > 1$ is some fixed constant. Consider the PDE

$$\partial_t p^\beta(t,x) = \mathrm{div}\Big(\nabla \bar{f}\big(p^\beta(t,x)\big) + p^\beta(t,x)\nabla\beta(x)\Big), \quad \text{for } (t,x) \in (t_0,T] \times U \tag{3.40}$$

subject to the same initial and Neumann boundary conditions as in (3.20). Again since $\bar{f}'(u) > \epsilon^{-1}$ for all $u \ge 0$, this PDE is uniformly parabolic. Therefore, standard quasilinear theory implies that there exists a smooth solution $p^\beta \in C^\infty([t_0,T] \times \overline{U})$ to (3.40).

For a fixed $\delta > 0$, we define

$$\lambda_\delta := \max_{t \in [t_0,\, t_0+\delta],\, x \in \overline{U}} \left|\partial_t p^\beta(t,x)\right| < \infty \quad \text{and} \quad \tau := \min\left(\delta,\, \frac{1}{2\kappa\lambda_\delta}\right) > 0.$$

Let $T_\beta := T \wedge (t_0 + \tau)$. Since $p^\beta_{t_0} = p_{t_0}$ by construction in (3.20),

$$\left|p^\beta(t,x) - p(t_0,x)\right| = \left|p^\beta(t,x) - p^\beta(t_0,x)\right| \le \int_{t_0}^{T_\beta} \left|\partial_t p^\beta(s,x)\right| \mathrm{d}s \le \tau\lambda_\delta \le \frac{1}{2\kappa},$$

for every $(t,x) \in [t_0, T_\beta] \times \overline{U}$. From Lemma 3.2.3, we have $\kappa^{-1} \le p_{t_0} \le \kappa$, thus $(2\kappa)^{-1} \le p^\beta_t \le \kappa + (2\kappa)^{-1}$ for all $t \in [t_0, T_\beta]$. This implies that $\bar{f}\big(p^\beta(t,x)\big) = f\big(p^\beta(t,x)\big)$ holds for every $(t,x) \in [t_0, T_\beta] \times \overline{U}$ and therefore p^β also solves (3.20).

Finally, for mass conservation, note that integration by parts gives us

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int_U p^\beta(t,x)\,\mathrm{d}x &= \int_U \operatorname{div}\Big(\nabla f\big(p^\beta(t,x)\big) + p^\beta(t,x)\nabla\beta(x)\Big)\,\mathrm{d}x \\
&= \int_{\partial U} \Big\langle \nabla f\big(p^\beta(t,x)\big) + p^\beta(t,x)\nabla\beta(x),\, n(x)\Big\rangle\,\mathrm{d}x = 0,
\end{aligned}$$

where the last step follows from the no-flux boundary condition in (3.20) as well as Assumption 3.2.8. $\square$

3.3.2 Proofs of Lemmas 3.2.4 and 3.2.10

We will only prove Lemma 3.2.10, as the proof of Lemma 3.2.4 is completely analogous.

For every $(\tau, y) \in [t_0, T_\beta) \times \overline{U}$, consider the SDE with reflection (3.21) conditional on the initial position y at time τ:

$$\begin{cases} X_t^{\beta,\tau,y} = y - \displaystyle\int_{t_0}^t \nabla\beta(X_s^{\beta,\tau,y})\,\mathrm{d}s + \int_\tau^t \sqrt{\frac{2f\big(p^\beta(s,X_s^{\beta,\tau,y})\big)}{p^\beta(s,X_s^{\beta,\tau,y})}}\mathrm{d}B_s - \int_\tau^t n(X_s^{\beta,\tau,y})\,\mathrm{d}L_s^{\beta,\tau,y} \in \overline{U}, \qquad t\in[\tau,T_\beta], \\ L_t^{\beta,\tau,y} = \displaystyle\int_\tau^t 1_{\{X_s^{\beta,\tau,y}\in\partial U\}}\,\mathrm{d}L_s^{\beta,\tau,y}, \\ L_\tau^{\beta,\tau,y} = 0 \text{ and } t\mapsto L_t^{\beta,\tau,y} \text{ is nondecreasing and continuous.} \end{cases} \tag{3.41}$$

From Lemma 3.2.9 and Assumption 3.2.1(c), it is straightforward to check that the diffusion coefficient is uniformly Lipschitz in the spatial variable. Thus by [122, Theorem 3.1 and Remark 3.3], the SDE with reflection (3.41) has a pathwise unique, strong solution. Let ξ be an independent $\overline{U}$-valued random variable with distribution p_{t_0}. Consider the process X^β given by $X_{t_0}^\beta = \xi$ and $X_t^\beta = X_t^{\beta,t_0,\xi}$ for $t\in(t_0,T_\beta]$. Similarly, let L^β be specified by $L_{t_0}^\beta = 0$ and $L_t^\beta = L_t^{\beta,t_0,\xi}$ for $t\in(t_0,T_\beta]$. Then (X^β,L^β) is the unique strong solution to (3.21). This completes the proof of the first part of the lemma.

Turning to the proof of the second part, we borrow ideas from [115, Remark 3.1.2] and [123, Chapter 5.7.B]. Recall that p^β is fixed as the solution of (3.20), given in Lemma 3.2.9. Consider the following *backward Kolmogorov equation*:

$$\begin{cases} \partial_\tau q^\beta(\tau,y) + \dfrac{f(p^\beta(\tau,y))}{p^\beta(\tau,y)}\Delta_y q^\beta(\tau,y) - \big\langle \nabla_y\beta(y), \nabla_y q^\beta(\tau,y)\big\rangle = 0, & \text{for } (\tau,y)\in(t_0,T_\beta)\times U, \\ \dfrac{\partial q^\beta(\tau,y)}{\partial n(y)} = 0, & \text{for } y\in\partial U. \end{cases} \tag{3.42}$$

It follows from [124] that (3.42) has a *fundamental solution* $G^\beta(\tau,y;t,x)$ defined for $t_0\le\tau<t\le T_\beta$ and $x,y\in\overline{U}$. In particular, G^β is nonnegative and for every $\phi\in C(\overline{U})$ and $t\in(\tau,T_\beta]$, the function

$$q^\beta(\tau,y) \coloneqq \int_U G^\beta(\tau,y;t,x)\,\phi(x)\,\mathrm{d}x, \tag{3.43}$$

satisfies (3.42) and the terminal condition

$$\lim_{\tau \uparrow t} q^\beta(\tau, y) = \phi(y), \quad \text{for all } y \in U. \tag{3.44}$$

If furthermore ϕ satisfies the no-flux boundary condition $\partial\phi/\partial n = 0$, then the above convergence holds uniformly in $\overline{U}$.

From the Feynman-Kac representation [115, Theorem 3.1.1], for any $\phi \in C(\overline{U})$,

$$q^\beta(\tau, y) = \mathbb{E}\big[\phi(X_t^{\beta,\tau,y})\big]. \tag{3.45}$$

Comparing (3.43) with (3.45), we deduce that the transition probability density of $X^{\beta,\tau,y}$ is given by G^β, i.e.,

$$\mathbb{P}\big(X_t^{\beta,\tau,y} \in A\big) = \int_A G^\beta(\tau, y; t, x)\,\mathrm{d}x, \quad \text{for every Borel } A \subseteq \overline{U}. \tag{3.46}$$

For any fixed $y \in \overline{U}$, the function $\Psi^\beta(t,x) := G^\beta(t_0, y; t, x)$ satisfies the *forward Kolmogorov equation*

$$\partial_t \Psi^\beta(t,x) = \mathrm{div}_x\bigg(\nabla_x\Big(\frac{f\big(p^\beta(t,x)\big)}{p^\beta(t,x)}\Psi^\beta(t,x)\Big) + \Psi^\beta(t,x)\nabla_x\beta(t,x)\bigg), \tag{3.47}$$

which is the *adjoint* of (3.42). The probability density function of X_t^β with initial distribution p_{t_0} is then

$$\widehat{p}^\beta(t,x) := \int_U G^\beta(t_0, y; t, x)\, p_{t_0}(y)\,\mathrm{d}y. \tag{3.48}$$

Together with (3.12) of Assumption 3.2.1(c), we see that $\widehat{p}^\beta$ satisfies the linear uniformly parabolic

PDE

$$\begin{cases}\partial_t \widehat{p}^\beta(t,x) = \operatorname{div}\Big(\nabla\Big(\dfrac{f\big(p^\beta(t,x)\big)}{p^\beta(t,x)}\widehat{p}^\beta(t,x)\Big) + \widehat{p}^\beta(t,x)\nabla\beta(t,x)\Big), & \text{for } (t,x) \in (t_0, T_\beta] \times U,\\ \widehat{p}^\beta(t_0,x) = p(t_0,x), & \text{for } x \in \overline{U},\\ \dfrac{\partial \widehat{p}^\beta(t,x)}{\partial n(x)} = 0, & \text{for } x \in \partial U.\end{cases}$$

Since the solution to this PDE is unique, we deduce from the Neumann problem in (3.20) and Lemma 3.2.9 that $\widehat{p}^\beta = p^\beta$. □

3.3.3 Proof of Theorems 3.2.5 and 3.2.11

We will only prove Theorem 3.2.11, as similar arguments can be used to show Theorem 3.2.5. We first prove (3.24). Recall the function φ defined in (3.14) and note for later use the simple identities

$$\varphi(u) = h(u) - \frac{f(u)}{u}, \quad h(u) = \varphi'(u)u + \varphi(u), \quad \Phi''(u) = \varphi''(u)u + 2\varphi'(u) \quad \text{for all } u > 0. \tag{3.49}$$

By writing $v^\beta(t,x) = \varphi\big(p^\beta(t,x)\big)$, we deduce from (3.20) that

$$\partial_t v^\beta(t,x) = \varphi'\big(p^\beta(t,x)\big)\partial_t p^\beta(t,x) = \varphi'\big(p^\beta(t,x)\big)\operatorname{div}\Big(\nabla f\big(p^\beta(t,x)\big) + p^\beta(t,x)\nabla\beta(x)\Big). \tag{3.50}$$

Using Itô's lemma along with (3.21) and (3.50), we see that the dynamics of the perturbed entropy process satisfies

$$\begin{aligned} \mathrm{d}v^\beta(t,X_t^\beta) &= \Big(\varphi'(p^\beta)\operatorname{div}\big(\nabla f(p^\beta) + p^\beta\nabla\beta\big) + \frac{f(p^\beta)}{p^\beta}\Delta v^\beta - \langle \nabla v^\beta,\, \nabla\beta\rangle\Big)(t,X_t^\beta)\,\mathrm{d}t \\ &\qquad + \Big\langle \Big(\sqrt{\frac{2f(p^\beta)}{p^\beta}}\nabla v^\beta\Big)(t,X_t^\beta),\, \mathrm{d}B_t\Big\rangle - \big\langle \nabla v^\beta(t,X_t^\beta),\, n(X_t^\beta)\big\rangle\,\mathrm{d}L_t^\beta \\ &= D^\beta(t,X_t^\beta)\,\mathrm{d}t + M_t^\beta - \varphi'\big(p^\beta(t,X_t)\big)\big\langle \nabla p^\beta(t,X_t^\beta),\, n(X_t^\beta)\big\rangle\,\mathrm{d}L_t^\beta. \end{aligned}$$

Note that the last line in (3.13) ensures that the reflection term L^β only increases when X^β is on the boundary. In conjunction with the no-flux boundary condition in (3.20), we see that the last term above is zero. This completes the proof of (3.24).

Next, to see that the local martingale M^β in (3.25) is in fact a true L^2-martingale, note that the quadratic variation of M^β is given by

$$\mathbb{E}\Big[\langle M^\beta,\, M^\beta\rangle_{T_\beta}\Big] = \mathbb{E}\int_{t_0}^{T_\beta}\left(\frac{2f(p^\beta)}{p^\beta}|\nabla v^\beta|^2\right)(t, X_t^\beta)\,\mathrm{d}t = \mathbb{E}\int_{t_0}^{T_\beta}\left(\frac{2f(p^\beta)^3}{(p^\beta)^5}|\nabla p^\beta|^2\right)(t, X_t^\beta)\,\mathrm{d}t.$$

We claim that the above quantity is finite. Indeed, due to the properties of the solution p^β in Lemma 3.2.9, the above expectation is bounded by

$$2(T_\beta - t_0)(2\kappa)^5 f\left(\frac{3}{2\kappa}\right)^3 \max_{[t_0,T_\beta]\times\overline{U}} |\nabla p^\beta(t,x)|^2 < \infty.$$

Therefore, it follows from [106, Corollary IV.1.25] that $\big(M_t^\beta\big)$ is an L^2-martingale.

Finally, in order to show (3.26), we take expectation in the first equation in (3.25) and use Fubini's theorem, to obtain

$$\mathbb{E}[F_t^\beta] = \int_{t_0}^{t}\mathbb{E}\big[D^\beta(s, X_s^\beta)\big]\,\mathrm{d}s = \sum_{i=1}^{3}\int_{t_0}^{t}\mathbb{E}\big[D_i^\beta(s, X_s^\beta)\big]\,\mathrm{d}s, \tag{3.51}$$

where

$$D_1^\beta := \varphi'(p^\beta)\,\mathrm{div}\Big(\nabla f\big(p^\beta\big) + p^\beta\nabla\beta\Big), \quad D_2^\beta := \frac{f(p^\beta)}{p^\beta}\Delta v^\beta, \quad \text{and} \quad D_3^\beta := -\big\langle \nabla v^\beta,\, \nabla\beta\big\rangle.$$

We evaluate now each of the expectations in (3.51). Integrating by parts, we have

$$\begin{aligned}\mathbb{E}\big[D_1^\beta(t, X_t^\beta)\big] &= \int_U \varphi'(p^\beta)p^\beta\,\mathrm{div}\Big(\nabla f\big(p^\beta\big) + p^\beta\nabla\beta\Big)(t,x)\,\mathrm{d}x \\ &= -\int_U \Big\langle \nabla\Big(\varphi'(p^\beta)p^\beta\Big), \nabla f\big(p^\beta\big) + p^\beta\nabla\beta\Big\rangle(t,x)\,\mathrm{d}x + C,\end{aligned} \tag{3.52}$$

where C is the boundary term given by

$$C := \int_{\partial U} \varphi'(p^\beta) f'(p^\beta) p^\beta \Big\langle \nabla p^\beta + p^\beta \nabla \beta, n \Big\rangle(t,x)\,\mathrm{d}x.$$

From the no-flux boundary condition in (3.20) and Assumption 3.2.8, we see that $C = 0$. Similarly, we see that $\mathbb{E}\big[D_2^\beta(t, X_t^\beta)\big]$ and $\mathbb{E}\big[D_3^\beta(t, X_t^\beta)\big]$ are respectively equal to

$$-\int_U \big\langle \nabla f(p^\beta), \nabla v^\beta \big\rangle(t,x)\,\mathrm{d}x, \quad \text{and} \quad -\int_U \big\langle \nabla v^\beta, \nabla \beta \big\rangle(t,x)\,\mathrm{d}x. \tag{3.53}$$

Assembling them gives

$$\begin{aligned} \mathbb{E}[D^\beta(t, X_t^\beta)] = -\int_U \Big\langle \nabla\big(\varphi'(p^\beta)p^\beta + v^\beta\big), \nabla f(p^\beta) \Big\rangle(t,x)\,\mathrm{d}x \\ - \int_U \Big\langle \nabla\big(\varphi'(p^\beta)p^\beta\big) + \nabla v^\beta,\, p^\beta \nabla \beta \Big\rangle(t,x)\,\mathrm{d}x. \end{aligned} \tag{3.54}$$

Using the third identity in (3.49), we see that the first integrand above is equal to

$$\big(\varphi''(p^\beta)p^\beta + 2\varphi'(p^\beta)\big) f'(p^\beta) |\nabla p^\beta|^2 = \Phi''(p^\beta) f'(p^\beta) |\nabla p^\beta|^2,$$

so the first integral in (3.54) is

$$-\int_U \Big(\big|\Phi''(p^\beta)\nabla p^\beta\big|^2 p^\beta \Big)(t,x)\,\mathrm{d}x = -I\big(p_t^\beta\big) > -\infty,$$

where the last step follows from the boundedness of p^β and ∇p^β implied by Lemma 3.2.9. Similarly, using the second identity in (3.49), we see that the second integral in (3.54) is equal to

$$-\mathbb{E}\Big[\big\langle \nabla\big(h(p^\beta)\big),\, \nabla\beta \big\rangle(t, X_t^\beta)\Big] > -\infty.$$

Putting them together completes the proof of (3.26). □

3.3.4 Proofs of Corollaries 3.2.6 and 3.2.12

We will only prove Corollary 3.2.12, as the proof of Corollary 3.2.6 proceeds in the same way.

Taking expectation in (3.24) and using the martingale property of M^β in (3.25) as well as (3.26), we have

$$\begin{aligned}\mathscr{F}\left(p_t^\beta\right) - \mathscr{F}\left(p_{t_0}^\beta\right) &= \mathbb{E}\Big[v^\beta(t, X_t^\beta) - v^\beta(t_0, X_{t_0}^\beta)\Big] \\ &= -\int_{t_0}^t I(p_s^\beta)\,\mathrm{d}s - \int_{t_0}^t \mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(s, X_s^\beta)\big),\, \nabla\beta(X_s^\beta)\big\rangle\Big]\,\mathrm{d}s,\end{aligned} \tag{3.55}$$

which proves (3.27).

Turning to the proof of (3.28), note that from (3.2) we have

$$\mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(s, X_s^\beta)\big),\, \nabla\beta(X_s^\beta)\big\rangle\Big] = \int_U f'\big(p^\beta(s,x)\big)\big\langle \nabla p^\beta(s,x), \nabla\beta(x)\big\rangle\,\mathrm{d}x. \tag{3.56}$$

From the continuity of $(s,x) \mapsto f'\big(p^\beta(s,x)\big)\big\langle \nabla p^\beta(s,x), \nabla\beta(x)\big\rangle$, we see that the expression of (3.56) is continuous as a function of s, thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0^+} \int_{t_0}^t \mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(s, X_s^\beta)\big),\, \nabla\beta(X_s^\beta)\big\rangle\Big]\mathrm{d}s = \mathbb{E}\Big[\big\langle \nabla h\big(p^\beta(t_0, X_{t_0})\big),\, \nabla\beta(X_{t_0})\big\rangle\Big], \tag{3.57}$$

where the last equality is due to the fact that $X_{t_0}^\beta = X_{t_0}$ by construction in (3.20). Similarly, from the continuity of $t \mapsto I\big(p_t^\beta\big)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=t_0^+} \int_{t_0}^t I\big(p_u^\beta\big)\,\mathrm{d}u = I\big(p_{t_0}^\beta\big) = I(p_{t_0}), \tag{3.58}$$

and the identity (3.28) follows.

Finally, to show (3.29), the martingale property of M^β implies that the numerator on the left-hand side of (3.29) is equal to

$$\mathbb{E}\Big[F_t^\beta - F_{t_0}^\beta \,\big|\, \mathcal{F}_{t_0}\Big] = \mathbb{E}\left[\int_{t_0}^t D^\beta(u, X_u^\beta)\,\mathrm{d}u \,\bigg|\, \mathcal{F}_{t_0}\right].$$

From the continuity of $u \mapsto D^\beta(u, X_u)$, we have

$$\lim_{t \downarrow t_0} \frac{1}{t - t_0} \int_{t_0}^{t} D^\beta(u, X_u^\beta)\, \mathrm{d}u = D^\beta(t_0, X_{t_0}), \quad \text{a.s.}.$$

Moreover, the properties of p^β from Lemma 3.2.9 implies that D^β is uniformly bounded on $[t_0, T_\beta] \times \overline{U}$. Therefore, by the bounded convergence theorem,

$$\lim_{t \downarrow t_0} \mathbb{E}\left[\frac{1}{t - t_0} \int_{t_0}^{t} D^\beta(u, X_u^\beta) \mathrm{d}u\right] = \mathbb{E}\big[D^\beta(t_0, X_{t_0})\big].$$

We now apply Scheffé's lemma to obtain

$$\lim_{t \downarrow t_0} \left\| \frac{1}{t - t_0} \int_{t_0}^{t} D^\beta(u, X_u^\beta) \mathrm{d}u - D^\beta(t_0, X_{t_0}) \right\|_{L^1} = 0.$$

To complete the proof, use Jensen's inequality and the tower property to get

$$\begin{aligned}
&\left\| \mathbb{E}\left[\frac{1}{t - t_0} \int_{t_0}^{t} D^\beta(u, X_u^\beta)\, \mathrm{d}u \,\middle|\, \mathcal{F}_{t_0}\right] - D^\beta(t_0, X_{t_0}) \right\|_{L^1} \\
&\qquad\qquad \leq \left\| \frac{1}{t - t_0} \int_{t_0}^{t} D^\beta(u, X_u^\beta)\, \mathrm{d}u - D^\beta(t_0, X_{t_0}) \right\|_{L^1},
\end{aligned}$$

and the L^1-convergence in (3.29) follows. □

3.3.5 Proof of Lemma 3.2.13

Since the proofs of (3.30) and (3.31) are very similar, we will only prove (3.31). We first rewrite the PDE in (3.20) as a continuity equation

$$\partial_t p^\beta(t, x) + \operatorname{div}\Big(p^\beta(t, x)\, \widetilde{v}^\beta(t, x)\Big) = 0, \quad \text{for } (t, x) \in (t_0, T_\beta) \times U, \tag{3.59}$$

where $\widetilde{v}^\beta : [t_0, T_\beta] \times \overline{U} \to \mathbb{R}^d$ is the *velocity field*, defined by

$$\widetilde{v}^\beta(t,x) := -\nabla \left[\beta + h\big(p^\beta\big)\right](t,x). \tag{3.60}$$

We see from Assumption 3.2.8 and Lemma 3.2.9 that $\widetilde{v}^\beta(t_0,\cdot)$ is the gradient of a smooth function. For each $x \in \overline{U}$, consider the curved flow Λ_t^β associated with $\widetilde{v}^\beta$, specified by

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda_t^\beta(x) = \widetilde{v}^\beta\big(t, \Lambda_t^\beta(x)\big), \quad t \in [t_0, T_\beta], \quad \Lambda_{t_0}^\beta(x) = x. \tag{3.61}$$

By the Cauchy-Lipschitz theorem, there exists a unique solution $t \mapsto \Lambda_t^\beta(x) \in \overline{U}$ to (3.61). Moreover, it follows from [99, Theorem 5.34] that Λ_t^β pushes forward $p_{t_0}^\beta$ to p_t^β, in the sense that $p_{t_0}^\beta \circ (\Lambda_t^\beta)^{-1} = p_t^\beta$. Note also that the Jacobian of Λ_t^β is given by

$$\nabla\Lambda_t^\beta(x) = I - \int_{t_0}^t \nabla^2\left[\big(\beta + h(p^\beta)\big)\big(s, \Lambda_s^\beta(x)\big)\right]\mathrm{d}s.$$

Therefore, by setting

$$K := \max\left\{\left|\partial_{ij}\big(\beta + h(p^\beta)\big)(t,x)\right| : t \in [t_0, T_\beta], x \in \overline{U}, i,j = 1,\dots,n\right\},$$

we see that for any $t \in [t_0, t_0 + K^{-1})$, the Jacobian $\nabla\Lambda_t^\beta(x)$ is positive-semidefinite for all $x \in \overline{U}$, so Λ_t^β is the gradient of a convex function. Hence, by Brenier's theorem [99, Theorem 2.12(ii)], Λ_t^β is the optimal transport map from $p_{t_0}^\beta$ to p_t^β, i.e.,

$$\mathcal{W}_2^2\big(p_{t_0}^\beta, p_t^\beta\big) = \mathbb{E}\left[\left|\Lambda_t^\beta(X_{t_0}^\beta) - X_{t_0}\right|^2\right] = \mathbb{E}\left[\left|\int_{t_0}^t \widetilde{v}^\beta\big(s, \Lambda_s^\beta(X_{t_0}^\beta)\big)\,\mathrm{d}s\right|^2\right].$$

Now, the continuity of $t \mapsto \widetilde{v}^\beta\big(t, \Lambda_t^\beta(x)\big)$ implies

$$\left|\frac{1}{t - t_0}\int_{t_0}^t \widetilde{v}^\beta\big(s, \Lambda_s^\beta(X_{t_0}^\beta)\big)\,\mathrm{d}s\right|^2 \xrightarrow{t \downarrow t_0} \left|\widetilde{v}^\beta(t_0, X_{t_0})\right|^2, \quad \text{a.s.} \tag{3.62}$$

Moreover, by Jensen's inequality, the random variable on the left-hand side above is bounded by

$$\frac{1}{t-t_0}\int_{t_0}^{t}\left|\widetilde{v}^{\beta}(s,\Lambda_s^{\beta}(X_{t_0}^{\beta}))\right|^2\mathrm{d}s \le \max\left\{|\widetilde{v}^{\beta}(t,x)|^2 : t\in[t_0,T_\beta], x\in\overline{U}\right\}<\infty,$$

where the last step follows from the aforementioned fact that $\widetilde{v}^{\beta}(t,\cdot)$ is the gradient of a smooth function. Consequently, by the bounded convergence theorem,

$$\lim_{t\downarrow t_0}\frac{\mathcal{W}_2(p_{t_0}^{\beta},p_t^{\beta})}{t-t_0}=\lim_{t\downarrow t_0}\left\|\frac{1}{t-t_0}\int_{t_0}^{t}\widetilde{v}^{\beta}\big(s,\Lambda_s^{\beta}(X_{t_0}^{\beta})\big)\,\mathrm{d}s\right\|_{L^2}=\left\|\widetilde{v}^{\beta}(t_0,X_{t_0})\right\|_{L^2}, \tag{3.63}$$

where the last step follows from the fact that $X_{t_0}^{\beta}=X_{t_0}$ by construction in (3.20). Recalling the expression of $\widetilde{v}^{\beta}$ in (3.60), we arrive at (3.31). □

