universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

A probabilistic view of Wasserstein gradient flows and the Benamou–Brenier problem

verfasst von / submitted by

Bertram Thomas Tschiderer, BSc MSc

angestrebter akademischer Grad /
in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien / Vienna, 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 796 605 405
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Mathematik
Betreut von / Supervisors:	emer. o. Univ.-Prof. Mag. Dr. Walter Schachermayer
	Univ.-Prof. Dipl.-Ing. Dr.techn. Mathias Beiglböck

Abstrakt

Diffusionsprozesse, die als Gradientenflüsse bestimmter Energiefunktionale bezüglich quadratischer Wasserstein-Metriken interpretierbar sind, werden in der Fachliteratur vielfach untersucht und finden in zahlreichen unterschiedlichen Teilgebieten der Mathematik Anwendung. Im ersten Teil dieser Arbeit betrachten wir Gradientenflüsse und verwandte klassische Themen wie Dissipation der Entropie, variationelle Charakterisierungen und Funktionalungleichungen aus einer neuen Perspektive. Für verschiedene Diffusionsmodelle schaffen wir einen probabilistischen Rahmen, um pfadweise Interpretationen dieser Thematiken zu ermöglichen. Unser Ansatz beruht auf Methoden der stochastischen Analysis sowie Zeitumkehr. Das Vorliegen eines Gradientenflusses wird durch eine Störungsanalyse gezeigt, wobei entweder der Drift- oder der Volatilitätskoeffizient der zugrundeliegenden Diffusion als Kontrollvariable verwendet wird.

In der klassischen Theorie des optimalen Transports bilden die Arbeiten von Benamou–Brenier und McCann zum optimalen Transportproblem in stetiger Zeit die Grundlage dieses Fachgebiets und ermöglichen eine Vielzahl von Anwendungen in anderen mathematischen Bereichen. Das Konzept der "stretched Brownian motion" bietet ein Analogon für die Martingal-Version dieses Problems. Der zweite Teil dieser Arbeit beinhaltet eine Charakterisierung von Optimierern dieses Problems in Form von Gradienten konvexer Funktionen, ähnlich der Charakterisierung von Optimierern im klassischen Transportproblem für quadratische Kosten. Der verwendete Ansatz basiert auf Dualität sowie Argumenten der konvexen Analysis und wird durch eine variationelle Perspektive ergänzt.

Abstract

Various diffusion processes can be interpreted as gradient flows of certain energy functionals with respect to quadratic Wasserstein metrics. Such processes are well studied and have applications in many different fields of mathematics. The first goal of this work is to shed new light on Wasserstein gradient flows and related classical topics such as entropy dissipation, variational characterizations, and functional inequalities. For several models of diffusions, we set up a probabilistic framework in order to provide trajectorial interpretations of these topics. Our approach is based on stochastic analysis and time-reversal techniques. The gradient flow property is established via a perturbation analysis, where either the drift or volatility coefficient of the underlying diffusion is used as a control variable.

In classical optimal transport, the contributions of Benamou–Brenier and McCann regarding the time-dependent version of the problem are cornerstones of the field and form the basis for a variety of applications in other mathematical areas. Stretched Brownian motion provides an analogue for the martingale version of this problem. The second goal of this work is to provide a characterization of optimizers in terms of gradients of convex functions, similar to the characterization of optimizers in the classical transport problem for quadratic distance cost. This is achieved by means of a convex duality approach, complemented with a purely variational perspective.

Acknowledgements

First and foremost, I sincerely thank my supervisors Walter Schachermayer and Mathias Beiglböck for their great assistance during the process of writing this thesis.

Thank you Walter for your support, energy, motivation and positivity. I really cannot imagine a supervisor who devoted more time to their PhD student than you have. An email response or callback after more than 24 hours was a seldom exception and almost a reason for concern. I also will never forget our exciting and unique time we had together in New York when we visited Ioannis. Thank you for this experience and opportunity.

Thank you Mathias, you did much more for me than supervisors typically are supposed to do. Whenever it got mathematically serious, you were always ready. On top of that, your support of my climbing career and fatherly care concerning my climbing injuries will remain unmatched forever. I already look forward to our future MoonBoard sessions.

No less thanks are due to Julio Backhoff-Veraguas, who I can undoubtedly call my unofficial co-supervisor. Thank you Julio, you were like a big brother to me. You always assisted me with your advice and offered quick help with all kinds of mathematical problems.

I extend my sincere thanks to Ioannis Karatzas. Thank you Ioannis, for being such a good friend, mentor and co-author. I was able to learn so much from you and your continuous support was more than generous.

Many thanks go to Lane Chun Yeung for our wonderful time in New York and our successful collaboration.

I am grateful to Ben Robinson for reading my thesis and for his valuable feedback and comments.

I express my sincere thanks to Astrid Kollros-Spinka for her superb assistance whenever I had administrative or organizational issues.

I am also grateful to all friends and colleagues who have accompanied me on this journey.

Last but not least, I express my deepest gratitude to my family for their generous and never-ending support.

During my work as a doctoral student, I greatly appreciated the support by the Austrian Science Fund (FWF) through projects P28661, P35197, P35519 and by the Vienna Science and Technology Fund (WWTF) through grant MA16-021.