3.3.6 Proof of Theorem 3.2.14

The identity (3.33) follows from (3.18) of Corollary 3.2.6 and (3.30) of Lemma 3.2.13. Similarly, for (3.34), we deduce from (3.28) of Corollary 3.2.12 and (3.31) of Lemma 3.2.13 that

$$|\partial\mathscr{F}|_{\mathcal{W}_2}(p_{t_0}^{\beta})=-\frac{I(p_{t_0})+\mathbb{E}\Big[\big\langle\nabla h\big(p(t_0,X_{t_0})\big),\,\nabla\beta(X_{t_0})\big\rangle\Big]}{\Big\|\nabla h\big(p(t_0,X_{t_0})\big)+\nabla\beta(X_{t_0})\Big\|_{L^2}}. \tag{3.64}$$

Moreover, we see from (3.5) and (3.2) that the entropy dissipation functional can be expressed as

$$I(p_{t_0})=\mathbb{E}\Big[\big|\nabla h\big(p(t_0,X_{t_0})\big)\big|^2\Big].$$

Putting this back into (3.64) yields (3.34). Finally, the inequality (3.35) follows from the Cauchy-Schwarz inequality. □

3.3.7 Proof of Proposition 3.2.17

For $t \in [0,1)$, let us denote by

$$T_t \coloneqq (1-t)\mathrm{Id} + t\nabla\psi \tag{3.65}$$

the optimal transport map from ρ_0 to ρ_t. Note that T_t is injective by [99, Theorem 5.49], so its inverse exists. By [99, Theorem 5.51(ii)], the probability density functions $(\rho_t)_{t\in(0,1)}$ satisfy the continuity equation

$$\partial_t\rho_t(x) + \mathrm{div}\big(\rho_t(x)\,\widehat{v}_t(x)\big) = 0, \tag{3.66}$$

where $\widehat{v} : [0,1) \times \overline{U} \to \mathbb{R}^d$ is the velocity field defined by

$$\widehat{v}_t(x) \coloneqq (\nabla\psi - \mathrm{Id}) \circ (T_t)^{-1}(x). \tag{3.67}$$

In conjunction with (3.65), we see that T_t satisfies the integral equation

$$T_t(x) = x + \int_0^t \widehat{v}_s\big(T_s(x)\big)\mathrm{d}s. \tag{3.68}$$

We now switch to probabilistic notations. On a sufficiently rich probability space, let Z_0 be a random variable with distribution ρ_0 and let $Z_t \coloneqq T_t(Z_0)$ for $t \in (0,1)$. On account of (3.68), we have

$$Z_t = Z_0 + \int_0^t \widehat{v}_s(Z_s)\,\mathrm{d}s. \tag{3.69}$$

Together with (3.66), we deduce

$$\mathrm{d}\rho_t(Z_t) = \partial_t\rho_t(Z_t)\,\mathrm{d}t + \big\langle \nabla\rho_t(Z_t), \mathrm{d}Z_t\big\rangle = -\rho_t(Z_t)\,\mathrm{div}\big(\widehat{v}_t(Z_t)\big)\mathrm{d}t.$$

Recalling the function φ in (3.14), we have

$$\mathrm{d}\varphi\big(\rho_t(Z_t)\big) = -\varphi'\big(\rho_t(Z_t)\big)\rho_t(Z_t)\,\mathrm{div}\big(\widehat{v}_t(Z_t)\big)\mathrm{d}t = -\frac{f\big(\rho_t(Z_t)\big)}{\rho_t(Z_t)}\mathrm{div}\big(\widehat{v}_t(Z_t)\big)\mathrm{d}t,$$

where the last step follows from the identity $f(u) = uh(u) - \Phi(u)$, valid for all $u \geq 0$. Integrating from 0 to t and taking expectation yield

$$\begin{aligned}\mathscr{F}(\rho_t) - \mathscr{F}(\rho_0) &= -\int_0^t \int_U f\big(\rho_s(z)\big)\mathrm{div}\big(\widehat{v}_s(z)\big)\mathrm{d}z\mathrm{d}s \\ &= \int_0^t \int_U \big\langle \nabla f\big(\rho_s(z)\big), \widehat{v}_s(z)\big\rangle\,\mathrm{d}z\mathrm{d}s - \int_0^t \int_{\partial U} f\big(\rho_s(z)\big)\big\langle \widehat{v}_s(z), n(z)\big\rangle\,\mathrm{d}z\,\mathrm{d}s.\end{aligned} \tag{3.70}$$

It follows from Assumption 3.2.15(b) that $\nabla\psi(x) = x$ for all $x \in \partial U$. Therefore, we see from (3.67) that $\widehat{v}_t(x) = 0$ for all $x \in \partial U$. Hence, the last integral in (3.70) vanishes. Letting $t \downarrow 0$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} \mathscr{F}(\rho_t) = \int_U \big\langle \nabla f\big(\rho_0(z)\big), \widehat{v}_0(z)\big\rangle\,\mathrm{d}z$$

Recalling the definition of $\widehat{v}_0$ in (3.67) completes the proof. □

3.3.8 Proof of Theorem 3.2.18

From [99, Theorem 5.15(i)], Assumption 3.2.15(c) implies that the entropy functional $\mathscr{F}$ is *displacement convex*. In other words,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathscr{F}(\rho_t) \geq 0, \quad \text{for} \quad t \in [0,1].$$

Therefore, Taylor's theorem and Proposition 3.2.17 give us

$$\begin{aligned}\mathscr{F}(\rho_1) &= \mathscr{F}(\rho_0) + \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} \mathscr{F}(\rho_t) + \int_0^1 (1-t)\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathscr{F}(\rho_t)\,\mathrm{d}t \\ &\geq \mathscr{F}(\rho_0) + \int_U \big\langle \nabla f\big(\rho_0(z)\big), \nabla\psi(z) - z\big\rangle\,\mathrm{d}z,\end{aligned}$$

which proves the first inequality in (3.38). The second inequality is a simple consequence of the Cauchy-Schwarz inequality; the expression in the middle of (3.38) is bounded from above by

$$\sqrt{\int_U \frac{\big|\nabla f\big(\rho_0(z)\big)\big|^2}{\rho_0(z)}\mathrm{d}z}\sqrt{\int_U \big|\nabla\psi(z) - z\big|^2 \rho_0(z)dz} = \sqrt{I(\rho_0)}\,\mathcal{W}_2(\rho_0, \rho_1).$$

□

Chapter 4: Mean field approximations via log concavity

In this chapter, we propose a new approach to deriving quantitative mean field approximations for any probability measure P on $\mathbb{R}^n$ with density proportional to $e^{f(x)}$, for f strongly concave. We bound the mean field approximation for the log partition function $\log \int_{\mathbb{R}^n} e^{f(x)} \mathrm{d}x$ in terms of $\sum_{i \neq j} \mathbb{E}_{Q^*} |\partial_{ij} f|^2$, for a semi-explicit probability measure Q^* characterized as the unique mean field optimizer, or equivalently as the minimizer of the relative entropy $H(\cdot \mid P)$ over product measures. This notably does not involve metric-entropy or gradient-complexity concepts which are common in prior work on nonlinear large deviations. Three implications are discussed, in the contexts of continuous Gibbs measures on large graphs, high-dimensional Bayesian linear regression, and the construction of decentralized near-optimizers in high-dimensional stochastic control problems. Our arguments are based primarily on functional inequalities and the notion of displacement convexity from optimal transport.

4.1 Introduction

At the center of the recent theory of *nonlinear large deviations* is the problem of justifying the *mean field approximation* for the partition function of a Gibbs measure. Given a (reference) probability measure μ on $\mathbb{R}$, suppose a probability measure P on $\mathbb{R}^n$ takes the form

$$P(dx) = Z^{-1} e^{f(x)} \mu^{\otimes n}(\mathrm{d}x),$$

for a function $f : \mathbb{R}^n \to \mathbb{R}$ and normalizing constant Z, where $\mu^{\otimes n}$ denotes the n-fold product measure. A recurring problem in diverse applications is the approximation of the often intractable *partition function* Z. It obeys the well-known Gibbs variational principle

$$\log Z = \log \int_{\mathbb{R}^n} e^f \,\mathrm{d}\mu^{\otimes n} = \sup_{Q \in \mathcal{P}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f \,\mathrm{d}Q - H(Q \,|\, \mu^{\otimes n}) \right), \tag{4.1}$$

where $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures on $\mathbb{R}^n$, and H denotes the relative entropy

$$H(Q \,|\, Q') \coloneqq \int_{\mathbb{R}^n} \frac{\mathrm{d}Q}{\mathrm{d}Q'} \log \frac{\mathrm{d}Q}{\mathrm{d}Q'} \,\mathrm{d}Q' \ \text{ if } Q \ll Q', \qquad H(Q \,|\, Q') \coloneqq \infty \ \text{ if } Q \not\ll Q'.$$

Note that $Q = P$ is the unique optimizer in (4.1). Letting $\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ denote the set of product measures $Q = Q_1 \times \cdots \times Q_n$ in $\mathcal{P}(\mathbb{R}^n)$, the *mean field approximation* is

$$\log \int_{\mathbb{R}^n} e^f \,\mathrm{d}\mu^{\otimes n} \approx \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f \,\mathrm{d}Q - H(Q \,|\, \mu^{\otimes n}) \right). \tag{4.2}$$

In the cases studied in this chapter, the left-hand side is expected to be of order n; a precise formulation of (4.2) is then to find conditions under which the difference is $o(n)$, so that the mean field approximation becomes asymptotically correct at the leading order. Note that the right-hand side of (4.2) is trivially a lower bound for the left, because of (4.1), and it is only the upper bound that incurs an error which must be estimated.

The groundbreaking work of [24], motivated by applications to subgraph counts in sparse random graphs, showed how to justify the mean field approximation in the case that $\mu^{\otimes n}$ is the uniform measure on the hypercube $\{-1, 1\}^n$. Their key assumption is that the gradient of f has *low complexity*, as measured by the metric entropy of the range $\nabla f(\{-1, 1\}^n)$. A number of subsequent papers have since refined this approach and results on subgraph counts [25, 26, 27], in addition to other noteworthy applications such as Ising models [28, 34, 29, 30, 31, 125]. Most applications thus far involve discrete μ, but the theory has been extended to compactly supported measures [28, 126, 127]. Alternative and often more convenient estimates have appeared, still based on "gradient complexity" but quantifying it in a different way, eschewing covering number estimates in favor of the simpler and weaker Gaussian-width [31, 32, 33] or Rademacher-width [34].

In this chapter, we propose an alternative approach to the mean field approximation, designed most notably for the case where f is concave and the reference measure μ is strongly log-concave (see Theorem 4.1.1 and Corollary 4.1.4). In particular, we deal with continuous μ of unbounded support, which covers a rather different host of applications, discussed in Chapter 4.2, compared to the somewhat more discrete-oriented prior literature. Our approach is based on a semi-explicit representation for the mean field optimizer Q^* in (4.2), which we show to be unique as soon as P is strictly log-concave, and which is in fact also the unique minimizer of $H(\cdot\,|\,P)$ over product measures. We control the error in the approximation (4.2) by a constant times $\mathbb{E}_{Q^*}\sum_{i\neq j}|\partial_{ij}f|^2$, which is typically much simpler to work with compared to the aforementioned notions of gradient complexity. Eldan [31, 32] and Austin [126] also analyze the mean field approximation by approximating P by product measures in entropy, but our methods and bounds are very different from theirs; notably, they approximate P not by a single product measure but by a mixture, which is natural when the mean field optimizer is not unique, as is explained well in [31]. The uniqueness of the mean field optimizer in our setting means that we expect P to concentrate around a single *pure state*, rather than a mixture of states.

In the rest of this section, we describe our general results on mean field approximations for log-concave measures, along with some related ideas and generalizations, with proofs deferred to Chapter 4.3. Chapter 4.2 develops three applications: Gibbs measures with heterogeneous interactions, high-dimensional Bayesian linear regression, and high-dimensional stochastic control problems.

4.1.1 Main results

Recall for $\kappa \in \mathbb{R}$ that a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ is said to be *κ-concave* if $x \mapsto f(x) + \frac{\kappa}{2}|x|^2$ is concave. If f is finite-valued and C^2, i.e., twice continuously differentiable, then f is κ-concave if and only if $\nabla^2 f(x) \leq -\kappa I$ in semidefinite order, for each $x \in \mathbb{R}^n$. We say that a probability measure P on $\mathbb{R}^n$ is κ-log-concave if it takes the form $P(\mathrm{d}x) = e^{f(x)}\,\mathrm{d}x$ for some

κ-concave function f. We will work with the (negative of the) differential entropy

$$H(Q) := \int_{\mathbb{R}^n} Q(x) \log Q(x) \, \mathrm{d}x,$$

for an absolutely continuous probability measure $Q(dx) = Q(x)\mathrm{d}x$ on a Euclidean space, well-defined in $(-\infty, \infty]$ whenever the negative part of $Q \log Q$ is integrable; we adopt the convention that $H(Q) = \infty$ if Q is not absolutely continuous, or if $(Q \log Q)^-$ is not integrable. Let $\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ denote the set of product measures on $\mathbb{R}^n$. Let $X = (X_1, \ldots, X_n) : \mathbb{R}^n \to \mathbb{R}^n$ denote the identity map, so that we may write $\mathbb{E}_Q[g(X)] = \int_{\mathbb{R}^n} g \, \mathrm{d}Q$ for the expectation under Q.

Theorem 4.1.1. *Consider a C^2 and κ-log-concave probability measure $P(\mathrm{d}x) = Z^{-1} e^{f(x)} \, \mathrm{d}x$, for some $\kappa > 0$. Assume there exist $c_1 \geq 0$ and $0 \leq c_2 < \kappa/2$ such that $|f(x)| \leq c_1 e^{c_2 |x|^2}$ for all $x \in \mathbb{R}^n$. Then the following conclusions hold:*

(1) There exists a unique product measure $Q^* = Q_1^* \times \cdots \times Q_n^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ with strictly positive density a.e. satisfying $f \in L^1(Q^*)$ and the fixed point equation

$$Q_i^*(\mathrm{d}x_i) = Z_i^{-1} \exp\left(\mathbb{E}_{Q^*}[f(X) \,|\, X_i = x_i]\right) \mathrm{d}x_i, \qquad Z_i > 0, \ i = 1, \ldots, n. \tag{4.3}$$

(2) Q^* is κ-log-concave.

(3) Q^* is the unique optimizer in

$$\sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}Q - H(Q) \right). \tag{4.4}$$

(4) If we define

$$R_f := \log \int_{\mathbb{R}^n} e^{f(x)} \, \mathrm{d}x - \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}Q - H(Q) \right),$$

then

$$0 \leq R_f \leq \frac{1}{2\kappa}\mathbb{E}_{Q^*}\sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i) \leq \frac{1}{\kappa^2}\sum_{1\leq i<j\leq n}\mathbb{E}_{Q^*}[|\partial_{ij}f(X)|^2]. \tag{4.5}$$

The supremum in (4.4) is finite, as we will see in Lemma 4.3.4. Also, as will be seen in the proof of Proposition 4.3.9, our assumptions ensure that $\partial_i f(x_i,\cdot) \in L^1\big(\prod_{j\neq i} Q_j^*\big)$, so the conditional variance in (4.5) is well-defined in $[0,\infty]$. The final quantity in (4.5) controlling our mean field approximation error involves only the cross-derivatives $i \neq j$, which are insensitive to additively separable perturbations $f(x) \to f(x) + \sum_{i=1}^n \tilde{f}_i(x_i)$. On the other hand, the measure Q^* is sensitive to these perturbations, but in the tractable sense that $Q_i^*(\mathrm{d}x_i)$ must be multiplied by $\exp \tilde{f}_i(x_i)$ (and a new normalizing constant). In particular, both upper bounds in (4.5) vanish if f is already additively separable, i.e., if P is a product measure.

In Theorem 4.1.1, the measure Q^* is defined implicitly, which can make bounding R_f difficult. In the simplest case where $\nabla^2 f$ is bounded, we need no knowledge of Q^* to obtain

$$R_f \leq \frac{1}{\kappa^2}\sup_{x\in\mathbb{R}^n}\sum_{1\leq i<j\leq n}|\partial_{ij}f(x)|^2,$$

which is sharp enough for many applications. But even when $\nabla^2 f$ is unbounded, we can take advantage of the fact that Q^* is κ-log-concave by Theorem 4.1.1(2), which implies in particular that it has finite moments of all orders controlled in terms of κ.

A guiding example is the class of Gibbs measures with pairwise interactions of the form

$$f(x) = \sum_{i=1}^n V(x_i) + \sum_{1\leq i<j\leq n} J_{ij}K(x_i - x_j), \tag{4.6}$$

where V is κ-concave, K is even and concave, and J is a symmetric matrix with nonnegative entries. Then $\partial_{ij}f(x) = -J_{ij}K''(x_i - x_j)$ for $i \neq j$, and for K'' bounded we immediately deduce $R_f \leq \mathrm{Tr}(J^2)\|K''\|_\infty^2/2\kappa^2$ from Theorem 4.1.1. Corollary 4.2.3 below proves a similar $O(\mathrm{Tr}(J^2))$ bound merely assuming that K'' has at most exponential growth, plus a symmetry assumption.

Since $\log \int_{\mathbb{R}^n} e^f \mathrm{d}x$ is order n in this case, we obtain a successful mean field approximation whenever J satisfies $\mathrm{Tr}(J^2) = o(n)$, which is, in a sense, optimal. For instance, in the noteworthy case that J is $1/d$ times the adjacency matrix of a d-regular graph, we have $\mathrm{Tr}(J^2) = o(n)$ precisely when $d \to \infty$, and the mean field approximation fails in general in the sparsest case where $d = O(1)$. See Chapter 4.2.1 for further discussion.

As a first corollary of Theorem 4.1.1, we deduce the following non-asymptotic law of large numbers for the empirical measure.

Corollary 4.1.2. *Under the assumptions of Theorem 4.1.1, for any 1-Lipschitz function $\varphi : \mathbb{R} \to \mathbb{R}$, we have*

$$\mathbb{E}_P\left[\left(\frac{1}{n}\sum_{i=1}^n \varphi(X_i) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{Q^*}[\varphi(X_i)]\right)^2\right] \le \frac{(1+\sqrt{2R_f})^2}{\kappa n}. \tag{4.7}$$

Remark 4.1.3. Corollary 4.1.2 can be interpreted as a form of concentration of the empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ around the measure $\frac{1}{n}\sum_{i=1}^n Q_i^*$. Alternatively, the Poincaré inequality for P implies $\mathrm{Var}_P(\frac{1}{n}\sum_{i=1}^n \varphi(X_i)) \le 1/\kappa n$ for 1-Lipschitz φ, which in turn implies a form of concentration of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ around its mean $\frac{1}{n}\sum_{i=1}^n P_i$, where P_i is the i^{th} marginal of P. However, the latter is normally not as useful, because the marginals of P are typically not as tractable as the various characterizations of Q^* provided by Theorem 4.1.1.

It is often convenient to work with a probability measure as a reference measure, in place of Lebesgue measure, as is common in the literature on mean field approximations (see for example [28, 126, 24, 31, 127]). Theorem 4.1.1 implies a similar result in terms of reference probability measures.

Corollary 4.1.4. *Let $V_i : \mathbb{R} \to \mathbb{R}$ be C^2 and κ-concave for some $\kappa > 0$, such that $\rho_i(\mathrm{d}x) = e^{V_i(x)}\,\mathrm{d}x$ is a probability measure, for $i = 1, \ldots, n$. Let $\rho = \rho_1 \times \cdots \times \rho_n$. Let $g : \mathbb{R}^n \to \mathbb{R}$ be C^2 and concave. Assume there exist $c_1 \ge 0$ and $0 \le c_2 < \kappa/2$ such that $|g(x)| \le c_1 e^{c_2|x|^2}$ for all $x \in \mathbb{R}^n$. Then the following conclusions hold:*

(1) There exists a unique product measure $Q^* = Q_1^* \times \cdots \times Q_n^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ with strictly positive

density a.e. satisfying $g \in L^1(Q^*)$ *and*

$$Q_i^*(\mathrm{d}x_i) = Z_i^{-1} \exp\left(\mathbb{E}_{Q^*}[g(X) \,|\, X_i = x_i]\right) \rho_i(\mathrm{d}x_i), \qquad Z_i > 0,\ i = 1, \dots, n. \tag{4.8}$$

(2) Q^* *is* κ*-log-concave.*

(3) Q^* *is the unique optimizer in*

$$\sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g \,\mathrm{d}Q - H(Q \,|\, \rho) \right). \tag{4.9}$$

(4) If we define

$$R_g^\rho := \log \int_{\mathbb{R}^n} e^g \,\mathrm{d}\rho - \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g \,\mathrm{d}Q - H(Q \,|\, \rho) \right),$$

then

$$0 \le R_g^\rho \le \frac{1}{2\kappa} \mathbb{E}_{Q^*} \sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i g(X) \,|\, X_i) \le \frac{1}{\kappa^2} \sum_{1 \le i < j \le n} \mathbb{E}_{Q^*}[|\partial_{ij} g(X)|^2]. \tag{4.10}$$

For certain symmetric choices of g, the bound (4.10) is related to the theorems of Cramér and Sanov on large deviations, which are settings in which the Gibbs variational principle is well known to be nearly saturated by product measures. For instance, if $g(x) = nG\big(\frac{1}{n}\sum_{k=1}^n x_k\big)$ for some continuous concave G, we obtain $R_g^\rho \le \|G''\|_\infty^2/2\kappa^2$, which is certainly $o(n)$ when G'' is bounded.

4.1.2 Overview and proof ideas

We explain here some key ideas behind Theorem 4.1.1 and its corollaries. The simple identity

$$\log \int_{\mathbb{R}^n} e^{f(x)} \,\mathrm{d}x - \int_{\mathbb{R}^n} f \, dQ + H(Q) = H(Q \,|\, P) \tag{4.11}$$

is valid for probability measures Q with finite entropy and implies (see Lemma 4.3.4 for details)

$$\log \int_{\mathbb{R}^n} e^{f(x)}\,\mathrm{d}x - \sup_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q) \right) = \inf_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} H(Q\,|\,P), \tag{4.12}$$

and also that optimizing (4.4) is equivalent to optimizing

$$\inf_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} H(Q\,|\,P). \tag{4.13}$$

That is, Q^* from Theorem 4.1.1 is the optimizer in (4.13). This can be seen as an *entropic projection*, in the sense of Csiszar [128], onto the set of product measures. A minimizer in (4.13) always exists, because the set of product measures is weakly closed and $H(\cdot\,|\,P)$ has weakly compact sub-level sets. But uniqueness is not obvious and in fact fails in general, because the set of product measures is not convex. We establish the uniqueness of the optimizer in Lemma 4.3.6 in the case where P is strictly log-concave, by exploiting the notion of displacement convexity from the theory of optimal transport, with similarities to the work of McCann [129].

Once we know that the optimizer Q^* for (4.4) takes the form (4.3), the proof of the mean field approximation (4.5) is fairly quick, if we ignore certain technical points: The right-hand side of the identity (4.12) is precisely $H(Q^*\,|\,P)$. We first use the log-Sobolev inequality for P, which is ensured by κ-log-concavity and the famous result of Bakry-Émery [130], to get

$$H(Q^*\,|\,P) \le \frac{1}{2\kappa} \int_{\mathbb{R}^n} \left| \nabla \log \frac{dQ^*}{dP} \right|^2 \mathrm{d}Q^*.$$

Since $Q^* = Q_1^* \times \cdots \times Q_n^*$ is a product measure, the formula (4.3) implies

$$\partial_i \log Q^*(x) = \partial_i \log Q_i^*(x_i) = \partial_i \mathbb{E}_{Q^*}[f(X)\,|\,X_i = x_i] = \mathbb{E}_{Q^*}[\partial_i f(X)\,|\,X_i = x_i]. \tag{4.14}$$

Thus,

$$H(Q^* \,|\, P) \le \frac{1}{2\kappa}\mathbb{E}_{Q^*}\sum_{i=1}^n \left(\mathbb{E}_{Q^*}[\partial_i f(X) \,|\, X_i] - \partial_i f(X)\right)^2 = \frac{1}{2\kappa}\mathbb{E}_{Q^*}\sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i).$$

Differentiating (4.14) again shows easily that Q^* is κ-log-concave since f is concave. Hence, Q^* and its marginals obey a Poincaré inequality, and we deduce

$$\mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i) \le \frac{1}{\kappa}\sum_{j \ne i}\mathbb{E}_{Q^*}\left[|\partial_{ij} f(X)|^2 \,|\, X_i\right].$$

Combining the last two inequalities yields (4.5). See the part titled "Generalization of the main theorem" in Chapter 4.1.3 below for a discussion of a generalization of this argument beyond the strongly log-concave case.

The proof of Corollary 4.1.2 begins with the observation that the κ-log-concavity of P in Theorem 4.1.1 implies the quadratic transport inequality [131, Theorems 1 and 2]

$$\mathcal{W}_2^2(Q^*, P) \le \frac{2}{\kappa} H(Q^* \,|\, P), \tag{4.15}$$

where $\mathcal{W}_2$ denotes the quadratic Wasserstein distance defined by

$$\mathcal{W}_2^2(Q^*, P) = \inf_\pi \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \,\pi(\mathrm{d}x, \mathrm{d}y),$$

where the infimum is over $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals Q^* and P. Combining (4.15) with the inequality $H(Q^* \,|\, P) \le R_f$ discussed above, we arrive at $\mathcal{W}_2^2(Q^*, P) \le 2R_f/\kappa$. The quadratic Wasserstein distance enjoys a useful and fairly well known *subadditivity* inequality, which we prove in Chapter 4.3.4 for the sake of completeness: If P_S denotes the marginal law of $(X_i)_{i \in S}$ under P for a set $S \subset [n] := \{1, \dots, n\}$, and similarly for Q^*_S, then we have

$$\binom{n}{k}^{-1} \sum_{S \subset [n],\, |S| = k} \mathcal{W}_2^2(Q^*_S, P_S) \le \frac{1}{\lfloor n/k \rfloor}\mathcal{W}_2^2(Q^*, P) \le \frac{2}{\kappa \lfloor n/k \rfloor} R_f \le \frac{4k}{n\kappa} R_f \tag{4.16}$$

for any $1 \leq k \leq n$. With (4.16) in hand, the proof of Corollary 4.1.2 is straightforward. Moreover, in our cases of interest where $R_f = o(n)$, the bound (4.16) quantifies a form of *approximate independence*: Most k-particle marginals of P are $\mathcal{W}_2$-close to product measures, if $k = o(n/R_f)$.

Remark 4.1.5. We work throughout the paper with state space $\mathbb{R}$, for simplicity. That is, we study approximations of measures on $\mathbb{R}^n$ by n-fold products of measures on $\mathbb{R}$, as opposed to, say, approximations of measures on $(\mathbb{R}^d)^n$ by n-fold products of measures on $\mathbb{R}^d$. Most of our arguments, based primarily on convexity and functional inequalities, extend to the case of $\mathbb{R}^d$ or even Riemannian manifolds with lower curvature bounds in the spirit of Bakry-Émery [130, 44]. The only difficulty is in the uniqueness claimed in Theorem 4.1.1 (proven in Proposition 4.3.9), which would require a finer analysis involving regularity of certain optimal transport maps.

4.1.3 Additional discussion and results

The remaining results presented in this section will not be used in the rest of the paper but serve to elaborate on the structure of the main theorem. The reader mainly interested in applications or proofs of the above results may skip to Chapters 4.2 or 4.3, respectively, with no loss of continuity.

More on entropic projections

Reversing the order of arguments in the relative entropy in (4.13) leads to a very different optimization problem, but it is instructive to compare the two. The infimum

$$\inf_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} H(P \,|\, Q) \tag{4.17}$$

is uniquely attained by taking $Q = P^* := P_1 \times \cdots \times P_n$ to be the product of the marginals of P. Indeed, from the simple identity $H(P \,|\, Q) = H(P \,|\, P^*) + H(P^* \,|\, Q)$, it follows that $H(P \,|\, Q) \geq H(P \,|\, P^*)$ for all Q, with equality if any only if $Q = P^*$.

The Gaussian case highlights the difference between (4.17) and (4.13). Suppose P is a centered Gaussian with nonsingular covariance matrix Σ. In this case it is easy to see that the (unique) minimizer of $H(Q \mid P)$ among product measures Q is the centered Gaussian with covariance matrix $\widetilde{\Sigma}$, where $\widetilde{\Sigma}^{-1}$ is the diagonal matrix obtained by deleting the off-diagonal entries of Σ^{-1}. On the other hand, the unique minimizer of $H(P \mid Q)$ among product measures Q is the centered Gaussian with covariance matrix $\widehat{\Sigma}$ obtained by deleting the off-diagonal entries of Σ.

Tilts

A similar bound to Corollary 4.1.4 is available if one seeks a stronger mean field approximation, in which $\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ is replaced by the sub-class of product measures given by *tilts* of a given reference measure. We focus on the case of Gaussian reference measure, as it is not obvious how to extend the argument to a general reference measure. For $y \in \mathbb{R}^n$, let $\gamma_{y,t}$ denote the Gaussian with mean y and covariance matrix tI, with $\gamma_t := \gamma_{0,t}$, noting that $\gamma_{y,t} \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$.

Proposition 4.1.6. *Let $t > 0$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be C^2 and concave. Assume there exist $c_1 \geq 0$ and $0 \leq c_2 < 1/2t$ such that $|f(x)| \leq c_1 e^{c_2|x|^2}$. Then there is a unique $y^* \in \mathbb{R}^n$ satisfying*

$$y^* = t \int_{\mathbb{R}^n} \nabla f \, \mathrm{d}\gamma_{y^*,t}, \tag{4.18}$$

and it holds that

$$\log \int_{\mathbb{R}^n} e^f \, \mathrm{d}\gamma_t \leq \sup_{y \in \mathbb{R}^n} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}\gamma_{y,t} - H(\gamma_{y,t} \mid \gamma_t) \right) + \frac{t^2}{2} \sum_{i,j=1}^n \int_{\mathbb{R}^n} |\partial_{ij} f|^2 \, \mathrm{d}\gamma_{y^*,t}. \tag{4.19}$$

Noting that $H(\gamma_{y,t} \mid \gamma_t) = |y|^2/2t$, a simple calculation shows that y^* uniquely attains the supremum in (4.19). The difference between (4.19) and (4.10) is that the former includes the diagonal terms $i = j$ in the sum. This is natural; an additively separable function $f(x) = \sum_{i=1}^n f_i(x_i)$ yields a product measure $P(\mathrm{d}x) = Z^{-1} e^{f(x)} \gamma_t(\mathrm{d}x)$, but it takes an *affine* function f for P to be a Gaussian. Small off-diagonal derivatives $\partial_{ij} f$ can be naturally interpreted as meaning f is close to being additively separable, but the full Hessian matrix $\nabla^2 f$ must to be small in order for f to be close to affine.

The above proposition is worth comparing with prior results based on gradient complexity. It was shown in [34, Proposition 3.4, arXiv version] that if $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 then

$$\log \int_{\mathbb{R}^n} e^f \, \mathrm{d}\gamma_t \le \sup_{y \in \mathbb{R}^n} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}\gamma_{y,t} - H(\gamma_{y,t} \,|\, \gamma_t) \right) + \sqrt{2} \int_{\mathbb{R}^n} \sup_{y \in \mathbb{R}^n} \big(x \cdot \nabla f(y) \big) \, \gamma_t(\mathrm{d}x). \quad (4.20)$$

The last integral is ($\sqrt{t}$ times) the *Gaussian mean-width* of the set $\nabla f(\mathbb{R}^n)$. This estimate (4.20) has the advantage of applying to non-concave functions f, but it is only meaningful if ∇f is bounded. Proposition 4.1.6, on the other hand, can accommodate non-Lipschitz but concave functions f.

Generalization of the main theorem

We briefly discuss how Theorem 4.1.1 can generalize beyond the strongly log-concave setting. Essentially, strong log-concavity is needed only for the uniqueness claims and to justify the log-Sobolev and Poincaré inequalities as explained in Chapter 4.1.2. *Uniqueness* of Q^* is actually not essential if one is interested only in a bound like (4.5). The *existence* of an optimizer Q^* is automatic, and it is not hard to show that it must satisfy the fixed point equation (4.3), modulo technical conditions. If it can be shown that Q^* admits a strictly positive C^2 density, and that P and Q^* obey a log-Sobolev and Poincaré inequality, respectively, with constants C_1 and C_2, then the following bound can be proven as in Chapter 4.1.2:

$$0 \le R_f \le C_1 \mathbb{E}_{Q^*} \sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i) \le 2C_1 C_2 \sum_{1 \le i < j \le n} \mathbb{E}_{Q^*}[|\partial_{ij} f(X)|^2].$$

It is unclear if our assumed bound on $|f(x)|$ is needed or merely an artifact of our proof technique. We use the assumed bound on $|f(x)|$ in the proof of Theorem 1.1 only to show that Q^* is strictly positive a.e., but this can be shown directly in many particular cases, such as when f is symmetric.

4.1.4 Outline of the chapter

In Chapter 4.2, we will present in detail the three main applications of Theorem 4.1.1, which pertain to Gibbs measures, high-dimensional Bayesian linear regression, and high-dimensional stochastic optimal control. The proof of Theorem 4.1.1 is given in Chapter 4.3.1, followed by the proof of Corollary 4.1.4 in Chapter 4.3.2. Chapter 4.3.3 contains the proof of Proposition 4.1.6, while Chapter 4.3.4 contains the proofs of the subadditivity inequality (4.16) and Corollary 4.1.2. Finally, the proofs of the applications are given in Chapters 4.4 and 4.5.

4.2 Applications

4.2.1 Gibbs measures with pairwise interactions

First, we study Gibbs measures with pairwise interaction potentials of the form (4.6), where the following assumption holds:

Assumptions 4.2.1. $V : \mathbb{R} \to \mathbb{R}$ is C^2 and κ-concave for some $\kappa > 0$, $K : \mathbb{R} \to \mathbb{R}$ is even, C^2, and concave, and J is a symmetric matrix with nonnegative entries and $J_{ii} = 0$ for all $i = 1, \ldots, n$. Assume there exists $a, b, c \geq 0$ and $0 \leq d < \kappa/2$ such that $|V(x)| \leq ce^{dx^2}$ and $|K''(x)|^2 \leq ae^{b|x|}$ hold for all $x \in \mathbb{R}$.

Note since K is even that there is no loss of generality in assuming that J is zero on the diagonal. The most traditional *mean field* setting is when $J_{ij} = 1/n$ for all (i, j), so that all particles interact equally, and there is a vast literature on the large-n behavior; see [132, 133] for some recent results and references. In general, the matrix J represents *disorder* or *heterogeneous interactions*, and a common situation is when J is the rescaled adjacency matrix of a graph. A notable strength of the non-asymptotic perspective of our work, and the theory of nonlinear large deviations more broadly, is that it can seamlessly handle this kind of heterogeneity. Gibbs measures with pairwise interactions on large graphs have been studied in many contexts, primarily on finite state space (see [29, 134, 135, 136] and references therein). In the continuous context we study here, these Gibbs measures appear as invariant measures of locally interacting diffusion processes whose large-scale behavior has recently been the subject of active research [137, 138].

To work toward applying Theorem 4.1.1 with f as in (4.6), we first record the simple observation that f is strongly concave under Assumption 4.2.1. The proof of this and other results in Chapter 4.2.1 are given in Chapter 4.4.

Lemma 4.2.2. *Define f by* (4.6)*, and suppose Assumption 4.2.1 holds. Then f is κ-concave.*

The following corollary will allow us to cover the case of unbounded K'', but only if we can control the barycenter of Q^* in the sense that $\mathbb{E}_{Q^*}[X_i - X_j] = 0$. This symmetry condition is justified in different ways in the following applications and is explained further in the part titled “On the symmetry of Q^*" in Chapter 4.2.1.

Corollary 4.2.3. *Define f by* (4.6)*, and suppose Assumption 4.2.1 holds. With Q^* denoting the unique optimizer of* (4.4)*, assume further that $\mathbb{E}_{Q^*}[X_i - X_j] = 0$. Then*

$$R_f \le \mathrm{Tr}(J^2) a \kappa^{-2} e^{b^2/\kappa}.$$

Remark 4.2.4. Corollary 4.2.3 shows that $R_f = o(n)$ as long as $\mathrm{Tr}(J^2) = o(n)$. The assumption $\mathrm{Tr}(J^2) = o(n)$ has been used in the literature as a *mean field condition* for quadratic interaction models, first in [29, Theorem 1.1] and then in [127, Theorem 4]. Both cases are limited to measures with compact support. Moreover, in their setting, neither uniqueness of the optimizer nor convergence of the empirical measure hold in general. In contrast, in our setting we can allow measures of unbounded support, and we show both uniqueness of the optimizer and the convergence of the empirical measure in Theorems 4.2.5 and 4.2.8 below. On the other hand, our results require concavity assumptions which were not needed in [29, 127].

Using Corollary 4.2.3, one can study the weak law of large numbers of the empirical measure under P, by studying the corresponding weak law under the product measure Q^*. Under additional assumptions on the matrix J, the mean field optimization problem can be shown to converge as

$n \to \infty$, allowing us to characterize the weak law under P in terms of the limiting optimization problem. Below we illustrate this in two special cases.

Doubly stochastic matrices

In the following $n \to \infty$ results, note that the dependence of f, $P(\mathrm{d}x) = Z^{-1}e^{f(x)}\mathrm{d}x$ and J on n is suppressed.

Theorem 4.2.5. *Define f by* (4.6)*, and suppose Assumption 4.2.1 holds. Assume there exist $a, b \geq 0$ such that $|K''(x)|^2 \leq ae^{b|x|}$ for all x. Assume further that the symmetric matrix J is doubly stochastic (i.e., $\sum_{j=1}^n J_{ij} = 1$ for all i), and obeys the mean field condition* $\mathrm{Tr}(J^2) = o(n)$*. Then we have the following conclusions:*

(1)

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^n} e^{f(x)}\,\mathrm{d}x = \sup_{Q\in\mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}} V\,\mathrm{d}Q + \frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} K(x-y)Q(\mathrm{d}x)Q(\mathrm{d}y) - H(Q) \right). \tag{4.21}$$

(2) The supremum in (4.21) *is attained by a unique $Q \in \mathcal{P}(\mathbb{R})$, and if $(X_1, \ldots, X_n) \sim P$ then*

$$\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \to Q, \text{ weakly in law.} \tag{4.22}$$

The above theorem applies when $J = A/d$ and A is the adjacency matrix of a d-regular graph. In this case we get $\mathrm{Tr}(J^2) = n/d$, which is $o(n)$ as long as $d \to \infty$. The above theorem is similar in spirit to [29, Theorem 2.1], which dealt with Ising and Potts models, and a comment similar to Remark 4.2.4 applies. Note that one cannot expect a mean field approximation to be valid in the sparsest (diluted) case, where d stays bounded as $n \to \infty$. The framework of local weak convergence has proven to be successful in this context [139], and we refer also to [140, Sections 2 and B] for continuous models encompassing the form studied here, and for a detailed derivation of the (folklore) limit of the empirical measure for locally convergent graph sequences, which requires uniqueness of the infinite-volume Gibbs measure on the limiting graph.

Graphons

Another case in which we can derive asymptotics of the log partition function is when the matrix J converges to a graphon W in cut metric. Below we introduce the relevant notions, deferring to [141, 142, 143, 144] for additional background:

Definition 4.2.6. Let $\mathcal{W}$ denote the space of all symmetric measurable functions from $[0,1]^2$ to $[0,\infty)$ which are integrable. For $W_1, W_2 \in \mathcal{W}$, define the strong cut (pseudo-)metric by

$$d_\square(W_1, W_2) := \sup_{S,T\subset[0,1]} \left| \int_{S\times T} \big(W_1(u,v) - W_2(u,v)\big)\mathrm{d}u\mathrm{d}v \right|,$$

and their weak cut (pseudo-)metric by

$$\delta_\square(W_1, W_2) := \inf_{\varphi} d_\square(W_1, W_2^\varphi),$$

where the infimum is over all invertible measure-preserving maps $\varphi : [0,1] \to [0,1]$, and $W_2^\varphi(u,v) := W_2(\varphi(u), \varphi(v))$. Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$ with nonnegative entries, we define a function $W_A \in \mathcal{W}$ by setting $W_A(u,v) := A_{\lceil nu\rceil, \lceil nv\rceil}$. We say that a sequence of symmetric matrices $\{A_n\}$ *converges in weak cut metric* to a function $W \in \mathcal{W}$ if $\delta_\square(W_{A_n}, W) \to 0$.

Remark 4.2.7. Suppose G_n is the adjacency matrix of an Erdős-Rényi random graph on n vertices with parameter p_n, such that $np_n \to \infty$. If $J_n = \frac{1}{np_n} G_n$, then nJ_n converges in strong cut metric to the constant function 1 (see [142, Example 3.3.1]). Similar convergences hold if G_n arises from a stochastic block model, where the edge probability matrix has a block structure, in which case the limiting W retains the same block structure. For more examples of convergent sequence of graphs in cut metric, we refer again to [141, 142, 143, 144] and references therein.

Let $\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ denote the space of all probability measures on $[0,1]\times\mathbb{R}$ with uniform first marginal. Note that any $\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ admits the disintegration $\mu(\mathrm{d}u,\mathrm{d}x) = \mathrm{d}u\mu_u(\mathrm{d}x)$.

Theorem 4.2.8. *Define f by* (4.6)*, and suppose Assumption 4.2.1 holds. Assume there exist $a, b \geq 0$ such that $|K''(x)|^2 \leq ae^{b|x|}$ for all x. Assume further that V is even, K is nonpositive, $\int_{\mathbb{R}} e^{V(x)}\,\mathrm{d}x = 1$, and $J = \{J_n\}$ is a sequence of matrices such that $\{nJ_n\}$ converges in weak cut metric to a function $W \in \mathcal{W}$. Assume also that $\mathrm{Tr}(J_n^2) = o(n)$.*

(1) Defining the probability measure $\rho(\mathrm{d}x) = e^{V(x)}\,\mathrm{d}x$, we have

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{R}^n} e^{f(x)}\,\mathrm{d}x \\ &\quad = \sup_{\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})}\left(\frac{1}{2}\int_{([0,1]\times\mathbb{R})^2} W(u,v)K(x-y)\mu(\mathrm{d}u,\mathrm{d}x)\mu(\mathrm{d}v,\mathrm{d}y) - \int_0^1 H(\mu_u\,|\,\rho)\,\mathrm{d}u\right).\end{aligned} \tag{4.23}$$

(2) The supremum in (4.23) *is attained by a unique $\mu^* \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$, and if $(X_1,\ldots,X_n) \sim P$, then*

$$\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \to \int_0^1 \mu_u^*\,\mathrm{d}u, \ \textit{weakly in law.} \tag{4.24}$$

Remark 4.2.9. It follows from [141, Propositions C.5 and C.15] that the condition $\mathrm{Tr}(J_n^2) = o(n)$ holds automatically if J_n is the adjacency matrix of a simple graph $G_n = ([n], E_n)$ multiplied by $n/(2|E_n|)$, and nJ_n converges in cut metric. However, if J_n is a general matrix, we need the added assumption $\mathrm{Tr}(J_n^2) = o(n)$ in Theorem 4.2.8.

On the symmetry of Q^*

This short section elaborates on conditions under which one can check that $\mathbb{E}_{Q^*}[X_i - X_j] = 0$, which was needed in Corollary 4.2.3. The main two conditions we found are evenness and a weak form of permutation invariance.

Definition 4.2.10. Let S be a set of permutations of $[n]$. We say that S is *transitive* if for every $i, j \in [n]$ there exists $\pi \in S$ such that $\pi(i) = j$. We say also that a function f on $\mathbb{R}^n$ is *invariant under S* if $f(x_1,\ldots,x_n) = f(x_{\pi(1)},\ldots,x_{\pi(n)})$ for every $x \in \mathbb{R}^n$ and $\pi \in S$.

Lemma 4.2.11. *In the setting of Theorem 4.1.1, the following implications hold:*

(1) If f is even, meaning $f(-x) = f(x)$ for all x, then Q_i^* is even for each $i = 1, \ldots, n$.

(2) Suppose f is invariant under a transitive set of permutations. Then $Q_1^* = Q_2^* = \cdots = Q_n^*$.

In both cases, we have $\mathbb{E}_{Q^*}[X_i - X_j] = 0$ for all $i, j \in [n]$.

When f is of the form (4.6), it is clear that f is even if K and V are, and indeed V is assumed even in Theorem 4.2.5 to enable an application of Lemma 4.2.11(1). We will not apply Lemma 4.2.11(2), but we find it interesting in its own right. For instance, (2) holds if f is symmetric, i.e., invariant under all permutations. Another natural case covered by (2) is where f is of the form (4.6) and J is a scalar multiple of the adjacency matrix of a *vertex transitive* graph.

4.2.2 High dimensional Bayesian linear regression

Our next application is concerned with high dimensional Bayesian linear regression. Suppose we observe a set of data $\{(y_i, X_i)\}_{i=1}^n$, where $y_i \in \mathbb{R}$ and $X_i \in \mathbb{R}^p$. Let $y = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$ and $\mathbf{X}^\top = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$. Consider the linear regression model

$$y = \mathbf{X}\beta + \varepsilon, \quad \varepsilon \sim \gamma_{\sigma^2},$$

where γ_{σ^2} denotes the Gaussian with mean 0 and covariance matrix $\sigma^2 I$. Here $\beta \in \mathbb{R}^p$ is the unknown parameter.

Following a Bayesian approach, assume that $\beta = (\beta_1, \ldots, \beta_p)^\top \overset{i.i.d.}{\sim} \pi$, where π is a prior distribution on $\mathbb{R}$ with density proportional to $e^V \in L^1(\mathbb{R})$ for some $V : \mathbb{R} \to \mathbb{R}$. The posterior density $\pi_{y,\mathbf{X}}$ of β given y and $\mathbf{X}$ is then proportional to $e^{f_{y,\mathbf{X}}}$, where

$$f_{y,\mathbf{X}}(\beta) := \sum_{i=1}^p V(\beta_i) - \frac{1}{2\sigma^2} |y - \mathbf{X}\beta|^2 .$$

The posterior distribution is the central object of inference in Bayesian statistics. Note that even though β has independent coordinates under the prior, the coordinates of β are no longer independent under the posterior. Frequently, mean-field techniques are used to approximate such complex posterior distributions, including and beyond the set up of Bayesian linear regression (see [145, 146, 147, 148, 149] and references therein). In particular, it is useful to understand what conditions guarantee the validity of a mean field approximation, showing that the posterior is close to a product measure. Using Theorem 4.1.1, the following corollary provides sufficient conditions under which the posterior is indeed mean-field. Leveraging this, it also derives a law of large numbers for the empirical measure under the true posterior distribution.

Corollary 4.2.12. *Assume V is κ_1-concave for some $\kappa_1 \in \mathbb{R}$, and that there exists $c_1 \geq 0$ and $0 \leq c_2 < \kappa/2$ such that $|V(x)| \leq c_1 e^{c_2 x^2}$ for all $x \in \mathbb{R}$. Set $J = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{p \times p}$, and assume that $J \geq \kappa_2 I$ for some $\kappa_2 \in \mathbb{R}$ such that $\kappa_1 + \kappa_2 \sigma^{-2} > 0$. Then*

$$\sup_{y \in \mathbb{R}^n} \left| \log \int_{\mathbb{R}^p} e^{f_{y,\mathbf{X}}(\beta)} \mathrm{d}\beta - \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^p)} \left(\int_{\mathbb{R}^p} f_{y,\mathbf{X}} \, \mathrm{d}Q - H(Q) \right) \right| \leq \frac{1}{(\kappa_1 \sigma^2 + \kappa_2)^2} \sum_{1 \leq i < j \leq p} J_{ij}^2. \tag{4.25}$$

Moreover, for every $y \in \mathbb{R}^n$, the inner supremum in (4.25) *is attained by a unique $Q_y^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^p)$, and for any 1-Lipschitz function $\varphi : \mathbb{R} \to \mathbb{R}$, we have*

$$\sup_{y \in \mathbb{R}^n} \mathbb{E}_{\pi_{y,\mathbf{X}}} \left[\left(\frac{1}{p} \sum_{i=1}^p \varphi(\beta_i) - \frac{1}{p} \sum_{i=1}^p \mathbb{E}_{Q_y^*} \left[\varphi(\beta_i) \right] \right)^2 \right] \leq \frac{\sigma^2 \left(\kappa_1 \sigma^2 + \kappa_2 + \sqrt{2 \sum_{1 \leq i < j \leq p} J_{ij}^2} \right)^2}{p (\kappa_1 \sigma^2 + \kappa_2)^3}. \tag{4.26}$$

The proof of this corollary is by a direct application of Theorem 4.1.1 and Corollary 4.1.2, and is hence omitted. Indeed, the concavity assumption on V and the lower bound on J ensure that $\nabla^2 f_{y,\mathbf{X}}(\beta) \leq -(\kappa_1 + \kappa_2 \sigma^{-2}) I$ for all β.

Remark 4.2.13. The uniformity in y in (4.25) implies that the mean field approximation continues to hold with high probability, under any distributional assumption on y. Note that when $n, p \to \infty$ in any arbitrary manner, the right-hand side of (4.25) and (4.26) are $o(p)$ and $o(1)$ respectively, as long as $\sum_{1 \leq i < j \leq p} J_{ij}^2 = o(p)$ when $n, p \to \infty$. We also point out that the same conclusion as in

(4.25) above was derived in [150, Theorem 1] using very different techniques, under the assumption that the prior distribution π is compactly supported. In our setup, we allow the support to be non-compact, but instead assume that the prior distribution is strongly log-concave. One added advantage of our setup is that we also get the law of large numbers under no extra assumptions.

4.2.3 Stochastic control

This section describes an application of Corollary 4.1.4 to a class of high-dimensional stochastic optimal control problems. Let $T > 0$, and let $g : \mathbb{R}^n \to \mathbb{R}$ be C^2 and concave. Consider the stochastic control problem

$$V_{\text{orig}} := \sup \mathbb{E}\left[g(X_T) - \frac{1}{2n} \sum_{i=1}^{n} \int_0^T |\alpha_i(t, X_t)|^2 \, \mathrm{d}t \right], \tag{4.27}$$

where the supremum is over pairs (α, X), where $\alpha = (\alpha_1, \ldots, \alpha_n) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ is a measurable function and $X = (X^1, \ldots, X^n)$ a weak solution of the stochastic differential equation (SDE)

$$\mathrm{d}X_t^i = \alpha_i(t, X_t) \, \mathrm{d}t + \mathrm{d}B_t^i, \qquad X_0^i = 0, \;\; i = 1, \ldots, n, \tag{4.28}$$

defined on an arbitrary filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, satisfying also $\int_0^T |\alpha(t, X_t)|^2 \, \mathrm{d}t < \infty$ a.s. Here $B = (B^1, \ldots, B^n)$ is an n-dimensional $\mathbb{F}$-Brownian motion, and X is required to be $\mathbb{F}$-adapted. We call such a pair (α, X) *admissible*. There is a well known semi-explicit solution to (4.27) which has come to be known as the *Föllmer drift*, which we will discuss in Remark 4.2.15 below.

We interpret $i = 1, \ldots, n$ as the indices of different "players," each facing an independent source of randomness B^i, and each choosing a control α_i which can depend on the full information of all n players. Players "cooperate" in the sense that $(\alpha_1, \ldots, \alpha_n)$ are chosen together to optimize

(4.27). When g is of the form

$$g(x) = G\bigg(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}\bigg), \quad \text{for some } G : \mathcal{P}(\mathbb{R}) \to \mathbb{R}, \tag{4.29}$$

we recover a well-studied class of problems which goes under the name *mean field control* in the cooperative setting [151], or *mean field games* in the competitive (Nash equilibrium) setting [152, 153]; see [154] for an overview. In this setting, it is typically argued that V_{orig} converges to the value of a limiting “mean field" control problem, and the optimal control $\widehat{\alpha}$ from this limiting problem can be used to construct *distributed* controls $\alpha_i(t, x_1, \ldots, x_n) = \widehat{\alpha}(t, x_i)$ which are provably approximately optimal for the n-player problem for n large. This is a very desirable outcome, because distributed controls are much simpler (lower-dimensional).

Our results give a new non-asymptotic perspective on control problems of this form, by showing how to construct approximately optimal distributed controls for much more general g than in (4.29). The link between (4.27) and the setting of Chapter 4.1 is the formula

$$V_{\mathrm{orig}} = \sup_{Q\in\mathcal{P}(\mathbb{R}^n)} \bigg(\int_{\mathbb{R}^n} g\,\mathrm{d}Q - \frac{1}{n}H(Q\,|\,\gamma_T)\bigg) = \frac{1}{n}\log\int_{\mathbb{R}^n} e^{ng}\,\mathrm{d}\gamma_T, \tag{4.30}$$

where we recall that γ_T denotes the centered Gaussian with covariance matrix TI. This formula is essentially a well known consequence of Girsanov's theorem.[1] The mean field approximation also admits a natural control-theoretic interpretation. Define

$$V_{\mathrm{dstr}} := \sup\,\mathbb{E}\Bigg[g(X_T) - \frac{1}{2n}\sum_{i=1}^{n}\int_0^T |\alpha_i(t, X_t)|^2\,\mathrm{d}t\Bigg], \tag{4.31}$$

where the supremum is now over admissible pairs (α, X) for which $\alpha = (\alpha_1, \ldots, \alpha_n)$ is of the

[1] Experts might recognize a similarity with a famous formula often named after Boué-Dupuis [155] or Borell [156], though the form we present here is simpler because of our restriction to Markovian controls, whereas [155, 156] work with *open-loop* controls, i.e., controls specified as arbitrary progressively measurable processes.

form

$$\alpha_i(t, x_1, \dots, x_n) = \widehat{\alpha}_i(t, x_i),$$

for some measurable $\widehat{\alpha}_i : [0, T] \times \mathbb{R} \to \mathbb{R}$, and also for which $X_t^1, \dots, X_t^n$ are independent for each $t \in [0, T]$ (this second statement being redundant if the SDE (4.28) driven by this α is known to be unique in law). Let us call any such pair (α, X) a *distributed admissible pair*. We will derive the following result from Corollary 4.1.4, after first showing that V_{dstr} is nothing but the mean field approximation of (4.30), in the sense that

$$V_{\mathrm{dstr}} = \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g \, \mathrm{d}Q - \frac{1}{n} H(Q \,|\, \gamma_T) \right). \tag{4.32}$$

Corollary 4.2.14. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be C^2 and concave, and let $T > 0$. Assume there exists $c_1 \geq 0$ and $0 \leq c_2 < 1/2T$ such that $|g(x)| \leq c_1 e^{c_2 |x|^2}$ for all $x \in \mathbb{R}^n$. Define V_{orig} and V_{dstr} by* (4.27) *and* (4.31)*, respectively. Then the formulas* (4.30) *and* (4.32) *hold, and*

$$0 \leq V_{\mathrm{orig}} - V_{\mathrm{dstr}} \leq nT^2 \sum_{1 \leq i < j \leq n} \mathbb{E}_{Q^*}[|\partial_{ij} g(X)|^2], \tag{4.33}$$

where $Q^* = Q_1^* \times \cdots \times Q_n^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ is the unique product measure with strictly positive density a.e. satisfying $g \in L^1(Q^*)$ and the fixed point equation

$$Q_i^*(\mathrm{d}x_i) = Z_i^{-1} \exp\left(n \mathbb{E}_{Q^*}[g(X) \,|\, X_i = x_i] \right) \gamma_T(\mathrm{d}x_i), \qquad Z_i > 0, \; i = 1, \dots, n.$$

The proof is given in Chapter 4.5. Corollary 4.2.14 shows that distributed controls are approximately optimal for large n if $n \| \sum_{i \neq j} \partial_{ij} g \|_\infty^2 = o(1)$. As an example, if g is of the form (4.29) and G is twice continuously *Wasserstein-* or *L-differentiable* in the sense of [154, Chapter 5.2], then

$$\partial_i g(x) = \frac{1}{n} D_m G\left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, x_i \right), \quad \partial_{ij} g(x) = \frac{1}{n^2} D_m^2 G\left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, x_i, x_j \right), \;\; i \neq j.$$

Hence, if $D_m^2 G$ is bounded, then the right-hand side of (4.33) is bounded by $T^2 \|D_m^2 G\|_\infty^2 / 2n$.

Remark 4.2.15. In fact, the proof of Corollary 4.2.14 also yields an explicit characterization of the optimal distributed control in (4.31), which we summarize as follows. For a measure $Q \ll \gamma_T$, consider a process $X = (X_t)_{t \in [0,T]}$ such that $X_T \sim Q$ and the conditional law of the trajectory $(X_t)_{t \in [0,T]}$ given $X_T = x$ coincides with the law of the Brownian bridge from 0 to x on the time interval $[0, T]$. This process might be called the *Brownian* (or *Schrödinger*) *bridge with terminal law* Q. The associated control α is given by $\alpha(t, x) = \nabla_x \log \mathbb{E}[\frac{\mathrm{d}Q}{\mathrm{d}\gamma_T}(x + B_T - B_t)]$, as shown in full generality by Föllmer [157, 158]. Note that the associated SDE (4.28) may not be pathwise unique in general, but it always admits a weak solution X with the law just described. The optimizer for the original control problem (4.27) is nothing but the Brownian bridge with terminal law $P(\mathrm{d}x) = Z^{-1} e^{ng(x)} \gamma_T(\mathrm{d}x)$. Similarly, the optimizer for the distributed control problem (4.31) is the Brownian bridge with terminal law Q^*.

Remark 4.2.16. Proposition 4.1.6 admits a similar control-theoretic formulation in terms of *deterministic* controls. Let V_{det} denote the value of the stochastic control problem (4.27) but with the supremum limited to those admissible pairs (α, X) in which the control is non-random, i.e., $\alpha_i(t, x) = \tilde{\alpha}_i(t)$ for some $\tilde{\alpha}_i \in L^2[0, T]$. For these controls, X_t is Gaussian with covariance matrix tI for each $t \in [0, T]$. It can then be shown that

$$V_{\mathrm{det}} = \sup_{y \in \mathbb{R}^n} \left(\int_{\mathbb{R}^n} g \, \mathrm{d}\gamma_{y,T} - \frac{1}{n} H(\gamma_{y,T} \,|\, \gamma_T) \right) = \sup_{y \in \mathbb{R}^n} \left(\int_{\mathbb{R}^n} g \, \mathrm{d}\gamma_{y,T} - \frac{|y|^2}{2nT} \right),$$

and Proposition 4.1.6 yields the following analogue of (4.33):

$$0 \le V_{\mathrm{orig}} - V_{\mathrm{det}} \le \frac{nT^2}{2} \sum_{i,j=1}^{n} \int_{\mathbb{R}^n} |\partial_{ij} g|^2 \, \mathrm{d}\gamma_{y^*,T},$$

where $y^* \in \mathbb{R}^n$ is the unique solution of $y^* = T \int_{\mathbb{R}^n} \nabla g \, \mathrm{d}\gamma_{y^*,T}$.

4.3 Proof of the main theorem

The proofs will make use of the well known log-Sobolev and Poincaré inequalities for strongly log-concave measures, recalled here for convenience as we will use them in several parts of the paper. The former is due to Bakry-Émery (see [130] or [44, Corollary 5.7.2]), and the latter is a consequence of the Brascamp-Lieb inequality [159, Theorem 4.1].

Theorem 4.3.1 (Log-Sobolev inequality)**.** *If $h : \mathbb{R}^n \to \mathbb{R}$ is C^2 and κ-concave, and $R(\mathrm{d}x) = e^{h(x)}\mathrm{d}x$ is a probability measure, then R satisfies the log-Sobolev inequality,*

$$H(Q\,|\,R) \leq \frac{1}{2\kappa}\int_{\mathbb{R}^n} \left|\nabla \log \frac{\mathrm{d}Q}{\mathrm{d}R}\right|^2 \mathrm{d}Q,$$

for every $Q \in \mathcal{P}(\mathbb{R}^n)$ such that $Q \ll R$ and the weak gradient of $\log \mathrm{d}Q/\mathrm{d}R$ exists in $L^2(Q)$.

Theorem 4.3.2 (Poincaré inequality)**.** *If $h : \mathbb{R}^n \to \mathbb{R}$ is κ-concave, and $R(\mathrm{d}x) = e^{h(x)}\,\mathrm{d}x$ is a probability measure, then R satisfies the Poincaré inequality,*

$$\mathrm{Var}_R(\varphi) := \int_{\mathbb{R}^n} \varphi^2\,\mathrm{d}R - \left(\int_{\mathbb{R}^n} \varphi\,\mathrm{d}R\right)^2 \leq \frac{1}{\kappa}\int_{\mathbb{R}^n} |\nabla\varphi|^2\,\mathrm{d}R,$$

for every continuously differentiable function $\varphi : \mathbb{R}^n \to \mathbb{R}$ in $L^1(R)$.

The above Poincaré inequality is normally stated with the additional assumptions that h is C^2, which is easily removed by mollification by a Gaussian, and that $\varphi \in L^2(R)$, which can be weakened to $L^1(R)$ by monotone approximation, though both sides may be infinite.

We will also make use of the Gibbs variational principle, which is well known, but we give the proof as we need a non-standard form which is careful about edge cases. Recall our convention that $H(Q) := \infty$ if Q is not absolutely continuous or if $Q \log Q \notin L^1(\mathbb{R}^n)$.

Theorem 4.3.3 (Gibbs variational principle)**.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ be measurable, bounded from above and such that $Z := \int_{\mathbb{R}^n} e^f\,\mathrm{d}x \in (0, \infty)$. Define $P \in \mathcal{P}(\mathbb{R}^n)$ by $P(\mathrm{d}x) = Z^{-1}e^{f(x)}\,\mathrm{d}x$.*

Then

$$\sup_{Q\in\mathcal{P}(\mathbb{R}^n)}\left(\int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q)\right) = \log Z \in (-\infty,\infty), \tag{4.34}$$

and the following are equivalent:

(1) $H(P) < \infty$.

(2) The supremum in (4.34) *is attained uniquely by* P.

(3) There exists a maximizer in (4.34).

Proof. We first prove (4.34). Since f is bounded from above, $\int_{\mathbb{R}^n} f\,\mathrm{d}Q \in [-\infty,\infty)$ is well-defined for all $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$. We may thus restrict the supremum in (4.34) to those Q with $H(Q) < \infty$. For $H(Q) < \infty$, we have the simple identity

$$\int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q) = -H(Q\,|\,P) + \log Z. \tag{4.35}$$

Therefore,

$$\sup_{Q\in\mathcal{P}(\mathbb{R}^n)}\left(\int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q)\right) = -\inf\left\{H(Q\,|\,P) : Q\in\mathcal{P}(\mathbb{R}^n),\ H(Q)<\infty\right\} + \log Z,$$

and it suffices to show that the infimum on the right-hand side is zero. We proceed by approximation. For each $k \in \mathbb{N}$, let $B_k \subset \mathbb{R}^n$ denote the centered ball of radius k, and define the probability density $Q_k = P1_{B_k}/P(B_k)$. Since f is bounded from above, the density Q_k is bounded and supported on the bounded set B_k. Thus $Q_k \log Q_k \in L^1(\mathbb{R}^n)$, or $H(Q_k) < \infty$, and we conclude that $H(Q\,|\,P) \le \liminf_k H(Q_k\,|\,P)$. Finally, since $P(B_k) \to 1$,

$$H(Q_k\,|\,P) = -\log P(B_k) \to 0.$$

This proves the claim (4.34).

Turning to the equivalence of (1–3), the implication (1) $\Rightarrow$ (2) follows by taking $Q = P$ in (4.35). The implication (2) $\Rightarrow$ (3) is trivial. Lastly, for the implication (3) $\Rightarrow$ (1), suppose

$Q \in \mathcal{P}(\mathbb{R}^n)$ attains the supremum in (4.34). We know from (4.34) that the supremum is not $-\infty$, so $H(Q) < \infty$. Then, for any $R \in \mathcal{P}(\mathbb{R}^n)$ with $H(R) < \infty$, the identity (4.35) implies

$$-H(R\,|\,P) + \log Z = \int_{\mathbb{R}^n} f \,\mathrm{d}R - H(R) \leq \int_{\mathbb{R}^n} f \,\mathrm{d}Q - H(Q) = -H(Q\,|\,P) + \log Z.$$

Rearrange and minimize over R to get

$$H(Q\,|\,P) \leq \inf \Big\{ H(R\,|\,P) : R \in \mathcal{P}(\mathbb{R}^n),\ H(R) < \infty \Big\} = 0,$$

where the last equality was shown just above while proving (4.34). It follows that $H(Q\,|\,P) = 0$, so $Q = P$, and $H(P) = H(Q) < \infty$. This completes the proof.

$\square$

4.3.1 Proof of Theorem 4.1.1

This section proves Theorem 4.1.1 in several parts, and we assume throughout that f satisfies the assumptions therein. Since f is C^2 and κ-concave,

$$f(x) \leq a - b|x|^2, \qquad \text{for all } x \in \mathbb{R}^n, \ \text{ where } a := f(0) + \kappa^{-1}|\nabla f(0)|^2, \ b := \kappa/4. \tag{4.36}$$

This implies that $Z := \int_{\mathbb{R}^n} e^{f(x)}\mathrm{d}x < \infty$, so $P(dx) = Z^{-1}e^{f(x)}\mathrm{d}x$ is well defined. Moreover, f is bounded from above, so $\int_{\mathbb{R}^n} f\,\mathrm{d}Q$ is well defined in $[-\infty, \infty)$ for every $Q \in \mathcal{P}(\mathbb{R}^n)$. Note lastly that $fe^f \in L^1(\mathbb{R}^n)$, or equivalently $H(P) < \infty$, which follows from the growth assumption on $|f|$ and the fact that the κ-log-concave measure P satisfies $\int_{\mathbb{R}^n} e^{c|x|^2} P(\mathrm{d}x) < \infty$ for each $c < \kappa/2$. (In fact, every absolutely continuous log-concave measure has finite entropy [160, Theorem I.1].) We first establish some properties of the optimization and fixed point problems appearing in Theorem 4.1.1.

Lemma 4.3.4. *It holds that*

$$-\infty < \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}Q - H(Q) \right) < \infty, \tag{4.37}$$

and any $Q^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ attaining the supremum satisfies $f \in L^1(Q^*)$. Also, equation (4.12) *is valid.*

Proof. The Gibbs variational formula (Theorem 4.3.3) implies that the supremum in (4.37) is no greater than $\log Z < \infty$. To see that it is not $-\infty$, note that f is locally bounded because it is concave and real-valued. Hence, if Q is any product measure with bounded support and finite entropy (such as the uniform measure on $[0,1]^n$), we can bound the supremum from below by $\int_{\mathbb{R}^n} f \, \mathrm{d}Q - H(Q) > -\infty$. Now, if Q^* is an optimizer, then $H(Q^*) < \infty$ and $\int_{\mathbb{R}^n} f \mathrm{d}Q^* > -\infty$, the latter implying that $f \in L^1(Q^*)$ since f is bounded from above.

To prove (4.12), note that the simple calculation (4.11) is valid for any $Q \in \mathcal{P}(\mathbb{R}^n)$ with $H(Q) < \infty$, though both sides are $+\infty$ if and only if $\int_{\mathbb{R}^n} f \, \mathrm{d}Q = -\infty$. Since $\int_{\mathbb{R}^n} f \, \mathrm{d}Q$ always exists in $[-\infty, \infty)$, the supremum in (4.37) remains the same when restricted to those Q with $H(Q) < \infty$. By infimizing (4.11) over $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ with finite entropy, we deduce that the left-hand side of (4.12) is finite and equals $\inf\{H(Q \,|\, P) : Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n), \, H(Q) < \infty\}$. To complete the proof, we claim that if $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ satisfies $H(Q \,|\, P) < \infty$ and $H(Q) = \infty$, then there exists $Q_k \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ such that $H(Q_k) < \infty$ for each k and $H(Q_k \,|\, P) \to H(Q \,|\, P)$. Indeed, define the probability density $Q_k = Q 1_{B_k} / Q(B_k)$, where $B_k = [-k, k]^n$, for k large enough that $Q(B_k) > 0$. Then

$$H(Q_k \,|\, P) = \frac{1}{Q(B_k)} \int_{B_k} \log \frac{\mathrm{d}Q}{\mathrm{d}P} \, \mathrm{d}Q - \log Q(B_k)$$

is finite and converges to $H(Q \,|\, P)$ as $k \to \infty$. In particular, $\log(\mathrm{d}Q_k/\mathrm{d}P) \in L^1(Q_k)$. We also have $\log P = f - \log Z \in L^1(Q_k)$ because f is locally bounded and Q_k has compact support. We deduce that $\log Q_k \in L^1(Q_k)$, or $H(Q_k) < \infty$, which completes the proof.

$\square$

The following proposition shows essentially that the fixed point problem (4.3) is the first order condition for optimality in (4.4). This extends naturally to much more general settings, with $(\mathbb{R}^n, dx)$ replaced by a general σ-finite product measure space, but we will not need this.

Proposition 4.3.5 (Optimality to fixed point)**.** *Suppose $Q^* = Q_1^* \times \cdots \times Q_n^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ attains the supremum in* (4.37)*. Then $f \in L^1(Q^*)$ and Q^* satisfies the fixed point equation*

$$\begin{aligned} Q_i(dx_i) &= Z_i^{-1} e^{\hat{f}_i(x_i)}\,\mathrm{d}x_i, \qquad \textit{where } \hat{f}_i : \mathbb{R} \to \mathbb{R} \cup \{-\infty\} \textit{ is defined by} \\ \hat{f}_i(x_i) &:= \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_n) \prod_{j \neq i} Q_j^*(\mathrm{d}x_j), \qquad i \in [n]. \end{aligned} \tag{4.38}$$

Proof. Note that $f \in L^1(Q^*)$ by Lemma 4.3.4. By assumption, $(Q_1^*, \ldots, Q_n^*)$ attains the supremum

$$\sup_{Q_1,\ldots,Q_n \in \mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}^n} f\,\mathrm{d}(Q_1 \times \cdots \times Q_n) - H(Q_1 \times \cdots \times Q_n) \right).$$

Clearly, $\hat{f}_i(x_i) = \mathbb{E}_{Q^*}[f(X)\,|\,X_i = x_i]$ for Q_i^*-a.e. $x_i \in \mathbb{R}$. Also, it is well known that entropy tensorizes for product measures: $H(Q_1 \times \cdots \times Q_n) = \sum_{i=1}^n H(Q_i)$. From these and the tower property it follows for each $i \in [n]$ that Q_i^* attains the supremum

$$S_i := \sup_{Q_i \in \mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}} \hat{f}_i\,\mathrm{d}Q_i - H(Q_i) \right). \tag{4.39}$$

We wish to invoke the Gibbs variational principle (Theorem 4.3.3) to deduce that this supremum is uniquely attained by the probability measure with density proportional to $e^{\hat{f}_i}$, and thus $Q_i^*(\mathrm{d}x_i) = Z_i^{-1} e^{\hat{f}_i(x_i)}\,\mathrm{d}x_i$, which yields (4.38). It remains to carefully check the conditions of Theorem 4.3.3. We know that Q_i^* attains the supremum (4.39), so we must just check that $Z_i \in (0, \infty)$. Note that (4.36) implies $f(x) \leq a - bx_i^2$ for all $x \in \mathbb{R}^n$, and thus $\hat{f}_i(x_i) \leq a - bx_i^2$ for all $x_i \in \mathbb{R}$, which implies $Z_i = \int_{\mathbb{R}} e^{\hat{f}_i(x_i)}\,\mathrm{d}x_i < \infty$. Next, recall from Lemma 4.3.4 that $f \in L^1(Q^*)$, so by Fubini's theorem, $Q_i^*(|\hat{f}_i| < \infty) = 1$. Note that Q_i^* is absolutely continuous since $H(Q_i^*) < \infty$. Hence, $\{|\hat{f}_i| < \infty\}$ has nonzero Lebesgue measure, and so $Z_i > 0$.

□

Lemma 4.3.6. *There exists a unique maximizer in* (4.37).

Proof. We first prove existence. Recalling the identity (4.12), the optimizers of (4.37) are in one-to-one correspondence with the optimizers of $\inf_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} H(Q \,|\, P)$. The latter exist because $\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ is a weakly closed subset of $\mathcal{P}(\mathbb{R}^n)$ and because $H(\cdot \,|\, P)$ has weakly compact sub-level sets.

We next prove uniqueness. Let $Q^* = Q_1^* \times \cdots \times Q_n^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ denote any optimizer of (4.5). Define $G_1, G_2 : (\mathcal{P}(\mathbb{R}))^n \to \mathbb{R}$ by

$$G_1(Q_1, \ldots, Q_n) \coloneqq \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) \prod_{i=1}^n Q_i(\mathrm{d}x_i), \quad G_2(Q_1, \ldots, Q_n) \coloneqq H(Q_1 \times \cdots \times Q_n).$$

That is, Q^* is a maximizer of $G = G_1 - G_2$, and we will show it must be the only one. Let $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ be distinct from Q^*. We denote by $M(t) = (M_1(t), \ldots, M_n(t))$ the displacement interpolations between the marginals, i.e.,

$$M_i(t) = Q_i^* \circ ((1-t)\mathrm{Id} + tT_i)^{-1},$$

where $T_i : \mathbb{R} \to \mathbb{R}$ is the Q_i^*-a.s. unique nondecreasing function satisfying $Q_i^* \circ T_i^{-1} = Q_i$. Since Q_i^* and Q_i are distinct, there exists i such that T_i is different from the identity map on a set with strictly positive Q_i^*-measure. Writing out the expression of G_1,

$$G_1(M(t)) = \int_{\mathbb{R}^n} f\Big((1-t)x_1 + tT_1(x_1), \ldots, (1-t)x_n + tT_n(x_n)\Big) \prod_{i=1}^n Q_i^*(\mathrm{d}x_i),$$

we see that $t \mapsto G_1(M(t))$ is strictly concave because f is strictly concave and $Q \neq Q^*$. Tensorization of entropy yields $G_2(M(t)) = \sum_{i=1}^n H(M_i(t))$, and it is well known that differential entropy is displacement convex [55, Theorem 5.15(i)]. That is, $t \mapsto H(M_i(t))$ is convex for each i. We deduce that $t \mapsto G(M(t))$ is strictly concave. This proves uniqueness: if Q were also an optimizer, then $G(M(1)) = G(Q) = G(Q^*) = G(M(0))$ would imply $G(M(t)) > G(Q^*)$ for some $t \in (0,1)$. $\square$

Remark 4.3.7. We do not expect uniqueness in Lemma 4.3.6 to hold under mere concavity of f. The challenge is that the differential entropy functional is displacement convex, but not strictly so..

In some of the following proofs, some shorthand notation will be useful. For $Q \in \mathcal{P}(\mathbb{R}^n)$, let us write Q_{-i} for the marginal of $(X_j)_{j\neq i}$ under Q. For $x \in \mathbb{R}^n$ let us write $x_{-i} = (x_j)_{j\neq i}$ and, with some abuse of notation, $f(x) = f(x_i, x_{-i})$.

Lemma 4.3.8. *If $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ satisfies the fixed point equation* (4.38)*, then Q is κ-log-concave.*

Proof. By (4.38), the density of Q is proportional to $e^{\hat{F}}$, where $\hat{F}(x) = \sum_{i=1}^n \hat{f}_i(x_i)$ and $\hat{f}_i$ is given by (4.38). By the κ-concavity of f, for any $y, z \in \mathbb{R}$ and $t \in [0,1]$, we have

$$\begin{aligned}
&\hat{f}_i(tz + (1-t)y) + \frac{\kappa}{2}(tz + (1-t)y)^2 \\
&\quad = \int_{\mathbb{R}^{n-1}} \Big[f(tz + (1-t)y, x_{-i}) + \frac{\kappa}{2}(tz + (1-t)y)^2 \Big] Q_{-i}(x_{-i})\,\mathrm{d}x_{-i} \\
&\quad \geq \int_{\mathbb{R}^{n-1}} \Big[tf(z, x_{-i}) + t\frac{\kappa}{2}z^2 + (1-t)f(y, x_{-i}) + (1-t)\frac{\kappa}{2}y^2 \Big] Q_{-i}(x_{-i})\,\mathrm{d}x_{-i} \\
&\quad = t\hat{f}_i(z) + t\frac{\kappa}{2}z^2 + (1-t)\hat{f}_i(y) + (1-t)\frac{\kappa}{2}y^2.
\end{aligned}$$

This shows that $\hat{f}_i$ is κ-concave, and thus so is $\hat{F}$. $\square$

The next proposition, in conjunction with Proposition 4.3.5, shows that the optimizers of (4.37) and the solutions of the fixed point problem (4.3) are exactly the same.

Proposition 4.3.9 (Fixed point to optimality)**.** *Let $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ satisfy $f \in L^1(Q)$ and the fixed point problem* (4.38)*. Then Q has strictly positive density a.e. and is a maximizer of* (4.37).

Proof. We first show that Q has strictly positive density a.e. Since $Q = Q_1 \times \cdots \times Q_n$ satisfies the fixed point equation (4.38), each Q_i has a density with exponent

$$\hat{f}_i(x_i) = \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_n) \prod_{j\neq i} Q_j(\mathrm{d}x_j) \geq -c_1 e^{c_2 x_i^2} \prod_{j\neq i} \int_{\mathbb{R}} e^{c_2 x_j^2} Q_j(x_j)\,\mathrm{d}x_j$$

for every $x_i \in \mathbb{R}$. From Lemma 4.3.8 we know that Q is κ-log-concave. Since $c_2 < \kappa/2$, we deduce that $\int_{\mathbb{R}} e^{c_2 x_j^2} Q_j(x_j)\mathrm{d}x_j < \infty$. Thus $\hat{f}_i(x_i) > -\infty$ for all $x_i \in \mathbb{R}$.

Define $G(R) \coloneqq \int_{\mathbb{R}^n} f\,\mathrm{d}R - H(R)$ for $R \in \mathcal{P}(\mathbb{R}^n)$. Let Q^* be an optimizer of $\sup\{G(R) : R \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)\}$, which exists uniquely by Lemma 4.3.6. By Proposition 4.3.5, we have $f \in L^1(Q^*)$, and Q^* satisfies the fixed point equation (4.38). The argument given in the previous paragraph implies that Q^* has a strictly positive density a.e. To complete the proof, we must show that $G(Q) \geq G(Q^*)$.

For $i = 1, \ldots, n$, let $T_i : \mathbb{R} \to \mathbb{R}$ denote the unique nondecreasing function satisfying $Q_i \circ T_i^{-1} = Q_i^*$, and define $M_i(t) = Q_i \circ ((1-t)\mathrm{Id} + tT_i)^{-1}$. Let $M(t) = M_1(t) \times \cdots \times M_n(t)$, so that $G(M(t)) = g_1(t) - g_2(t)$, where

$$g_1(t) := \int_{\mathbb{R}^n} f\big((1-t)x_1 + tT_1(x_1), \ldots, (1-t)x_n + tT_n(x_1)\big) \prod_{i=1}^n Q_i(\mathrm{d}x_i),$$
$$g_2(t) := H\big(M_1(t) \times \cdots \times M_n(t)\big) = \sum_{i=1}^n H(M_i(t)).$$

Let us write g'^{+} for the right-derivative of a real-valued function g, when it exists. Note that T_i is a.e. differentiable, as it is monotone. Using [55, Theorem 5.30], we may compute the right-derivatives at zero as

$$g_1'^{+}(0) = \sum_{i=1}^n \int_{\mathbb{R}^n} \partial_i f(x)\big(T_i(x_i) - x_i\big)Q(x)\,\mathrm{d}x,$$
$$g_2'^{+}(0) = -\sum_{i=1}^n \int_{\mathbb{R}} \left(T_i'(x_i) - 1\right) Q_i(x_i)\,\mathrm{d}x_i.$$

We wish to rewrite both terms in more useful forms.

We first claim that

$$\int_{\mathbb{R}^n} \partial_i f(x)\big(T_i(x_i) - x_i\big)Q(x)\,\mathrm{d}x = \int_{\mathbb{R}} \hat{f}_i'(x_i)\big(T_i(x_i) - x_i\big)Q_i(x_i)\,\mathrm{d}x_i, \tag{4.40}$$

where $\hat{f}_i$ is defined as in (4.38). To see this, note that $\hat{f}_i(x_i) = \mathbb{E}_Q[f(x_i, X_{-i})]$ for all $x_i \in \mathbb{R}$, so

$$\hat{f}_i'^{+}(x_i) = \lim_{h \downarrow 0} h^{-1} \mathbb{E}_Q[f(x_i + h, X_{-i}) - f(x_i, X_{-i})].$$

By the concavity of f, the difference quotient $[f(x_i + h, X_{-i}) - f(x_i, X_{-i})]/h$ increases as $h \downarrow 0$, and it is bounded from below for $0 < h \leq h_0$ by $[f(x_i + h_0, X_{-i}) - f(x_i, X_{-i})]/h_0$, which has finite Q-expectation for a.e. choice of $h_0 > 0$ by Fubini's theorem since $f \in L^1(Q)$. Hence, by monotone convergence,

$$\hat{f}_i'^{+}(x_i) = \mathbb{E}_Q[\partial_i f(x_i, X_{-i})]. \tag{4.41}$$

Moreover, this quantity is finite and nonincreasing in x_i because $\hat{f}_i$ is a concave real-valued function. In addition, $\hat{f}_i' = \hat{f}_i'^{+}$ a.e. since concave functions are a.e. differentiable. Using (4.41), we see that the right-hand side of (4.40) equals $\mathbb{E}_Q[\mathbb{E}_Q[\partial_i f(X) \mid X_i](T_i(X_i) - X_i)]$, which yields (4.40).

We next integrate by parts to get

$$-\int_{\mathbb{R}} (T_i'(x_i) - 1)\, Q_i(x_i)\, \mathrm{d}x_i = \int_{\mathbb{R}} (T_i(x_i) - x_i)\, Q_i'(x_i)\, \mathrm{d}x_i. \tag{4.42}$$

To justify this carefully, we use Lebesgue-Stieltjes integration by parts: Note that the probability density function of Q_i is absolutely continuous because it is proportional to $e^{\hat{f}_i}$, and $\hat{f}_i$ is absolutely continuous as a concave function. Let F_{Q_i} and $F_{Q_i^*}$ denote the CDFs of Q_i and Q_i^* respectively. Recalling that $T_i = F_{Q_i^*}^{-1} \circ F_{Q_i}$ is the monotone map pushing Q_i forward to Q_i^*, and that both Q_i and Q_i^* admit strictly positive densities, the function T_i is absolutely continuous. Hence, there is no jump term in the integration by parts, and we must only show that the boundary terms vanish. For this it suffices to show that there exist sequences $x_n^{\pm} \to \pm\infty$ such that

$$\lim_{n \to \infty} (T_i(x_n^{\pm}) - x_n^{\pm}) Q_i(x_n^{\pm}) = 0.$$

If this were not the case, it would imply that $|T_i(x) - x| Q_i(x) \leq (|T_i(x)| + |x|) Q_i(x)$ is bounded away from zero for $|x|$ sufficiently large. This would in turn imply that $\int_{\mathbb{R}} (|T_i(x_i)| + |x_i|) Q_i(x_i)\, \mathrm{d}x_i =$

∞, contradicting the fact that

$$\int_{\mathbb{R}} (|T_i(x_i)| + |x_i|) Q_i(x_i)\,\mathrm{d}x_i = \int_{\mathbb{R}} |x_i| Q_i^*(x_i)\,\mathrm{d}x_i + \int_{\mathbb{R}} |x_i| Q_i(x_i)\,\mathrm{d}x_i < \infty.$$

Both integrals are finite because Q_i and Q_i^* are κ-log-concave by Lemma 4.3.8 and thus admit finite moments of every order. With (4.42) and (4.40) now justified, we see that the right-derivative of $G(M(t))$ at $t = 0$ is

$$g_1'^{+}(0) - g_2'^{+}(0) = \sum_{i=1}^n \int_{\mathbb{R}} \big(\hat{f}_i'(x_i) Q_i(x_i) - Q_i'(x_i)\big)\big(T_i(x_i) - x_i\big)\,\mathrm{d}x_i.$$

This is in fact zero, because Q_i is proportional to $e^{\hat{f}_i}$. We saw in the proof of Lemma 4.3.6 that $G(M(t))$ is concave. Since we now know that it has vanishing right-derivative at $t = 0$, it follows that $G(M(1)) \le G(M(0))$. That is, $G(Q^*) \le G(Q)$, which completes the proof. $\square$

Proof of Theorem 4.1.1. Let S_{opt} denote the set of maximizers in (4.37), and let S_{fix} denote the set of $Q^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ satisfying $f \in L^1(Q^*)$ and the fixed point equation (4.38). Proposition 4.3.5 shows that $S_{\mathrm{opt}} \subset S_{\mathrm{fix}}$. Proposition 4.3.9 shows conversely that $S_{\mathrm{opt}} \supset S_{\mathrm{fix}}$, so in fact $S_{\mathrm{opt}} = S_{\mathrm{fix}}$. Lemma 4.3.6 shows that this set is a singleton. Its unique element Q^* is κ-log-concave by Lemma 4.3.8 and has strictly positive density a.e. by Proposition 4.3.9. This proves claims (1)–(3) of Theorem 4.1.1.

To prove (4), recall the identity (4.12), which shows that

$$R_f = \log \int_{\mathbb{R}^n} e^{f(x)}\,\mathrm{d}x - \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q) \right) = H(Q^* \,|\, P).$$

The κ-log-concavity of P and the log-Sobolev inequality (Theorem 4.3.1) imply

$$H(Q^* \,|\, P) \le \frac{1}{2\kappa} \int_{\mathbb{R}^n} \left| \nabla \log \frac{dQ^*}{dP} \right|^2 \mathrm{d}Q^*.$$

Since $Q^* = Q_1^* \times \cdots \times Q_n^*$ is a product measure, we have $\partial_i \log Q^*(x) = \partial_i \log Q_i^*(x_i)$ for $x \in \mathbb{R}^n$

and note that the derivative exists almost everywhere because $\log Q_i^*$ is concave. We saw in (4.41) in the proof of Proposition 4.3.9 that the following identity is valid for almost every $x_i \in \mathbb{R}$, with the expectation on the right-hand side being finite:

$$\partial_i \log Q_i^*(x_i) = \partial_i \mathbb{E}_{Q^*}[f(X) \,|\, X_i = x_i] = \mathbb{E}_{Q^*}[\partial_i f(X) \,|\, X_i = x_i].$$

Thus,

$$\begin{aligned}
H(Q^* \,|\, P) &\leq \frac{1}{2\kappa} \int_{\mathbb{R}^n} \sum_{i=1}^n \Big| \partial_i \log Q_i^*(x_i) - \partial_i f(x) \Big|^2 Q^*(\mathrm{d}x) \\
&= \frac{1}{2\kappa} \mathbb{E}_{Q^*} \sum_{i=1}^n \left(\mathbb{E}_{Q^*}[\partial_i f(X) \,|\, X_i] - \partial_i f(X) \right)^2 \\
&= \frac{1}{2\kappa} \mathbb{E}_{Q^*} \sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i).
\end{aligned}$$

This yields the first bound in (4.5). Recall that Q_{-i}^* denotes the law of $(X_j)_{j \neq i}$, which equals the conditional law of $(X_j)_{j \neq i}$ given X_i under Q^* by independence. The measure Q_{-i}^* is κ-log-concave because Q_j^* is for each j. Hence, it obeys a Poincaré inequality (Theorem 4.3.2), $\mathrm{Var}_{Q_{-i}^*}(\varphi) \leq \kappa^{-1} \int_{\mathbb{R}^{n-1}} |\nabla \varphi|^2 \, \mathrm{d}Q_{-i}^*$, for any C^1 function $\varphi \in L^1(Q_{-i}^*)$. Applying this to $\partial_i f$ with coordinate i fixed,

$$\mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i) \leq \frac{1}{\kappa} \sum_{j \neq i} \mathbb{E}_{Q^*}[|\partial_{ij} f(X)|^2 \,|\, X_i].$$

Complete the proof of the second inequality of (4.5) by using the tower property to get

$$\frac{1}{2\kappa} \mathbb{E}_{Q^*} \sum_{i=1}^n \mathrm{Var}_{Q^*}(\partial_i f(X) \,|\, X_i) \leq \frac{1}{2\kappa^2} \sum_{i=1}^n \sum_{j \neq i} \mathbb{E}_{Q^*}[|\partial_{ij} f(X)|^2]. \qquad \square$$

4.3.2 Proof of Corollary 4.1.4

Let $f(x) := g(x) + \sum_{i=1}^n V_i(x_i)$. Then $\int_{\mathbb{R}^n} e^g \,\mathrm{d}\rho = \int_{\mathbb{R}^n} e^{f(x)}\,\mathrm{d}x$, and the concavity of g and κ-concavity of V_i imply that f is κ-concave. Note also that for any $Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$,

$$\int_{\mathbb{R}^n} g\,\mathrm{d}Q - H(Q\,|\,\rho) = \int_{\mathbb{R}^n} f\,\mathrm{d}Q - H(Q).$$

This shows that the optimization problems (4.4) and (4.9) are the same. Moreover, the fixed point problems (4.8) and (4.3) admit exactly the same solutions: Q_i^* solves (4.3) if and only if it solves (4.8). With these identifications, applying Theorem 4.1.1 to f immediately proves claims (1–3) of Corollary 4.1.4. Finally, with Q_i^* solving (4.38) (or equivalently (4.8)), we have

$$R_g^\rho = R_f \le \frac{1}{\kappa^2}\sum_{1\le i<j\le n}\mathbb{E}_{Q^*}[|\partial_{ij}f(X)|^2] = \frac{1}{\kappa^2}\sum_{1\le i<j\le n}\mathbb{E}_{Q^*}[|\partial_{ij}g(X)|^2],$$

because $\partial_{ij}f = \partial_{ij}g$ for all $i \neq j$. This proves claim (4) of Corollary 4.1.4. □

4.3.3 Proof of Proposition 4.1.6

Note that $\int_{\mathbb{R}^n} f(x+y)\,\gamma_t(\mathrm{d}x) < \infty$ for each $y \in \mathbb{R}^n$ by the growth assumption on f. The function

$$y \mapsto \int_{\mathbb{R}^n} f\,\mathrm{d}\gamma_{y,t} - H(\gamma_{y,t}\,|\,\gamma_t) = \int_{\mathbb{R}^n} f(x+y)\,\gamma_t(\mathrm{d}x) - \frac{1}{2t}|y|^2$$

is $(1/t)$-concave and thus bounded from above. It admits a unique maximizer obtained by setting the gradient equal to zero; the first order condition is precisely (4.18). Let $P(\mathrm{d}x) = Z^{-1}e^{f(x)}\gamma_t(\mathrm{d}x)$. The simple identity

$$\log\int_{\mathbb{R}^n} e^f\,\mathrm{d}\gamma_t - \left(\int_{\mathbb{R}^n} f\,\mathrm{d}\gamma_{y,t} - H(\gamma_{y,t}\,|\,\gamma_t)\right) = H(\gamma_{y,t}\,|\,P),$$

valid for all $y \in \mathbb{R}^n$, implies that

$$\log \int_{\mathbb{R}^n} e^f \, \mathrm{d}\gamma_t - \sup_{y \in \mathbb{R}^n} \left(\int_{\mathbb{R}^n} f \, \mathrm{d}\gamma_{y,t} - H(\gamma_{y,t} \,|\, \gamma_t) \right) = \inf_{y \in \mathbb{R}^n} H(\gamma_{y,t} \,|\, P).$$

The right-hand side is equal to $H(\gamma_{y^*,t} \,|\, P)$. The measure P is $(1/t)$-log-concave, so we may use the log-Sobolev inequality (Theorem 4.3.1) to get

$$\begin{aligned} H(\gamma_{y^*,t} \,|\, P) &\leq \frac{t}{2} \int_{\mathbb{R}^n} \left| \nabla \log \frac{\mathrm{d}\gamma_{y^*,t}}{\mathrm{d}P} \right|^2 \mathrm{d}\gamma_{y^*,t} = \frac{t}{2} \int_{\mathbb{R}^n} \left| \nabla \log \frac{\mathrm{d}\gamma_{y^*,t}}{\mathrm{d}\gamma_t} - \nabla \log \frac{\mathrm{d}P}{\mathrm{d}\gamma_t} \right|^2 \mathrm{d}\gamma_{y^*,t} \\ &= \frac{t}{2} \int_{\mathbb{R}^n} \left| \frac{1}{t} y^* - \nabla f(x) \right|^2 \gamma_{y^*,t}(\mathrm{d}x) \\ &= \frac{t}{2} \sum_{i=1}^n \mathrm{Var}_{\gamma_{y^*,t}}(\partial_i f), \end{aligned}$$

where the last step follows from (4.18). Using the Gaussian Poincaré inequality (or Theorem 4.3.2), this is bounded by the second term on the right-hand side of (4.19). □

4.3.4 Asymptotic independence

Proof of first inequality in (4.16). Let $P, Q \in \mathcal{P}(\mathbb{R}^n)$. Let $k_1, \ldots, k_m$ be positive integers summing to n. Suppose $P_1, \ldots, P_m$ are the marginals of P on $\mathbb{R}^{k_1}, \ldots, \mathbb{R}^{k_m}$, and define the marginals $Q_1, \ldots, Q_m$ similarly. Then

$$\sum_{i=1}^m \mathcal{W}_2^2(P_i, Q_i) \leq \mathcal{W}_2^2(P, Q).$$

Indeed, to prove this, let (X, Y) be an optimal coupling of (P, Q). Let X_i be the $\mathbb{R}^{k_i}$ coordinate, for $i = 1, \ldots, m$, and similarly define Y_i. Then (X_i, Y_i) is a coupling of (P_i, Q_i), and so

$$\mathcal{W}_2^2(P, Q) = \mathbb{E}\left[|X - Y|^2\right] = \mathbb{E}\left[\sum_{i=1}^m |X_i - Y_i|^2\right] \geq \sum_{i=1}^m \mathcal{W}_2^2(P_i, Q_i).$$

Now, let $1 \leq k \leq n$, and let $m = \lfloor n/k \rfloor$. Let Π be the set of vectors $(S_1, \ldots, S_m)$ of disjoint k-

element subsets of $[n]$. Let Q_{S_i} and P_{S_i} denote the corresponding marginals, on those coordinates in $S_i \subset [n]$. Note that $\mathcal{W}_2^2(P_{S_i}, Q_{S_i})$ does not depend on the order of the elements of S_i. Then

$$\sum_{i=1}^{m} \mathcal{W}_2^2(P_{S_i}, Q_{S_i}) \le \mathcal{W}_2^2(P_{S_1\cup\cdots\cup S_m}, Q_{S_1\cup\cdots\cup S_m}) \le \mathcal{W}_2^2(P, Q).$$

If $(S_1, \ldots, S_m)$ is chosen uniformly at random from Π and i is chosen uniformly at random from $[m]$, then the marginal law of S_i is the same as the law of a uniformly random choice of k-element subset of $[n]$. In particular,

$$\frac{1}{\binom{n}{k}} \sum_{S\subset[n],\, |S|=k} \mathcal{W}_2^2(P_S, Q_S) = \frac{1}{|\Pi|} \sum_{(S_1,\ldots,S_m)\in\Pi} \frac{1}{m} \sum_{i=1}^{m} \mathcal{W}_2^2(P_{S_i}, Q_{S_i}).$$

Combining the two previous inequalities yields

$$\frac{1}{\binom{n}{k}} \sum_{S\subset[n],\, |S|=k} \mathcal{W}_2^2(P_S, Q_S) \le \frac{1}{m}\mathcal{W}_2^2(P, Q) = \frac{1}{\lfloor n/k \rfloor}\mathcal{W}_2^2(P, Q). \qquad \square$$

Proof of Corollary 4.1.2. By the triangle inequality, the square root of the left-hand side of (4.7) is no more than $A_1 + A_2$, where we define

$$A_1 := \mathbb{E}_P\left[\left(\frac{1}{n}\sum_{i=1}^{n}\varphi(X_i) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_P[\varphi(X_i)]\right)^2\right]^{1/2},$$
$$A_2 := \left|\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}_P[\varphi(X_i)] - \mathbb{E}_{Q^*}[\varphi(X_i)])\right|.$$

Recall that $|\varphi'| \le 1$. Using Kantorovich duality and (4.16) with $k = 1$,

$$A_2^2 \le \frac{1}{n}\sum_{i=1}^{n}\mathcal{W}_1^2(P_i, Q_i^*) \le \frac{1}{n}\sum_{i=1}^{n}\mathcal{W}_2^2(P_i, Q_i^*) \le \frac{2R_f}{\kappa n}.$$

Apply the Poincaré inequality (Theorem 4.3.2) to the function $x \mapsto (1/n)\sum_{i=1}^n \varphi(x_i)$ to get

$$A_1^2 = \mathrm{Var}_P\left(\frac{1}{n}\sum_{i=1}^n \varphi(X_i)\right) \le \frac{1}{\kappa n^2}\sum_{i=1}^n \mathbb{E}_P[|\varphi'(X_i)|^2] \le \frac{1}{\kappa n}.$$

Combine these two bounds to complete the proof. $\square$

4.4 Gibbs measure proofs

This section proves the results of Chapter 4.2.1. Throughout, the function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as in (4.6) and satisfies Assumption 4.2.1.

Proof of Lemma 4.2.2. Compute two derivatives to find, for all $i \neq j$,

$$\partial_{ii} f(x) = V''(x_i) + \sum_{j\neq i} J_{ij}K''(x_i - x_j), \qquad \partial_{ij} f(x) = -J_{ij}K''(x_i - x_j).$$

Hence, for any $x, z \in \mathbb{R}^n$,

$$z^\top \nabla^2 f(x) z = \sum_{i,j=1}^n z_i z_j \partial_{ij} f(x) = \sum_{i=1}^n z_i^2 V''(x_i) + \sum_{i,j=1}^n \left(z_i^2 - z_i z_j\right) J_{ij} K''(x_i - x_j).$$

Using the evenness of K'' and the symmetry of J,

$$\sum_{i,j=1}^n \left(z_i^2 - z_i z_j\right) J_{ij} K''(x_i - x_j) = \frac{1}{2}\sum_{i,j=1}^n (z_i - z_j)^2 J_{ij} K''(x_i - x_j).$$

Since $K'' \le 0$ and $J_{ij} \ge 0$, we find that this quantity is nonpositive. By κ-concavity of V,

$$z^\top \nabla^2 f(x) z \le \sum_{i=1}^n z_i^2 V''(x_i) \le -\kappa |z|^2,$$

which shows that f is κ-concave. $\square$

Proof of Corollary 4.2.3. Note that f is C^2 and κ-concave. Also, the assumptions on $|V|$ and $|K''|$ in Assumption 4.2.1 clearly imply that $|f|$ satisfies the growth assumption in Theorem 4.1.1.

Therefore, Theorem 4.1.1 applies. Let Q^* be given as therein. Computing derivatives as above, we have

$$R_f \leq \frac{1}{\kappa^2} \sum_{1 \leq i < j \leq n} \mathbb{E}_{Q^*}[|\partial_{ij} f(X)|^2] = \frac{1}{\kappa^2} \sum_{1 \leq i < j \leq n} J_{ij}^2 \mathbb{E}_{Q^*}\left[|K''(X_i - X_j)|^2\right]. \tag{4.43}$$

Using the assumption on K'', we find

$$\mathbb{E}_{Q^*}\left[|K''(X_i - X_j)|^2\right] \leq a \mathbb{E}_{Q^*}\left[e^{b|X_i - X_j|}\right]. \tag{4.44}$$

By assumption, $X_i - X_j$ has mean zero under Q^*. It follows from the κ-log-concavity of Q^* that the law of $X_i - X_j$ is $(\kappa/2)$-log-concave (see, e.g., [161, Theorem 3.7(a) and Theorem 3.8]). This implies that it is subgaussian in the sense that

$$\mathbb{E}_{Q^*}[e^{s(X_i - X_j)}] \leq e^{s^2/\kappa}, \qquad \forall s \in \mathbb{R}.$$

Indeed, this can be deduced from the log-Sobolev inequality (Theorem 4.3.1) via Herbst's argument or [3, Theorem 1.3]. Thus, using (4.44),

$$\mathbb{E}_{Q^*}\left[|K''(X_i - X_j)|^2\right] \leq a \mathbb{E}_{Q^*}\left[e^{b(X_i - X_j)} + e^{b(X_j - X_i)}\right] \leq 2ae^{b^2/\kappa}.$$

Combine this with (4.43) to complete the proof. □

4.4.1 Doubly stochastic matrices

We now turn to the proof of Theorem 4.2.5. We first need a straightforward lemma about displacement convexity, which is likely known.

Lemma 4.4.1. *Let $Q_1, \ldots, Q_n \in \mathcal{P}(\mathbb{R})$ and $t_1, \ldots, t_n \in [0, 1]$ be such that $\sum_{i=1}^n t_i = 1$. Then* *there exists a random vector $X = (X_1, \ldots, X_n)$ such that $X_i \sim Q_i$ for each i and*

$$H\biggl(\mathrm{Law}\biggl(\sum_{i=1}^{n} t_i X_i\biggr)\biggr) \le \sum_{i=1}^{n} t_i H(Q_i).$$

Proof. The proof is by induction on n, with the case $n = 1$ holding trivially. Assume that the statement of the lemma is true for some n. Let $Q_1, \ldots, Q_{n+1} \in \mathcal{P}(\mathbb{R})$ and $t_1, \ldots, t_{n+1} \in [0, 1]$ be such that $\sum_{i=1}^{n+1} t_i = 1$. Without loss of generality, assume that $t_{n+1} < 1$ and that $Q_1, \ldots, Q_{n+1}$ have finite entropy, as otherwise there is nothing to prove. For $i = 1, \ldots, n$, define $\tilde{t}_i := t_i/(1 - t_{n+1})$, so that $\sum_{i=1}^{n} \tilde{t}_i = 1$. By assumption, we may find a random vector $(X_1, \ldots, X_n)$ such that $X_i \sim Q_i$ for each $i = 1, \ldots, n$ and

$$H(\widetilde{Q}) \le \sum_{i=1}^{n} \tilde{t}_i H(Q_i), \tag{4.45}$$

where $\widetilde{Q}$ denotes the law of $\widetilde{X} := \sum_{i=1}^{n} \tilde{t}_i X_i$. By absolute continuity, there is a unique nondecreasing function $T : \mathbb{R} \to \mathbb{R}$ such that $\widetilde{Q} \circ T^{-1} = Q_{n+1}$. The entropy functional is displacement convex [55, Theorem 5.15(i)], which means that the function

$$[0, 1] \ni t \mapsto H\bigl(\widetilde{Q} \circ (tT + (1 - t)\mathrm{Id})^{-1}\bigr)$$

is convex. In particular, letting $X_{n+1} = T(\widetilde{X})$, we find

$$\begin{aligned} H\bigl(\mathrm{Law}(t_{n+1}X_{n+1} + (1 - t_{n+1})\widetilde{X})\bigr) &= H\bigl(\widetilde{Q} \circ (t_{n+1}T + (1 - t_{n+1})\mathrm{Id})^{-1}\bigr) \\ &\le t_{n+1}H(Q_{n+1}) + (1 - t_{n+1})H(\widetilde{Q}). \end{aligned}$$

By (4.45) and the definition of $\tilde{t}_i$, we have $(1 - t_{n+1})H(\widetilde{Q}) \le \sum_{i=1}^{n} t_i H(Q_i)$, completing the proof. $\square$

Proof of Theorem 4.2.5(1). Let us abbreviate

$$M_n := \sup_{Q \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} M_n(Q), \tag{4.46}$$

where we define

$$\begin{aligned} M_n(Q) &:= \int_{\mathbb{R}^n} f \, \mathrm{d}Q - H(Q) \\ &= \sum_{i=1}^n \int_{\mathbb{R}} V(x) \, Q_i(\mathrm{d}x) + \frac{1}{2} \sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y) \, Q_i(\mathrm{d}x) Q_j(\mathrm{d}y) - \sum_{i=1}^n H(Q_i), \end{aligned}$$

where the last equality used the symmetry of J and K, the fact that the diagonal entries of J are zero, and the tensorization of entropy. Recall that $\log \int_{\mathbb{R}^n} e^f \, \mathrm{d}x = M_n + R_f$, by definition of R_f. We will complete the proof by showing that

$$M_n = n \sup_{Q \in \mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}} V \, \mathrm{d}Q + \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y) Q(\mathrm{d}x) Q(\mathrm{d}y) - H(Q) \right), \tag{4.47}$$

and that the optimizer $Q^* = Q_1^* \times \cdots \times Q_n^*$ in (4.4) must be i.i.d., or $Q_1^* = \cdots = Q_n^*$. Indeed, the i.i.d. form of Q^* implies $\mathbb{E}_{Q^*}[X_i - X_j] = 0$ for all i, j. Using this and the assumption $\mathrm{Tr}(J^2) = o(n)$, we may apply Corollary 4.2.3 to deduce that $R_f / n \to 0$, and Theorem 4.2.5(1) follows.

The proof of the inequality $(\geq)$ in (4.47) is immediate upon restricting the supremum in (4.46) to i.i.d. measures and using $\sum_{i,j=1}^n J_{ij} = n$:

$$\begin{aligned} M_n &\geq \sup_{Q \in \mathcal{P}(\mathbb{R})} \left(n \int_{\mathbb{R}} V(x) \, Q(\mathrm{d}x) + \frac{1}{2} \sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y) \, Q(\mathrm{d}x) Q(\mathrm{d}y) - nH(Q) \right) \\ &= n \sup_{Q \in \mathcal{P}(\mathbb{R})} \left(\int_{\mathbb{R}} V \, \mathrm{d}Q + \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y) Q(\mathrm{d}x) Q(\mathrm{d}y) - H(Q) \right). \end{aligned}$$

To prove the inequality $(\leq)$ in (4.47), fix $Q = Q_1 \times \cdots \times Q_n \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ arbitrarily. By Lemma

4.4.1, there exists a random vector $X = (X_1, \ldots, X_n)$ such that $X_i \sim Q_i$ for all i and

$$H(\overline{Q}) \le \frac{1}{n} \sum_{i=1}^n H(Q_i). \tag{4.48}$$

where $\overline{Q}$ denotes the law of $\frac{1}{n}\sum_{i=1}^n X_i$. Using the concavity V, we find

$$\frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} V(x)\, Q_i(\mathrm{d}x) = \mathbb{E}\Bigg[\frac{1}{n}\sum_{i=1}^n V(X_i)\Bigg] \le \int_{\mathbb{R}} V \,\mathrm{d}\overline{Q}. \tag{4.49}$$

Let $Y = (Y_1, \ldots, Y_n)$ be an independent copy of X. Using the concavity of K and the fact that $\sum_i J_{ij} = \sum_j J_{ij} = 1$, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y)\, Q_i(\mathrm{d}x) Q_j(\mathrm{d}y) \\
&\qquad = \mathbb{E}\Bigg[\frac{1}{n} \sum_{i,j=1}^n J_{ij} K(X_i - Y_j)\Bigg] \le \mathbb{E}\Bigg[K\Bigg(\frac{1}{n}\sum_{i,j=1}^n J_{ij}(X_i - Y_j)\Bigg)\Bigg] \\
&\qquad = \mathbb{E}\Bigg[K\Bigg(\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{j=1}^n Y_j\Bigg)\Bigg] = \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y)\, \overline{Q}(\mathrm{d}x)\overline{Q}(\mathrm{d}y).
\end{aligned} \tag{4.50}$$

Combining (4.48), (4.49), and (4.50), we see that

$$M_n(Q)/n \le \int_{\mathbb{R}} V \,\mathrm{d}\overline{Q} + \frac{1}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y)\, \overline{Q}(\mathrm{d}x)\overline{Q}(\mathrm{d}y) - H(\overline{Q}) = M_n(\overline{Q}^{\otimes n})/n.$$

In other words, for an arbitrary choice of product measure Q, we may increase $M_n(Q)$ by replacing Q with the i.i.d. measure $\overline{Q}^{\otimes n}$. This completes the proof. $\square$

Proof of Theorem 4.2.5(2). We first justify the uniqueness claim. From part (3) of Theorem 4.1.1, we know that the optimizer $Q^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ in (4.46) is unique. It follows from the previous paragraph that this unique optimizer is in fact i.i.d., i.e., $Q^* = Q^{\otimes n}$, where $Q \in \mathcal{P}(\mathbb{R})$ is the (necessarily unique) optimizer of (4.47), which does not depend on n. This proves the desired uniqueness.

Turning to the proof of (4.22), recall that $R_f/n \to 0$, and use Corollary 4.1.2 and the aforementioned i.i.d. form of the optimizer $Q^* = Q^{\otimes n}$ to deduce that, for any 1-Lipschitz function φ,

$$\mathbb{E}_P\left[\left(\frac{1}{n}\sum_{i=1}^n \varphi(X_i) - \int_{\mathbb{R}} \varphi \, \mathrm{d}Q\right)^2\right] \le \frac{(1+\sqrt{2R_f})^2}{\kappa n} \to 0, \quad \text{as } n \to \infty.$$

This is enough to deduce that $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges to Q weakly in law. □

4.4.2 Graphons proofs

This section is devoted to the proof of Theorem 4.2.8. For $W \in \mathcal{W}$ and any measurable function $\psi : \mathbb{R}^2 \to \mathbb{R}$ bounded from above, define $T_{W,\psi} : \mathcal{P}_{\mathrm{Unif}}([0,1] \times \mathbb{R}) \to \mathbb{R}$ by

$$T_{W,\psi}(\mu) := \mathbb{E}_{\mu^{\otimes 2}}\left[\psi(X_1, X_2) W(U_1, U_2)\right].$$

where (U_1, X_1) and (U_2, X_2) are independent with law μ. Note that $W \ge 0$ is integrable, so $T_{W,\psi}(\mu)$ is well-defined in $[-\infty, \infty)$. Let $\overline{\mu} := \mathrm{Unif}[0,1] \times \rho$, and define $I : \mathcal{P}_{\mathrm{Unif}}([0,1] \times \mathbb{R}) \to [0, \infty]$ by

$$I(\mu) := H(\mu \,|\, \overline{\mu}) = \int_0^1 H(\mu_u \,|\, \rho) \, \mathrm{d}u,$$

with the second identity coming from the chain rule for relative entropy [54, Theorem B.2.1], and we recall that $\rho(dx) = e^{V(x)}\mathrm{d}x$ is a probability measure. We begin with two lemmas pertaining to the continuity of $T_{W,\psi}$.

Lemma 4.4.2. *Let $\mathcal{K} \subset \mathbb{R}$ be a compact interval. Let $\psi : \mathbb{R}^2 \to \mathbb{R}$ be supported on $\mathcal{K}^2$ and continuous when restricted to $\mathcal{K}^2$.*

(1) If $\{W_\ell\}$ converges to W in strong cut metric and $W_\ell, W \ge 0$, then

$$\sup_{\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})} \left| T_{W_\ell,\psi}(\mu) - T_{W,\psi}(\mu) \right| \to 0.$$

(2) The map $\mu \to T_{W,\psi}(\mu)$ is continuous on $\{\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1] \times \mathbb{R}) : I(\mu) < \infty\}$, with respect

to the topology of weak convergence.

Proof. We begin with (1). Let $\mathcal{V}$ denote the space of functions $\phi : \mathcal{K}^2 \mapsto \mathbb{R}$ of the form

$$\phi(x,y) = \sum_{i=1}^{L} c_i a_i(x) b_i(y), \tag{4.51}$$

for some $L \in \mathbb{N}$, $c_i \in \mathbb{R}$, and continuous functions $a_i, b_i : \mathcal{K} \to [0,1]$. It is easy to check that $\mathcal{V}$ is closed under multiplication, contains the constant functions, separates points in $\mathcal{K}^2$, and is a vector subspace of the space $C(\mathcal{K}^2)$ of continuous real-valued functions on $\mathcal{K}^2$. By the Stone-Weierstrass Theorem, we deduce that $\mathcal{V}$ is dense in $C(\mathcal{K}^2)$ with the supremum norm. Let $\varepsilon > 0$, and find $\phi \in \mathcal{V}$ such that $|\psi - \phi| < \varepsilon$ uniformly on $\mathcal{K}^2$. Extend the domain of ϕ to $\mathbb{R}^2$ by setting $\phi = 0$ on the complement of $\mathcal{K}^2$. Then for all $\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1] \times \mathbb{R})$,

$$\left|T_{W_\ell,\psi}(\mu) - T_{W_\ell,\phi}(\mu)\right| \le \varepsilon \|W_\ell\|_{L_1[0,1]^2}, \quad \left|T_{W,\psi}(\mu) - T_{W,\phi}(\mu)\right| \le \varepsilon \|W\|_{L_1[0,1]^2}.$$

Consequently, using the triangle inequality, we have

$$\left|T_{W_\ell,\psi}(\mu) - T_{W,\psi}(\mu)\right| \le \varepsilon \|W_\ell\|_{L_1[0,1]^2} + \varepsilon \|W\|_{L_1[0,1]^2} + \left|T_{W_\ell,\phi}(\mu) - T_{W,\phi}(\mu)\right|. \tag{4.52}$$

Since ϕ is of the form (4.51), we have

$$T_{W_\ell,\phi}(\mu) = \sum_{i=1}^{L} c_i \int_{[0,1]^2} \bar{a}_i(u) \bar{b}_i(v) W_\ell(u,v) \mathrm{d}u \mathrm{d}v,$$

where we define $\bar{a}_i(u) := \mathbb{E}_\mu[a_i(X) \,|\, U = u]$, and $\bar{b}_i$ similarly. This yields

$$\left|T_{W_\ell,\phi}(\mu) - T_{W,\phi}(\mu)\right| \le \sum_{i=1}^{L} |c_i| d_\square(W_\ell, W). \tag{4.53}$$

Noting that $d_\square(W_\ell, W) \to 0$ implies $\|W_\ell\|_{L_1[0,1]^2} \to \|W\|_{L_1[0,1]^2}$, we may now combine (4.52) and (4.53), sending $\ell \to \infty$ and then $\varepsilon \to 0$, to prove the claim (1).

To prove (2), let μ_k be a sequence of measures in $\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ converging weakly to μ_∞, such that $I(\mu_\infty)<\infty$. Let W_ℓ be a sequence of continuous functions in $\mathcal{W}$ converging in $L_1[0,1]^2$ to W. By the triangle inequality,

$$\big|T_{W,\psi}(\mu_k)-T_{W,\psi}(\mu_\infty)\big|\le 2\sup_{\nu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})}\big|T_{W,\psi}(\nu)-T_{W_\ell,\psi}(\nu)\big|+\big|T_{W_\ell,\psi}(\mu_k)-T_{W_\ell,\psi}(\mu_\infty)\big|.$$

The first term converges to 0 as $\ell\to\infty$, by part (1) and the fact that convergence in $L_1[0,1]^2$ implies convergence in strong cut metric. The second term converges to 0 for fixed ℓ as $k\to\infty$, using the fact that μ_k converges weakly to μ_∞, and the set of discontinuity points of $W_\ell(\cdot,\cdot)\psi(\cdot,\cdot)$ is contained in $[0,1]^2\times\partial(\mathcal{K}^2)$, which has measure 0 under $\mu_\infty^{\otimes 2}$ (as μ_∞ is absolutely continuous with respect to Lebesgue measure on $[0,1]\times\mathbb{R}$). $\square$

Lemma 4.4.3. *Suppose μ_m is a sequence of measures in $\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ converging weakly to μ_∞. Let $\psi:\mathbb{R}^2\mapsto\mathbb{R}$ be a continuous function, and let $W\in L_1[0,1]^2$. For $1\le m\le\infty$, let $(U_1^m,X_1^m),(U_2^m,X_2^m)\overset{i.i.d.}{\sim}\mu_m$. Then*

$$W(U_1^m,U_2^m)\psi(X_1^m,X_2^m)\xrightarrow{d}W(U_1^\infty,U_2^\infty)\psi(X_1^\infty,X_2^\infty).$$

Proof. If W is continuous, then the claim is immediate. For a general W, we proceed as follows: Fix $\varepsilon>0$, and let $\mathcal{K}$ be a compact set such that $\mathbb{P}(X_1^m\in\mathcal{K},X_2^m\in\mathcal{K})\ge 1-\varepsilon$, which is again possible by tightness of $\{(X_1^m,X_2^m)\}_{m\in\mathbb{N}}$. Let g be a continuous function with

$$\|W-g\|_{L_1[0,1]^2}<\frac{\varepsilon}{1\vee\sup_{x,y\in\mathcal{K}}|\psi(x,y)|}.$$

Then on the event $\{X_1^m\in\mathcal{K},X_2^m\in\mathcal{K}\}$, we have

$$\Big|W(U_1^m,U_2^m)\psi(X_1^m,X_2^m)-g(U_1^m,U_2^m)\psi(X_1^m,X_2^m)\Big|\le\varepsilon.$$

Thus, for any continuous function $\phi : \mathbb{R} \to [0,1]$ which is 1-Lipschitz, we have

$$\Big|\mathbb{E}\phi\big(W(U_1^m, U_2^m)\psi(X_1^m, X_2^m)\big) - \mathbb{E}\phi\big(g(U_1^m, U_2^m)\psi(X_1^m, X_2^m)\big)\Big| \le 2\varepsilon.$$

Finally,

$$\mathbb{E}\phi\big(g(U_1^m, U_2^m)\psi(X_1^m, X_2^m)\big) \to \mathbb{E}\phi(g(U_1^\infty, U_2^\infty)\psi(X_1^\infty, X_2^\infty)),$$

by the result for continuous functions. Thus

$$\limsup_{m\to\infty}\Big|\mathbb{E}\phi(W(U_1^m, U_2^m)\psi(X_1^m, X_2^m)) - \mathbb{E}\phi(W(U_1^\infty, U_2^\infty)\psi(X_1^\infty, X_2^\infty))\Big| \le 4\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the proof of the lemma is complete. □

Proof of Theorem 4.2.8 (1).

We begin with some notation. For a measurable function $\psi : \mathbb{R}^2 \to \mathbb{R}$ which is bounded from above, define $M_n^\psi := \sup_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} M_n^\psi(Q)$, where

$$\begin{aligned} M_n^\psi(Q) &:= \sum_{i=1}^n \int_{\mathbb{R}} V(x)\, Q_i(\mathrm{d}x) + \sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}}\int_{\mathbb{R}} \psi(x,y)\, Q_i(\mathrm{d}x)Q_j(\mathrm{d}y) - \sum_{i=1}^n H(Q_i) \\ &= \sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}}\int_{\mathbb{R}} \psi(x,y)\, Q_i(\mathrm{d}x)Q_j(\mathrm{d}y) - \sum_{i=1}^n H(Q_i\,|\,\rho). \end{aligned} \tag{4.54}$$

Letting $\widetilde{K} : \mathbb{R}^2 \to \mathbb{R}$ by $\widetilde{K}(x,y) = K(x-y)/2$, we are most interested in the choice $\psi = \widetilde{K}$, but treating a general ψ will be helpful for a truncation argument. Let Q^* be as in Theorem 4.1.1. With this notation, we have $\log \int_{\mathbb{R}^n} e^{f(x)}\mathrm{d}x = M_n^{\widetilde{K}} + R_f$. Corollary 4.2.3 and the assumption that $\mathrm{Tr}(J^2) = o(n)$ imply that $R_f/n \to 0$, and to prove Theorem 4.2.8 it will thus suffice to show that

$$\lim_{n\to\infty} M_n^\psi/n = \sup_{\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})} \big(T_{W,\psi}(\mu) - I(\mu)\big) \tag{4.55}$$

for any continuous function $\psi \le 0$.

To this effect, use the assumption that $\{nJ\}_{n\ge1}$ converges in weak cut metric to W to conclude the existence of a sequence of permutations $\{\pi_n\}_{n\geq 1}$ with $\pi_n \in S_n$, such that $\{nJ^{(\pi_n)}\}_{n\geq 1}$ converges in strong cut metric to W, where $J^{(\pi_n)}_{ij} := J_{\pi_n(i)\pi_n(j)}$ for $1 \leq i, j \leq n$. Since π_n is a permutation, for any $Q = Q_1 \times \cdots \times Q_n \in \mathcal{P}_{\rm pr}(\mathbb{R}^n)$ we can write

$$M^\psi_n(Q) = \sum_{i=1}^n \int_{\mathbb{R}} V(x)\tilde{Q}_i(\mathrm{d}x) + \sum_{i,j=1}^n J_{\pi_n(i)\pi_n(j)} \int_{\mathbb{R}}\int_{\mathbb{R}} \psi(x,y)\tilde{Q}_i(\mathrm{d}x)\tilde{Q}_j(\mathrm{d}y) - \sum_{i=1}^n H(\tilde{Q}_i),$$

where $\tilde{Q}_i := Q_{\pi_n(i)} \in \mathcal{P}(\mathbb{R})$. Thus

$$\sup_{Q\in\mathcal{P}_{\rm pr}(\mathbb{R}^n)} M^\psi_n(Q) = \sup_{\tilde{Q}\in\mathcal{P}_{\rm pr}(\mathbb{R}^n)} \widetilde{M}^\psi_n(\tilde{Q}),$$

where $\widetilde{M}^\psi_n(\cdot)$ defined similarly to $M^\psi_n(\cdot)$ in (4.54), but with J replaced by $J^{(\pi_n)}$. Since $nJ^{(\pi_n)}$ converges to W in strong cut metric, by replacing J with $J^{(\pi_n)}$ without loss of generality we assume throughout the rest of the proof that nJ converges in strong cut metric to W.

To prove (4.55), we need the following construction which essentially embeds $\mathcal{P}_{\rm pr}(\mathbb{R}^n)$ into $\mathcal{P}_{\rm Unif}([0,1]\times\mathbb{R})$ for all n. For any $Q = Q_1 \times \cdots \times Q_n \in \mathcal{P}_{\rm pr}(\mathbb{R}^n)$, define a probability measure $\mu_n(Q) \in \mathcal{P}_{\rm Unif}([0,1]\times\mathbb{R})$ as follows: If $(U,X) \sim \mu_n(Q)$, then $U \sim \mathrm{Unif}[0,1]$, and the conditional law of X given $\{(i-1)/n < U \leq i/n]\}$ is given by Q_i. Then we have

$$\begin{aligned} T_{W_{nJ},\psi}(\mu_n(Q)) &= \frac{1}{n}\sum_{i,j=1}^n J_{ij} \int_{\mathbb{R}}\int_{\mathbb{R}} \psi(x,y)\, Q_i(\mathrm{d}x)Q_j(\mathrm{d}y), \\ I(\mu_n(Q)) &= \frac{1}{n}\sum_{i=1}^n H(Q_i|\rho), \end{aligned}$$

and so

$$M^\psi_n(Q)/n = T_{W_{nJ},\psi}(\mu_n(Q)) - I(\mu_n(Q)). \tag{4.56}$$

As a final preparation for the proof of (4.55), we argue that $\inf_n M^\psi_n/n > -\infty$. To see this, take $B \subset \mathbb{R}$ to be any compact set of positive ρ-measure, and define $\hat{\rho} \ll \rho$ by $d\hat{\rho}/d\rho = 1_B/\rho(B)$.

Let $Q_i = \hat{\rho}$ for $i = 1, \ldots, n$, and $Q = Q_1 \times \cdots \times Q_n$. Then

$$\frac{1}{n}\sum_{i=1}^{n} H(Q_i \,|\, \rho) = H(\hat{\rho} \,|\, \rho) = -\log \rho(B) < \infty,$$

and also

$$T_{W_{nJ},\psi}(\mu_n(Q)) \geq -\|W_{nJ}\|_{L_1[0,1]^2} \sup_{x,y\in B} |\psi(x,y)|.$$

Since ψ is continuous, it is bounded on the compact set B. Since W_{nJ} converges in strong cut metric to W, we have $\|W_{nJ}\|_{L_1[0,1]^2} \to \|W\|_{L_1[0,1]^2}$, and thus the right-hand side is bounded. This proves that $\inf_n M_n^\psi/n > -\infty$. We now prove the upper and lower bounds in (4.55) separately.

Proof of the upper bound in (4.55): Let $Q^n = Q_1^n \times \cdots \times Q_n^n \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ be any near-optimizer of $M_n^\psi(\cdot)$, meaning

$$M_n^\psi(Q^n) \geq M_n^\psi - o(n). \tag{4.57}$$

Note that $M_n^\psi(Q^n)/n$ is bounded from below by some constant C, as shown just above. Since J has nonnegative entries and $\psi \leq 0$, we have $T_{W_{nJ},\psi} \leq 0$ which implies

$$C \leq M_n^\psi(Q^n)/n = T_{W_{nJ},\psi}(\mu_n(Q^n)) - I(\mu_n(Q^n)) \leq -I(\mu_n(Q^n)).$$

This implies $\sup_n I(\mu_n(Q^n)) < \infty$. Since the sub-level sets of I are weakly compact, the sequence $(\mu_n(Q^n))$ has a limit point. Let μ_∞ be any limit point. Lower semicontinuity of $I(\cdot)$ gives $I(\mu_\infty) < \infty$. For each $m \in \mathbb{N}$, define $\psi_m(x,y) := \psi(x,y)1_{\{|x|,|y|\leq m\}}$. Note that $\psi \leq \psi_m \leq 0$, and thus $T_{W_{nJ},\psi} \leq T_{W_{nJ},\psi_m}$. By part (1) of Lemma 4.4.2,

$$\sup_{\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})} \left| T_{W_{nJ},\psi_m}(\mu) - T_{W,\psi_m}(\mu) \right| \to 0,$$

for all $m \in \mathbb{N}$. Therefore, for all m,

$$\limsup_{n\to\infty} T_{W_{nJ},\psi}(\mu_n(Q^n)) \leq \limsup_{n\to\infty} T_{W_{nJ},\psi_m}(\mu_n(Q^n)) \leq \limsup_{n\to\infty} T_{W,\psi_m}(\mu_n(Q^n)) = T_{W,\psi_m}(\mu_\infty),$$

where the last step uses part (2) of Lemma 4.4.2. The left-hand side above does not depend on m, and thus

$$\limsup_{n\to\infty} T_{W_{nJ},\psi}(\mu_n(Q^n)) \le \inf_{m\in\mathbb{N}} T_{W,\psi_m}(\mu_\infty) = T_{W,\psi}(\mu_\infty),$$

where the last equality follows from the monotone convergence theorem and the fact that $\psi_m \downarrow \psi$ pointwise. Using the lower semicontinuity of I, we deduce

$$T_{W,\psi}(\mu_\infty) - I(\mu_\infty) \ge \limsup_{n\to\infty} \big(T_{W_{nJ},\psi}(\mu_n(Q^n)) - I(\mu_n(Q^n))\big) = \limsup_{n\to\infty} M_n^\psi(Q^n)/n.$$

Bound the left-hand side by a supremum to prove the upper bound in (4.55). Moreover, once we prove (4.55), then this argument shows the following: for any near-optimizing sequence $Q^n = \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ in the sense of (4.57), the sequence $\{\mu_n(Q^n)\}$ is tight, and for any limit point μ_∞ of $\{\mu_n(Q^n)\}$ it holds that μ_∞ is an optimizer for the right-hand side of (4.55).

Proof of the lower bound in (4.55): To prove the lower bound in (4.55), we first claim that

$$\sup_{\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})} \big(T_{W,\psi}(\mu) - I(\mu)\big) = \sup_{\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R}),\ \text{compact support}} \big(T_{W,\psi}(\mu) - I(\mu)\big). \tag{4.58}$$

The inequality ($\ge$) is obvious. To prove the reverse, let $\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ such that $I(\mu) < \infty$, and define $\mu^m \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ with compact support by setting $\mathrm{d}\mu^m/\mathrm{d}\mu = 1_{[0,1]\times[-m,m]}/\mu([0,1]\times[-m,m])$, which is well defined for large enough m. Then

$$\begin{aligned} I(\mu_m) = H(\mu^m \,|\, \overline{\mu}) &= \int_{[0,1]\times\mathbb{R}} \log\frac{\mathrm{d}\mu^m}{\mathrm{d}\overline{\mu}}\,\mathrm{d}\mu^m \\ &= \int_{[0,1]\times\mathbb{R}} \log\frac{\mathrm{d}\mu^m}{\mathrm{d}\mu}\,\mathrm{d}\mu^m + \int_{[0,1]\times\mathbb{R}} \log\frac{\mathrm{d}\mu}{\mathrm{d}\overline{\mu}}\,\mathrm{d}\mu^m. \end{aligned}$$

The second term converges to $I(\mu)$ by dominated convergence. The first term equals $-\log\mu([0,1]\times[-m,m])$ and vanishes as $m\to\infty$. Finally, since $W\ge 0$ and $\psi\le 0$, it is straightforward to check by monotone convergence that $T_{W,\psi}(\mu_m) \to T_{W,\psi}(\mu)$, and thus $T_{W,\psi}(\mu_m) - I(\mu_m) \to$

$T_{W,\psi}(\mu) - I(\mu)$ as $m \to \infty$. This proves (4.58).

Now, to prove the lower bound in (4.55), we let $\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ with compact support and with $I(\mu) < \infty$, and note that necessarily $\mu \ll \overline{\mu}$. By defining $h(\cdot,\cdot) := \frac{\mathrm{d}\mu}{\mathrm{d}\overline{\mu}}$, we have $\int_{\mathbb{R}} h(u,\cdot)\,\mathrm{d}\rho = 1$ for a.e. $u \in [0,1]$ since both μ and $\overline{\mu}$ have uniform first marginal. For each $i \in [n]$, define $h_i^n : \mathbb{R} \to [0,\infty)$ by

$$h_i^n(x) := n \int_{(i-1)/n}^{i/n} h(u,x)\,\mathrm{d}u.$$

By Fubini's theorem, $\int_{\mathbb{R}} h_i^n\,\mathrm{d}\rho = 1$ for all i. We may thus define $Q^n = Q_1^n \times \cdots \times Q_n^n \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ by setting $\frac{\mathrm{d}Q_i^n}{\mathrm{d}\rho} = h_i^n$, and define $\mu_n(Q^n)$ as before; note for later use the key identity $\frac{\mathrm{d}\mu_n(Q^n)}{\mathrm{d}\overline{\mu}}(u,x) = h^n_{\lceil nu \rceil}(x)$. If $\mathcal{K}$ denotes a compact interval such that $[0,1]\times\mathcal{K}$ contains the support of μ, then $[0,1]\times\mathcal{K}$ also contains the support of $\mu_n(Q^n)$, and we may replace ψ by $\psi 1_{\mathcal{K}^2}$ in the following argument. Recalling the formula (4.56) for $M_n^{\psi}(Q)$, we may use part (1) of Lemma 4.4.2 to get

$$T_{W,\psi}(\mu_n(Q^n)) - \frac{1}{n} M_n^{\psi}(Q^n) - I(\mu_n(Q^n)) \to 0. \tag{4.59}$$

To complete the proof of the lower bound, we will show that

$$\lim_{n\to\infty} T_{W,\psi}(\mu_n(Q^n)) = T_{W,\psi}(\mu), \quad \text{and} \quad I(\mu_n(Q^n)) \le I(\mu),\ \forall n. \tag{4.60}$$

Once (4.60) is established, it will follow from the lower semicontinuity of I that $I(\mu_n(Q^n)) \to I(\mu)$, and we use (4.59) to deduce

$$\liminf_{n\to\infty} M_n^{\psi}/n \ge \lim_{n\to\infty} M_n^{\psi}(Q^n)/n = T_{W,\psi}(\mu) - I(\mu).$$

This holds for every $\mu \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ of compact support satisfying $I(\mu) < \infty$. Hence, taking the supremum and recalling (4.58) yields the desired lower bound in (4.55).

It remains to prove (4.60). Note that

$$\int_{[0,1]\times\mathbb{R}}\Big|\frac{\mathrm{d}\mu}{\mathrm{d}\overline{\mu}}-\frac{\mathrm{d}\mu_n(Q^n)}{\mathrm{d}\overline{\mu}}\Big|\,\mathrm{d}\overline{\mu}=\int_0^1\int_{\mathbb{R}}|h(u,x)-h^n_{\lceil nu\rceil}(x)|\,\rho(\mathrm{d}x)\,\mathrm{d}u$$
$$=\mathbb{E}_{\overline{\mu}}\big|h(U,X)-\mathbb{E}_{\overline{\mu}}[h(U,X)|\mathcal{F}_n]\big|,$$

where $\mathcal{F}_n$ is the σ-field generated by $(\lceil nU\rceil, X)$. The right-hand side converges to 0 by Levy's upwards convergence theorem, since $\mathbb{E}_{\overline{\mu}}|h(U,X)|=1<\infty$. Thus the probability measure $\mu_n(Q^n)$ converges in total variation to μ, and the first claim in (4.60) follows from part (2) of Lemma 4.4.2. To prove the second claim in (4.60), use convexity of $\varphi(x):=x\log x$ for $x\geq 0$, along with Jensen's inequality, to get

$$I(\mu_n(Q^n))=\mathbb{E}_{\overline{\mu}}\varphi(\mathbb{E}_{\overline{\mu}}[h(U,X)|\mathcal{F}_n])\leq\mathbb{E}_{\overline{\mu}}\varphi(h(U,X))=I(\mu).$$

This proves (4.60), completing the proof of the lower bound, and thus Theorem 4.2.8(1). $\square$

Proof of Theorem 4.2.8(2).

We first discuss the optimization problem. The functional to be optimized can be written as

$$\Phi(\mu):=\frac{1}{2}\int_{[0,1]\times\mathbb{R}}\int_{[0,1]\times\mathbb{R}}W(u,v)K(x-y)\,\mu(\mathrm{d}u,\mathrm{d}x)\mu(\mathrm{d}v,\mathrm{d}y)-\int_0^1H(\mu_u\,|\,\rho)\,\mathrm{d}u$$

We will show the existence of an optimizer via the weak upper semicontinuity: Since $W\geq 0$ and $K\leq 0$, monotone convergence yields

$$2T_{W,\widetilde{K}}(\mu)=\mathbb{E}_{\mu^{\otimes 2}}[W(U_1,U_2)K(X_1-X_2)]=\inf_{m>0}\mathbb{E}_{\mu^{\otimes 2}}\left[\big(W(U_1,U_2)K(X_1-X_2)\big)\vee(-m)\right].$$

For each m, the expectation appearing on the right-hand side is continuous as a function of $\mu\in\mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$, by Lemma 4.4.3. Hence, the left-hand side is upper semicontinuous. Since relative entropy is lower semicontinuous with compact sub-level sets, the existence of an optimizer follows.

We prove uniqueness of the optimizer via displacement convexity. Letting $\widetilde{K}(x,y) = (1/2)K(x-y)$, we may rewrite $\Phi(\mu) = \Phi_1(\mu) + T_{W,\widetilde{K}}(\mu) - \Phi_2(\mu)$, where we define

$$\Phi_1(\mu) := \int_0^1 \int_{\mathbb{R}} V(x)\,\mu_u(\mathrm{d}x)\,\mathrm{d}u, \qquad \Phi_2(\mu) := \int_0^1 H(\mu_u)\,\mathrm{d}u,$$

where we used the simple identity $H(\nu \,|\, \rho) = H(\nu) - \int_{\mathbb{R}^n} V\,\mathrm{d}\nu$. Let $\mu^0, \mu^1 \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$ be two optimizers, written in disintegrated form as $\mathrm{d}u\mu^i_u(\mathrm{d}x)$ for $i = 0,1$. Let $F^i_u(x) = \mu^u_i(-\infty, x]$ denote the CDF, with generalized inverse $\overline{F}^i_u(y) := \inf\{x \in \mathbb{R} : y \le F^i_u(x)\}$. Then, for each $u \in [0,1]$, $T_u(x) := \overline{F}^1_u(F^0_u(x))$ denotes the unique nondecreasing function with $\mu^0_u \circ T_u^{-1} = \mu^1_u$. Since $F^i_u(x)$ is right-continuous in x and measurable in u, it is jointly measurable in (u,x), and the same is easily seen to be true for $\overline{F}^i_u(x)$ and thus $T_u(x)$. Consider the map $\overline{T} : [0,1]\times\mathbb{R} \to [0,1]\times\mathbb{R}$ given by $\overline{T}(u,x) = (u, T_u(x))$. Define the interpolation $\mu^t := \mu^0 \circ ((1-t)\mathrm{Id} + t\overline{T})^{-1}$ for each $t \in [0,1]$. Then we have

$$\begin{aligned} T_{W,\widetilde{K}}(\mu^t) &= \frac{1}{2}\int_{[0,1]\times\mathbb{R}}\int_{[0,1]\times\mathbb{R}} W(u,v)K(x-y)\,\mu^t(\mathrm{d}u,\mathrm{d}x)\mu^t(\mathrm{d}v,\mathrm{d}y) \\ &= \frac{1}{2}\int_{[0,1]\times\mathbb{R}}\int_{[0,1]\times\mathbb{R}} W(u,v)K\big((1-t)(x-y) + t(T_u(x) - T_u(y))\big)\mu^0(\mathrm{d}u,\mathrm{d}x)\mu^0(\mathrm{d}v,\mathrm{d}y). \end{aligned}$$

Since K is concave and $W \ge 0$, $t \mapsto T_{W,\widetilde{K}}(\mu^t)$ is concave. Note also that

$$\Phi_2(\mu^t) = \int_0^1 H\big(\mu^0_u \circ ((1-t)\mathrm{Id} + tT_u)^{-1}\big)\,\mathrm{d}u$$

is a convex function of t, by the displacement convexity of entropy [55, Theorem 5.15(i)]. By the κ-concavity of V, the function $t \mapsto \Phi_1(\mu^t)$ is strictly concave, and we find that $t \mapsto \Phi(\mu^t)$ is strictly concave. Since μ^0 and μ^1 are both optimizers, we have $\Phi(\mu^0) = \Phi(\mu^1)$. Hence, we must have $\mu^0 = \mu^1$, as otherwise the strict concavity would be contradicted.

With existence and uniqueness of the optimizer settled, we lastly prove the claim (4.24) in part (2) of Theorem 4.2.8. Note that Theorem 4.1.1 implies uniqueness of the optimizer Q^n in $\sup_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} M_n(Q)$ for each n. Since Q^n is optimal and thus a fortiori near-optimal, we may use the following fact proven in the course of proving the upper bound in Theorem 4.2.8(1): The sequence $\{\mu_n(Q^n)\}$ is tight (since Q^n is), and any limit point is an optimizer for the right-hand side of (4.55). We have just shown the latter optimizer to be unique, and let us denote it $\mu^* \in \mathcal{P}_{\mathrm{Unif}}([0,1]\times\mathbb{R})$. Thus, $\mu_n(Q^n) \to \mu^*$ weakly. From part (1) and Corollary 4.1.2, for any bounded 1-Lipschitz function $\varphi : \mathbb{R} \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \mathbb{E}_P\left[\left(\frac{1}{n}\sum_{i=1}^n \varphi(X_i) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{Q^n}[\varphi(X_i)]\right)^2\right] = 0.$$

Note that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{Q^n}[\varphi(X_i)] = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} \varphi(x)\, Q_i^n(\mathrm{d}x) = \int_{[0,1]\times\mathbb{R}} \varphi(x)\, \mu_n(Q^n)(\mathrm{d}u, \mathrm{d}x).$$

Using the weak convergence $\mu_n(Q^n) \to \mu^*$, the right-hand side converges to

$$\int_{[0,1]\times\mathbb{R}} \varphi(x)\, \mu^*(\mathrm{d}u, \mathrm{d}x) = \int_{\mathbb{R}} \varphi\, \mathrm{d}R^*, \ \text{ where } R^* := \int_0^1 \mu_u^*\, \mathrm{d}u.$$

We deduce that $\frac{1}{n}\sum_{i=1}^n \varphi(X_i) \to \int_{\mathbb{R}} \varphi\, \mathrm{d}R^*$ in probability for each bounded Lipschitz φ. This is enough to deduce the convergence in distribution $\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \to R^*$. $\square$

4.4.3 Proof of Lemma 4.2.11

We first prove (1). When f is even, we claim that (the density of) Q^* is also even, which completes the proof because it implies $\mathbb{E}_{Q^*}[X_i] = 0$ for all i. To show that Q^* is even, let $R_i(x) := Q_i^*(-x)$ for each $x \in \mathbb{R}$ and $i = 1, \ldots, n$. Let $R = R_1 \times \cdots \times R_n$. Then $\int_{\mathbb{R}^n} f\, \mathrm{d}R = \int_{\mathbb{R}^n} f\, \mathrm{d}Q$ by evenness of f, and clearly $H(Q) = H(R)$. Hence, R is also an optimizer of (4.4), and we deduce $R = Q^*$ by uniqueness of the optimizer.

We prove (2) by showing in this case that $Q_i^* = Q_j^*$ for all i, j. Suppose f is invariant with respect to a transitive set S of permutations of $[n]$. Fix $i, j \in \{1, \ldots, n\}$. Choose $\pi \in S$ such that $\pi(i) = j$, which is possible by the assumed transitivity of S. Let $R_k = Q_{\pi(k)}^*$ for each $k = 1, \ldots, n$, and let $R = R_1 \times \cdots \times R_n$. The invariance of f under S ensures that $\int_{\mathbb{R}^n} f \, \mathrm{d}R = \int_{\mathbb{R}^n} f \, \mathrm{d}Q^*$. Clearly, $H(R) = H(Q^*)$. Hence, R is also an optimizer of (4.4), and we deduce that $R = Q^*$ by uniqueness. Since $\pi(i) = j$, this implies $Q_i^* = R_i = Q_j^*$.

4.5 Stochastic control proofs

As explained in Remark 4.2.15, the optimal admissible pair (α, X) for (4.27) is given by

$$\alpha_g(t,x) = \nabla_x \log \mathbb{E}[e^{ng(x+B_T-B_t)}], \tag{4.61}$$

with $X = (X_t)_{t\in[0,T]}$ being the Brownian bridge with terminal law $P(\mathrm{d}x) = Z^{-1}e^{ng(x)}\gamma_T(\mathrm{d}x)$. Letting $\mathbb{P}$ denote the Wiener measure on $C([0,T];\mathbb{R}^n)$, the law $\mathbb{Q}^P$ of this process X can be characterized as the unique minimizer of $\mathbb{Q} \mapsto H(\mathbb{Q}\,|\,\mathbb{P})$ among $\mathbb{Q}$ with time-T marginal equal to P; see [162, Proposition 6] or [163, Lemma 10]. This minimizer satisfies

$$H(\mathbb{Q}^P\,|\,\mathbb{P}) = H(P\,|\,\gamma_T) = \frac{1}{2}\mathbb{E}\left[\int_0^T |\alpha_g(t,X_t)|^2 \,\mathrm{d}t\right]. \tag{4.62}$$

Note that $H(P\,|\,\gamma_T) < \infty$, and so the pair (α_g, X) is admissible in the sense of Chapter 4.2.3.

Proof of Corollary 4.2.14. Once the formulas (4.30) and (4.32) are established, the final claim follows immediately from Corollary 4.1.4, applied with $V_i(x) = -x^2/(2T)$ for $i = 1, \ldots, n$ and $\kappa = 1/T$.

To prove (4.30) and (4.32), we begin with the inequality ($\le$). Let (α, X) denote any admissible pair, and let $\mathbb{Q}$ denote the law of $X = (X_t)_{t\in[0,T]}$. A well known argument using Girsanov's theorem [163, Proposition 1] yields

$$H(\mathbb{Q}\,|\,\mathbb{P}) \le \frac{1}{2}\mathbb{E}\int_0^T |\alpha(t,X_t)|^2 \,\mathrm{d}t = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\int_0^T |\alpha_i(t,X_t)|^2 \mathrm{d}t.$$

With $\mathbb{Q}_T$ denoting the law of X_T, note that marginalizing (at time T) does not increase entropy:

$H(\mathbb{Q}\,|\,\mathbb{P}) \geq H(\mathbb{Q}_T\,|\,\gamma_T)$. Thus,

$$\mathbb{E}\left[g(X_T) - \frac{1}{2n}\sum_{i=1}^n \int_0^T |\alpha_i(t,X_t)|^2 \mathrm{d}t\right] \leq \int_{\mathbb{R}^n} g\,\mathrm{d}\mathbb{Q}_T - \frac{1}{n}H(\mathbb{Q}_T\,|\,\gamma_T)$$
$$\leq \sup_{Q\in\mathcal{P}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g\,\mathrm{d}Q - \frac{1}{n}H(Q\,|\,\gamma_T)\right).$$

Taking a supremum over all admissible pairs (α, X) proves the inequality ($\leq$) in (4.30). Now, if (α, X) is an *distributed* admissible pair, then the same chain of inequalities holds, but also $\mathbb{Q}_T$ is a product measure. We can thus deduce (4.32) in the same manner.

The inequality ($\geq$) in (4.30) and (4.32) follows quickly from the entropy identity (4.62). Starting with (4.30), let $X = (X_t)_{t\in[0,T]}$ be the Brownian bridge with terminal law $P(\mathrm{d}x) = Z^{-1}e^{ng(x)}\gamma_T(\mathrm{d}x)$. Let α_g be given as in (4.61). By the Gibbs variational principle [54, Proposition 1.4.2], the supremum in (4.30) is attained by $Q = P$. Using $X_T \sim P$ and (4.62), we obtain

$$\sup_{Q\in\mathcal{P}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g\,\mathrm{d}Q - \frac{1}{n}H(Q\,|\,\gamma_T)\right) = \int_{\mathbb{R}^n} g\,dP - \frac{1}{n}H(P\,|\,\gamma_T)$$
$$= \mathbb{E}\left[g(X_T) - \frac{1}{2n}\int_0^T |\alpha_g(t,X_t)|^2\,\mathrm{d}t\right] \leq V_{\mathrm{orig}}.$$

This proves ($\geq$) in (4.30), and also proves that (α_g, X) is optimal. Similarly, to prove the inequality ($\geq$) in (4.32), let $Q^* \in \mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)$ be the unique optimizer in (4.32), which we know by Corollary 4.1.4 to take the form stated in Corollary 4.2.14. Let $X = (X_t)_{t\in[0,T]}$ be the Brownian bridge with terminal law Q^*. Define $\alpha_{\widehat{g}}$ as in (4.61), with $\widehat{g}(x) = \sum_{i=1}^n \mathbb{E}_{Q^*}[g(X)|X_i = x_i]$ in place of g. Using $X_T \sim Q^*$ and (4.62), we obtain

$$\sup_{Q\in\mathcal{P}_{\mathrm{pr}}(\mathbb{R}^n)} \left(\int_{\mathbb{R}^n} g\,\mathrm{d}Q - \frac{1}{n}H(Q\,|\,\gamma_T)\right) = \int_{\mathbb{R}^n} g\,\mathrm{d}Q^* - \frac{1}{n}H(Q^*\,|\,\gamma_T)$$
$$= \mathbb{E}\left[g(X_T) - \frac{1}{2n}\int_0^T |\alpha_{\widehat{g}}(t,X_t)|^2\,\mathrm{d}t\right] \leq V_{\mathrm{distr}}.$$

Indeed, note that $(\alpha_{\widehat{g}}, X)$ is an admissible distributed pair because Q^* is a product measure. This proves ($\geq$) in (4.32), and also proves that $(\alpha_{\widehat{g}}, X)$ is optimal. $\square$

References

[1] K. Marton, "Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration," *The Annals of Probability*, vol. 24, no. 2, pp. 857–866, 1996.

[2] M. Talagrand, "Transportation cost for gaussian and other product measures," *Geometric and Functional Analysis*, vol. 6, no. 3, pp. 587–600, 1996.

[3] S. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities," *Journal of Functional Analysis*, vol. 163, no. 1, pp. 1–28, 1999.

[4] N. Gozlan and C. Léonard, "A large deviation approach to some transportation cost inequalities," *Probability Theory and Related Fields*, vol. 139, no. 1-2, pp. 235–283, 2007.

[5] A. Guillin, C. Léonard, L. Wu, and N. Yao, "Transportation-information inequalities for Markov processes," *Probability theory and related fields*, vol. 144, no. 3-4, pp. 669–695, 2009.

[6] A. Guillin, C. Léonard, F. Wang, and L. Wu, "Transportation-information inequalities for Markov processes II: Relations with other functional inequalities," *arXiv preprint arXiv:0902.2101*, 2009.

[7] Y. Ma, R. Wang, and L. Wu, "Transportation-information inequalities for continuum gibbs measures," *Electronic Communications in Probability*, vol. 16, pp. 600–613, 2011.

[8] F. Gao, A. Guillin, and L. Wu, "Bernstein-type concentration inequalities for symmetric Markov processes," *Theory of Probability & Its Applications*, vol. 58, no. 3, pp. 358–382, 2014.

[9] P. Cattiaux and A. Guillin, "Semi log-concave Markov diffusions," in *Séminaire de probabilités XLVI*, Springer, 2014, pp. 231–292.

[10] Y. Liu, "A new characterization of quadratic transportation-information inequalities," *Probability Theory and Related Fields*, vol. 168, no. 3-4, pp. 675–689, 2017.

[11] N. Wang and L. Wu, "Transport-information inequalities for Markov chains," *The Annals of Applied Probability*, vol. 30, no. 3, pp. 1276–1320, Jun. 2020.

[12] N. Gozlan, "A characterization of dimension free concentration in terms of transportation inequalities," *The Annals of Probability*, vol. 37, no. 6, pp. 2480–2498, 2009.

[13] D. Lacker and L. C. Yeung, “A characterization of transportation-information inequalities for markov processes in terms of dimension-free concentration,” *Annales de l’Institut Henri Poincare (B) Probabilites et statistiques*, vol. 59, no. 1, pp. 364–377, 2023.

[14] R. Jordan, D. Kinderlehrer, and F. Otto, “The Variational Formulation of the Fokker–Planck Equation,” *SIAM J. Math. Anal.*, vol. 29, no. 1, pp. 1–17, 1998.

[15] I. Karatzas, W. Schachermayer, and B. Tschiderer, “A trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions,” *Theory Probab. Appl.*, vol. 66, no. 4, pp. 668–707, 2022.

[16] H. McKean, “A class of Markov processes associated with nonlinear parabolic equations,” *Proc. Natl. Acad. Sci. USA*, vol. 56, no. 6, pp. 1907–1911, 1966.

[17] J. Carrillo, R. McCann, and C. Villani, “Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates,” *Rev. Mat. Iberoam.*, vol. 19, no. 3, pp. 971–1018, 2003.

[18] ——, “Contractions in the 2-Wasserstein Length Space and Thermalization of Granular Media,” *Arch. Ration. Mech. Anal.*, vol. 179, no. 2, pp. 217–263, 2006.

[19] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, second, ser. Lect. Math. ETH Zürich. Basel: Birkhäuser, 2008.

[20] M. Agueh, N. Ghoussoub, and X. Kang, “Geometric inequalities via a general comparison principle for interacting gases,” *Geom. Funct. Anal.*, vol. 14, no. 1, pp. 215–244, 2004.

[21] F. Otto and C. Villani, “Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality,” *J. Funct. Anal.*, vol. 173, no. 2, pp. 361–400, 2000.

[22] B. Tschiderer and L. C. Yeung, “A trajectorial approach to relative entropy dissipation of mckean–vlasov diffusions: Gradient flows and hwbi inequalities,” *Bernoulli*, vol. 29, no. 1, pp. 725–756, 2023.

[23] D. Kim and L. C. Yeung, “A trajectorial approach to entropy dissipation of degenerate parabolic equations,” *arXiv preprint arXiv:2210.16158*, 2022.

[24] S. Chatterjee and A. Dembo, “Nonlinear large deviations,” *Advances in Mathematics*, vol. 299, pp. 396–450, 2016.

[25] N. Cook and A. Dembo, “Large deviations of subgraph counts for sparse Erdős–Rényi graphs,” *Advances in Mathematics*, vol. 373, p. 107 289, 2020.

[26] N. A Cook, A. Dembo, and H.-T. Pham, “Regularity method and large deviation principles for the Erdős–Rényi hypergraph,” *arXiv preprint arXiv:2102.09100*, 2021.

[27] E. Lubetzky and Y. Zhao, "On the variational problem for upper tails in sparse random graphs," *Random Structures & Algorithms*, vol. 50, no. 3, pp. 420–436, 2017.

[28] F. Augeri, "Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdös-Rényi graphs," *The Annals of Probability*, vol. 48, no. 5, pp. 2404–2448, 2020.

[29] A. Basak and S. Mukherjee, "Universality of the mean-field for the Potts model," *Probability Theory and Related Fields*, vol. 168, no. 3-4, pp. 557–600, 2017.

[30] N. Deb and S. Mukherjee, "Fluctuations in mean-field Ising models," *arXiv preprint arXiv:2005.00710*, 2020.

[31] R. Eldan, "Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations," *Geometric and Functional Analysis*, vol. 28, no. 6, pp. 1548–1596, 2018.

[32] ——, "Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation," *Probability Theory and Related Fields*, vol. 176, no. 3, pp. 737–755, 2020.

[33] R. Eldan and R. Gross, "Decomposition of mean-field Gibbs distributions into product measures," *Electronic Journal of Probability*, vol. 23, 2018.

[34] F. Augeri, "A transportation approach to the mean-field approximation," *Probability Theory and Related Fields*, vol. 180, no. 1, pp. 1–32, 2021.

[35] D. Lacker, S. Mukherjee, and L. C. Yeung, "Mean field approximation via log-concavity," *arXiv preprint arXiv:2206.01260*, 2022.

[36] N. Gozlan and C. Léonard, "Transport inequalities. A survey," *Markov Processes and Related Fields*, vol. 16, pp. 635–736, 2010.

[37] M. Ledoux, *The concentration of measure phenomenon*, 89. American Mathematical Soc., 2001.

[38] P.-M. Samson, "Concentration of measure principle and entropy-inequalities," in *Convexity and concentration*, Springer, 2017, pp. 55–105.

[39] L. Wu, "Feynman-Kac semigroups, ground state diffusions, and large deviations," *Journal of Functional Analysis*, vol. 123, no. 1, pp. 202–231, 1994.

[40] ——, "A deviation inequality for non-reversible Markov processes," in *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, Elsevier, vol. 36, 2000, pp. 435–445.

[41] L. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge University Press, 2000, vol. 1.

[42] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*. Walter de Gruyter, 2010, vol. 19.

[43] A. Guionnet and B. Zegarlinksi, "Lectures on logarithmic sobolev inequalities," in *Séminaire de probabilités XXXVI*, Springer, 2003, pp. 1–134.

[44] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Springer Science & Business Media, 2013, vol. 348.

[45] D. Lacker, "A non-exponential extension of Sanov's theorem via convex duality," *Advances in Applied Probability*, vol. 52, no. 1, pp. 61–101, 2020.

[46] M. Donsker and S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time, I," *Communications on Pure and Applied Mathematics*, vol. 28, no. 1, pp. 1–47, 1975.

[47] ——, "Asymptotic evaluation of certain Markov process expectations for large time, III," *Communications on Pure and Applied Mathematics*, vol. 29, no. 4, pp. 389–461, 1976.

[48] L. Wu, "Uniformly integrable operators and large deviations for Markov processes," *Journal of Functional Analysis*, vol. 172, no. 2, pp. 301–376, 2000.

[49] H. Djellout, A. Guillin, and L. Wu, "Transportation cost-information inequalities and applications to random dynamical systems and diffusions," *The Annals of Probability*, vol. 32, no. 3B, pp. 2702–2732, 2004.

[50] R. van Handel, "Probability in high dimension," *Lecture Notes (Princeton University)*, 2016.

[51] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, "Kantorovich duality for general transport costs and applications," *Journal of Functional Analysis*, vol. 273, no. 11, pp. 3327–3405, 2017.

[52] W. Rudin, *Functional analysis*. McGraw-Hill, New York, 1973.

[53] R. Dudley, *Real analysis and probability*. CRC Press, 2018.

[54] P. Dupuis and R. Ellis, *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011, vol. 902.

[55] C. Villani, *Topics in optimal transportation*, 58. American Mathematical Soc., 2003.

[56] S. Eckstein, “Extended Laplace principle for empirical measures of a Markov chain,” *Advances in Applied Probability*, vol. 51, no. 1, pp. 136–167, 2019.

[57] D. Lacker, “Liquidity, risk measures, and concentration of measure,” *Mathematics of Operations Research*, vol. 43, no. 3, pp. 813–837, 2018.

[58] J. Carrillo, A. Jüngel, P. Markowich, G. Toscani, and A. Unterreiter, “Entropy Dissipation Methods for Degenerate Parabolic Problems and Generalized Sobolev Inequalities,” *Monatsh. Math.*, vol. 133, no. 1, pp. 1–82, 2001.

[59] F. Malrieu, “Convergence to equilibrium for granular media equations and their Euler schemes,” *Ann. Appl. Probab.*, vol. 13, no. 2, pp. 540–560, 2003.

[60] F. Bolley, I. Gentil, and A. Guillin, “Uniform Convergence to Equilibrium for Granular Media,” *Arch. Ration. Mech. Anal.*, vol. 208, no. 2, pp. 429–445, 2013.

[61] J. Tugaut, “Convergence to the equilibria for self-stabilizing processes in double-well landscape,” *Ann. Probab.*, vol. 41, no. 3A, pp. 1427–1460, 2013.

[62] P. Cattiaux and G. A., “Semi Log-Concave Markov Diffusions,” in *Sémin. Probab. XLVI*, ser. Lecture Notes in Math. C. Donati-Martin, A. Lejay, and A. Rouault, Eds., vol. 2123, Berlin, Heidelberg: Springer, 2014, pp. 231–292.

[63] J. Carrillo, R. Gvalani, G. Pavliotis, and A. Schlichting, “Long-Time Behaviour and Phase Transitions for the McKean–Vlasov Equation on the Torus,” *Arch. Ration. Mech. Anal.*, vol. 235, no. 1, pp. 635–690, 2020.

[64] S. Mei, A. Montanari, and P.-M. Nguyen, “A mean field view of the landscape of two-layer neural networks,” *Proc. Natl. Acad. Sci. USA*, vol. 115, no. 33, E7665–E7671, 2018.

[65] K. Hu, Z. Ren, D. Šiška, and Ł. Szpruch, “Mean-field Langevin dynamics and energy landscape of neural networks,” *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 57, no. 4, pp. 2043–2065, 2021.

[66] J. Fontbona and B. Jourdain, “A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations,” *Ann. Probab.*, vol. 44, no. 1, pp. 131–170, 2016.

[67] C. Gardiner, *Stochastic Methods — A Handbook for the Natural and Social Sciences*, fourth, ser. Springer Ser. Synergetics. Berlin, Heidelberg: Springer, 2009, vol. 13.

[68] R. Jordan and D. Kinderlehrer, “An extended variational principle,” in *Partial Differential Equations and Applications: Collected Papers in Honor of Carlo Pucci*, ser. Lect. Notes Pure Appl. Math. Vol. 177, CRC Press, 1996, ch. 18, pp. 187–200.

[69] H. Risken, *The Fokker–Planck Equation — Methods of Solution and Applications*, second, ser. Springer Ser. Synergetics. Berlin, Heidelberg: Springer, 1996, vol. 18.

[70] S. Benachour, B. Roynette, D. Talay, and P. Vallois, "Nonlinear self-stabilizing processes — I Existence, invariant probability, propagation of chaos," *Stochastic Process. Appl.*, vol. 75, no. 2, pp. 173–201, 1998.

[71] A. Veretennikov, "On Ergodic Measures for McKean–Vlasov Stochastic Equations," in *Monte Carlo and Quasi-Monte Carlo Methods 2004*, H. Niederreiter and D. Talay, Eds., Berlin, Heidelberg: Springer, 2006, pp. 471–486.

[72] S. Herrmann and J. Tugaut, "Non-uniqueness of stationary measures for self-stabilizing processes," *Stochastic Process. Appl.*, vol. 120, no. 7, pp. 1215–1246, 2010.

[73] J. Tugaut, "Phase transitions of McKean–Vlasov processes in double-wells landscape," *Stochastics*, vol. 86, no. 2, pp. 257–284, 2014.

[74] ——, "Self-stabilizing Processes in Multi-wells Landscape in $\mathbb{R}^d$-Invariant Probabilities," *J. Theoret. Probab.*, vol. 27, no. 1, pp. 57–79, 2014.

[75] M. Duong and J. Tugaut, "Stationary solutions of the Vlasov–Fokker–Planck equation: Existence, characterization and phase-transition," *Appl. Math. Lett.*, vol. 52, pp. 38–45, 2016.

[76] ——, "The Vlasov–Fokker–Planck equation in non-convex landscapes: Convergence to equilibrium," *Electron. Commun. Probab.*, vol. 23, pp. 1–10, 2018.

[77] P.-L. Lions, "Théorie des jeux de champ moyen et applications," *Lectures at the Collège de France.* https://www.college-de-france.fr/site/pierre-louis-lions/course-2007-2008.htm, 2007–2008.

[78] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I — Mean Field FBSDEs, Control, and Games*, ser. Probab. Theory Stoch. Model. Switzerland: Springer International Publishing, 2018, vol. 83.

[79] D. Benedetto, E. Caglioti, J. Carrillo, and M. Pulvirenti, "A Non-Maxwellian Steady Distribution for One-Dimensional Granular Media," *J. Stat. Phys.*, vol. 91, no. 5–6, pp. 979–990, 1998.

[80] C. Villani, "Mathematics of Granular Materials," *J. Stat. Phys.*, vol. 124, no. 2–4, pp. 781–822, 2006.

[81] P. Cattiaux, A. Guillin, and F. Malrieu, "Probabilistic approach for granular media equations in the non-uniformly convex case," *Probab. Theory Related Fields*, vol. 140, no. 1–2, pp. 19–40, 2008.

[82] F. Otto, "The geometry of dissipative evolution equations: The porous medium equation," *Comm. Partial Differential Equations*, vol. 26, no. 1–2, pp. 101–174, 2001.

[83] I. Karatzas, J. Maas, and W. Schachermayer, "Trajectorial dissipation and gradient flow for the relative entropy in Markov chains," *Commun. Inf. Syst.*, vol. 21, no. 4, pp. 481–536, 2021.

[84] M. Erbar, M. Fathi, V. Laschos, and A. Schlichting, "Gradient flow structure for McKean–Vlasov equations on discrete spaces," *Discrete Contin. Dyn. Syst.*, vol. 36, no. 12, pp. 6799–6833, 2016.

[85] D. Cordero-Erausquin, W. Gangbo, and C. Houdré, "Inequalities for generalized entropy and optimal transportation," in *Recent Advances in the Theory and Applications of Mass Transport*, ser. Contemp. Math. J. Carvalho M.C. Rodrigues, Ed., vol. 353, Providence, RI: Amer. Math. Soc., 2004, pp. 73–94.

[86] J. Feng and T. Kurtz, *Large Deviations for Stochastic Processes*, ser. Math. Surveys Monogr. Providence, RI: Amer. Math. Soc., 2006, vol. 131.

[87] B. Acciaio, M. Beiglböck, F. Penkner, W. Schachermayer, and J. Temme, "A trajectorial interpretation of Doob's martingale inequalities," *Ann. Appl. Probab.*, vol. 23, no. 4, pp. 1494–1505, 2013.

[88] M. Beiglböck and P. Siorpaes, "Pathwise versions of the Burkholder–Davis–Gundy inequality," *Bernoulli*, vol. 21, no. 1, pp. 360–373, 2015.

[89] P. Cattiaux, "A Pathwise Approach of Some Classical Inequalities," *Potential Anal.*, vol. 20, no. 4, pp. 361–394, 204.

[90] J. Lehec, "Representation formula for the entropy and functional inequalities," *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 49, no. 3, pp. 885–899, 2013.

[91] J. Backhoff, G. Conforti, I. Gentil, and C. Léonard, "The mean field Schrödinger problem: Ergodic behavior, entropy estimates and functional inequalities," *Probab. Theory Related Fields*, vol. 178, no. 1–2, pp. 475–530, 2020.

[92] I. Gentil, C. Léonard, L. Ripani, and L. Tamanini, "An entropic interpolation proof of the HWI inequality," *Stochastic Process. Appl.*, vol. 130, no. 2, pp. 907–923, 2020.

[93] R. Eldan, J. Lehec, and Y. Shenfeld, "Stability of the logarithmic Sobolev inequality via the Föllmer process," *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 56, no. 3, pp. 2253–2269, 2020.

[94] R. Eldan and D. Mikulincer, "Stability of the Shannon–Stam inequality via the Föllmer process," *Probab. Theory Related Fields*, vol. 177, no. 3–4, pp. 891–922, 2020.

[95] I. Karatzas, W. Schachermayer, and B. Tschiderer, "A trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions," *Theory of Probability and Its Applications*, vol. 66, no. 4, pp. 668–707, 2022.

[96] C. Léonard, "Some Properties of Path Measures," in *Sémin. Probab. XLVI*, ser. Lecture Notes in Math. C. Donati-Martin, A. Lejay, and A. Rouault, Eds., vol. 2123, Switzerland: Springer International Publishing, 2014, pp. 207–230.

[97] Y. Brenier, "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions," *Comm. Pure Appl. Math.*, vol. 44, no. 4, pp. 375–417, 1991.

[98] R. McCann, "A Convexity Principle for Interacting Gases," *Adv. Math.*, vol. 128, no. 1, pp. 153–179, 1997.

[99] C. Villani, *Topics in Optimal Transportation*, ser. Grad. Stud. Math. Providence, RI: Amer. Math. Soc., 2003, vol. 58.

[100] U. Haussmann and É. Pardoux, "Time Reversal of Diffusions," *Ann. Probab.*, vol. 14, no. 4, pp. 1188–1205, 1986.

[101] R. Liptser and A. Shiryaev, *Statistics of Random Processes — I. General Theory*, second, ser. Stoch. Model. Appl. Probab. Berlin, Heidelberg: Springer, 2001, vol. 5.

[102] H. Föllmer, "An entropy approach to the time reversal of diffusion processes," in *Stochastic Differential Systems — Filtering and Control*, ser. Lect. Notes Control Inf. Sci. M. Métivier and É. Pardoux, Eds., vol. 69, Berlin, Heidelberg: Springer, 1985, pp. 156–163.

[103] ——, "Time reversal on Wiener space," in *Stochastic Processes — Mathematics and Physics*, ser. Lecture Notes in Math. S. Albeverio, P. Blanchard, and L. Streit, Eds., vol. 1158, Berlin, Heidelberg: Springer, 1986, pp. 119–129.

[104] ——, "Random fields and diffusion processes," in *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, ser. Lecture Notes in Math. P.-L. Hennequin, Ed., vol. 1362, Berlin, Heidelberg: Springer, 1988, pp. 101–203.

[105] C. Léonard, "A survey of the Schrödinger problem and some of its connections with optimal transport," *Discrete Contin. Dyn. Syst.*, vol. 34, no. 4, pp. 1533–1574, 2014.

[106] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, third, ser. Grundlehren Math. Wiss. Berlin, Heidelberg: Springer, 1999, vol. 293.

[107] L. Rogers, "Smooth Transition Densities for OneDimensional Diffusion," *Bull. Lond. Math. Soc.*, vol. 17, no. 2, pp. 157–161, 1985.

[108] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani, and A. Unterreiter, "Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities," *Monatshefte für Mathematik*, vol. 133, no. 1, pp. 1–82, 2001.

[109] J. L. Vázquez, *The Porous Medium Equation: Mathematical Theory*. Oxford University Press, 2007.

[110] J. A. Carrillo and G. Toscani, "*Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity*," *Indiana University Mathematics Journal*, pp. 113–142, 2000.

[111] K. Sato and H. Tanaka, "Local times on the boundary for multi-dimensional reflecting diffusion," *Proceedings of the Japan Academy*, vol. 38, no. 10, pp. 699–702, 1962.

[112] H. Wilhelm Alt and S. Luckhaus, "Quasilinear elliptic-parabolic differential equations," *Mathematische Zeitschrift*, vol. 183, no. 3, pp. 311–341, 1983.

[113] M. Erbar and J. Maas, "Gradient flow structures for discrete porous medium equations," *Discrete and Continuous Dynamical Systems*, vol. 34, no. 4, pp. 1355–1374, 2014.

[114] S. Lisini, E. Mainini, and A. Segatti, "A gradient flow approach to the porous medium equation with fractional pressure," *Archive for Rational Mechanics and Analysis*, vol. 227, no. 2, pp. 567–606, 2018.

[115] A. Pilipenko, *An introduction to stochastic differential equations with reflection*. Universitätsverlag Potsdam, 2014, vol. 1.

[116] H. M. Soner, "Stochastic representations for nonlinear parabolic PDEs," in *Handbook of Differential Equations: Evolutionary Equations*, vol. 3, Elsevier, 2007, pp. 477–526.

[117] S. Benachour, P. Chassaing, B. Roynette, and P. Vallois, "Processus associés à l'équation des milieux poreux," *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, vol. 23, no. 4, pp. 793–832, 1996.

[118] V. Barbu, M. Röckner, and F. Russo, "Probabilistic representation for solutions of an irregular porous media type equation: The degenerate case," *Probability Theory and Related Fields*, vol. 151, no. 1, pp. 1–43, 2011.

[119] P. Blanchard, M. Röckner, and F. Russo, "Probabilistic representation for solutions of an irregular porous media type equation," *The Annals of Probability*, vol. 38, no. 5, pp. 1870–1900, 2010.

[120] I. Ciotir and F. Russo, "Probabilistic representation for solutions of a porous media type equation with Neumann boundary condition: The case of the half-line," *Differential and Integral Equations*, vol. 27, no. 1/2, pp. 181–200, 2014.

[121] Y. Hu, Z. Qian, and Z. Zhang, "Gradient estimates for porous medium and fast diffusion equations by martingale method," in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Institut Henri Poincaré, vol. 53, 2017, pp. 1793–1820.

[122] P.-L. Lions and A.-S. Sznitman, "Stochastic differential equations with reflecting boundary conditions," *Communications on Pure and Applied Mathematics*, vol. 37, no. 4, pp. 511–537, 1984.

[123] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, ser. Grad. Texts in Math. New York: Springer, 1988, vol. 113.

[124] S. Itô, "A boundary value problem of partial differential equations of parabolic type," *Duke Mathematical Journal*, vol. 24, no. 3, pp. 299–312, 1957.

[125] V. Jain, F. Koehler, and A. Risteski, "Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: A unified perspective," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 1226–1236.

[126] T. Austin, "The structure of low-complexity Gibbs measures on product spaces," *The Annals of Probability*, vol. 47, no. 6, pp. 4002–4023, 2019.

[127] J. Yan, "Nonlinear large deviations: Beyond the hypercube," *The Annals of Applied Probability*, vol. 30, no. 2, pp. 812–846, 2020.

[128] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, pp. 146–158, 1975.

[129] R. McCann, "A convexity principle for interacting gases," *Advances in mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[130] D. Bakry and M. Émery, "Diffusions hypercontractives," in *Séminaire de Probabilités XIX 1983/84*, Springer, 1985, pp. 177–206.

[131] F. Otto and C. Villani, "Generalization of an inequality by talagrand and links with the logarithmic sobolev inequality," *Journal of Functional Analysis*, vol. 173, no. 2, pp. 361–400, 2000.

[132] D. Chafaï, N. Gozlan, and P.-A. Zitt, "First-order global asymptotics for confined particles with singular pair repulsion," *The Annals of Applied Probability*, vol. 24, no. 6, pp. 2371–2413, 2014.

[133] P. Dupuis, V. Laschos, and K. Ramanan, "Large deviations for configurations generated by gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials," *Electronic Journal of Probability*, vol. 25, pp. 1–41, 2020.

[134] M. Biskup and L. Chayes, “Rigorous analysis of discontinuous phase transitions via mean-field bounds,” *Communications in mathematical physics*, vol. 238, no. 1, pp. 53–93, 2003.

[135] A. Dembo and A. Montanari, “Ising models on locally tree-like graphs,” *The Annals of Applied Probability*, vol. 20, no. 2, pp. 565–592, 2010.

[136] A. Dembo, A. Montanari, A. Sly, and N. Sun, “The replica symmetric solution for potts models on d-regular graphs,” *Communications in Mathematical Physics*, vol. 327, no. 2, pp. 551–575, 2014.

[137] S. Delattre, G. Giacomin, and E. Luçon, “A note on dynamical models on random graphs and fokker–planck equations,” *Journal of Statistical Physics*, vol. 165, no. 4, pp. 785–798, 2016.

[138] R. Oliveira and G. Reis, “Interacting diffusions on random graphs with diverging average degrees: Hydrodynamics and large deviations,” *Journal of Statistical Physics*, vol. 176, no. 5, pp. 1057–1087, 2019.

[139] A. Dembo and A. M., “Gibbs measures and phase transitions on sparse random graphs,” *Brazilian Journal of Probability and Statistics*, vol. 24, no. 2, pp. 137–211, 2010.

[140] D. Lacker, K. Ramanan, and R. Wu, “Local weak convergence for sparse networks of interacting processes,” *The Annals of Applied Probability*, vol. 33, no. 2, pp. 643–688, 2023.

[141] C. Borgs, J. Chayes, H. Cohn, and Y. Zhao, “An L^p theory of sparse graph convergence i: Limits, sparse random graph models, and power law distributions,” *Transactions of the American Mathematical Society*, vol. 372, no. 5, pp. 3019–3062, 2019.

[142] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao, “An L^p theory of sparse graph convergence ii: Ld convergence, quotients and right convergence,” *The Annals of Probability*, vol. 46, no. 1, pp. 337–396, 2018.

[143] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi, “Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing,” *Advances in Mathematics*, vol. 219, no. 6, pp. 1801–1851, 2008.

[144] L. Lovász, *Large networks and graph limits*. American Mathematical Soc., 2012, vol. 60.

[145] P. Alquier, J. Ridgway, and N. Chopin, “On the properties of variational approximations of gibbs posteriors,” *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 8374–8414, 2016.

[146] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[147] K. Ray and B. Szabó, "Variational bayes for high-dimensional linear regression with sparse priors," *Journal of the American Statistical Association*, pp. 1–12, 2021.

[148] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[149] A. Y. Zhang and H. H. Zhou, "Theoretical and computational guarantees of mean field variational inference for community detection," *The Annals of Statistics*, vol. 48, no. 5, pp. 2575–2598, 2020.

[150] S. Mukherjee and S. Sen, "Variational inference in high-dimensional linear regression," *Journal of Machine Learning Research*, vol. 23, no. 304, pp. 1–56, 2022.

[151] R. Carmona and F. Delarue, "Forward–backward stochastic differential equations and controlled mckean–vlasov dynamics," *The Annals of Probability*, vol. 43, no. 5, pp. 2647–2700, 2015.

[152] M. Huang, R. Malhamé, and P. Caines, "Large population stochastic dynamic games: Closed-loop mckean-vlasov systems and the nash certainty equivalence principle," *Communications in Information & Systems*, vol. 6, no. 3, pp. 221–252, 2006.

[153] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[154] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2018.

[155] M. Boué and P. Dupuis, "A variational representation for certain functionals of Brownian motion," *The Annals of Probability*, vol. 26, no. 4, pp. 1641–1659, 1998.

[156] C. Borell, "Diffusion equations and geometric inequalities," *Potential Analysis*, vol. 12, no. 1, pp. 49–71, 2000.

[157] H. Föllmer, "An entropy approach to the time reversal of diffusion processes," in *Stochastic Differential Systems Filtering and Control*, Springer, 1985, pp. 156–163.

[158] ——, "Time reversal on Wiener space," in *Stochastic processesmathematics and physics*, Springer, 1986, pp. 119–129.

[159] H. Brascamp and E. Lieb, “On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation,” in *Inequalities*, Springer, 2002, pp. 441–464.

[160] S. Bobkov and M. Madiman, “The entropy per coordinate of a random vector is highly constrained under convexity conditions,” *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4940–4954, 2011.

[161] A. Saumard and J. Wellner, “Log-concavity and strong log-concavity: A review,” *Statistics surveys*, vol. 8, p. 45, 2014.

[162] F. Baudoin, “Conditioned stochastic differential equations: Theory, examples and application to finance,” *Stochastic Processes and their Applications*, vol. 100, no. 1-2, pp. 109–145, 2002.

[163] J. Lehec, “Representation formula for the entropy and functional inequalities,” in *Annales de l’IHP Probabilités et statistiques*, vol. 49, 2013, pp. 885–899.