Contents

Preamble 1
A probabilistic view of Wasserstein gradient flows . . . 1
A structural description of the martingale Benamou–Brenier problem . . . 5

Synopsis of the publications and manuscripts 10
Indication of the submitter's own contribution . . . 12

I. A trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions 14

1. Introduction 15
1.1. Preview . . . 16

2. The stochastic approach 18
2.1. Regularity assumptions . . . 19

3. The main theorems in aggregate form 21

4. The main theorems in trajectorial form 25
4.1. Consequences of the trajectorial results . . . 27
4.2. A trajectorial proof of the HWI inequality . . . 34

5. Details and proofs of the main results 38
5.1. The proof of Theorem 4.1 . . . 38
5.2. The proof of Theorem 4.2 . . . 42
5.3. Some useful lemmas . . . 45
5.4. Completing the proof of Proposition 4.7 . . . 49
5.5. The rate of growth for the Wasserstein distance . . . 51

6. Proofs of auxiliary results 58
6.1. Relative entropy with respect to the σ-finite reference measure Q . . . 58
6.2. The proofs of Lemmas 2.3 and 3.2 . . . 59
6.3. Some results in probability and martingale theory . . . 60
6.4. The proof of Lemma 4.10 . . . 62

II. A variational characterization of Langevin–Smoluchowski diffusions 65

1. Introduction 66
1.1. Preview . . . 67

2. The setting 68
2.1. Invariant measure, likelihood ratio, and relative entropy . . . 68
2.2. The probabilistic setting . . . 70
2.3. Reversal of time . . . 71

3. A stochastic control problem 72
3.1. Entropic interpretation of the expected cost when $Q(\mathbb{R}^n) < \infty$ . . . 76

4. From local to square-integrable martingales 77
4.1. Reversing time once again . . . 78
4.2. The dynamics of the relative entropy process . . . 79
4.3. Relative entropy dissipation . . . 79

5. From backwards dynamics "back" to forward dynamics 81
5.1. Entropic interpretation of the expected cost when $Q(\mathbb{R}^n) < \infty$ . . . 84

6. The case of finite invariant measure, and an iterative procedure 86
6.1. The decrease of the relative entropy without convexity assumption . . . 87
6.2. The triviality of the tail $\boldsymbol{\sigma}$-algebra $\mathcal{H}(\infty)$ . . . 89

III. A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: gradient flows and HWBI inequalities 92

1. Introduction 93
1.1. Gradient flow structure of the granular media equation . . . 95
1.2. The HWBI inequality . . . 96
1.3. Preview . . . 98

2. The probabilistic framework 99
2.1. The setting . . . 99
2.2. Regularity assumptions . . . 99
2.3. Probabilistic representations of gradient flow functionals . . . 100

3. Main results 102
3.1. Trajectorial dissipation of relative entropy for McKean–Vlasov diffusions . 102
3.1.1. Examples . . . 103
3.1.2. Consequences of Theorem 3.1 . . . 106

3.2. Gradient flow structure of the granular media equation 107
3.3. A trajectorial proof of the HWBI inequality 112

4. Proofs of the main results 114
4.1. The proofs of Theorem 3.1 and Theorem 3.7 114
4.2. The proofs of Proposition 3.14 and Theorem 3.16 122

5. Proofs of auxiliary results 125
5.1. The proof of Lemma 2.2 125
5.2. The proof of (4.45) 126

IV. Diffusion processes as Wasserstein gradient flows via stochastic control of the volatility matrix 129

1. Introduction 130
1.1. Invariant measure 130
1.2. Relative entropy and free energy 131
1.3. Riemannian structure 132
1.4. Relative Fisher information 132
1.5. Wasserstein distance 133
1.6. Class of admissible volatilities 133
1.7. Main results 134
1.8. Literature review 136

2. Dynamics of the relative entropy process 138
2.1. The proof of Theorem 2.1 140
2.2. Consequences of Theorem 2.1 143

3. Proof of the gradient flow property 145
3.1. Derivative of the Wasserstein distance 145
3.2. The proof of Theorem 1.2 146

V. Time reversal of diffusions 148

1. Introduction 149

2. The setting 149

3. Time reversal and the backwards filtration 151

4. Some remarkable Brownian motions 152

5. The diffusion property under time reversal 155

6. The backwards Itô integral 157

VI. The structure of martingale Benamou–Brenier in $\mathbb{R}^d$ 160

1. Introduction 161
1.1. Benamou–Brenier transport problem and McCann interpolation in probabilistic terms . . . 161
1.2. Martingale optimization problem . . . 162
1.3. The Bass martingale . . . 163
1.4. Structure of stretched Brownian motion . . . 163
1.5. Dual viewpoint . . . 165
1.6. Literature . . . 166
1.7. Preview . . . 167

2. Definitions and Notation 168

3. Duality for stretched Brownian motion 170

4. Proof of the first part of Theorem 1.7 175

5. Preparation for the proof of the second part of Theorem 1.7 181

6. Existence of dual optimizers and Bass martingales 187

7. Irreducibility and existence of dual optimizers 196
7.1. Proof of Theorem 7.6 . . . 197
7.1.1. Step 1 of the proof of Theorem 7.6 . . . 197
7.1.2. Step 2 of the proof of Theorem 7.6 . . . 200
7.2. Proof of Theorem 1.3 . . . 209
7.3. Proof of the second part of Theorem 1.7 . . . 209

VII. The Bass functional of martingale transport 212

1. Introduction 213
1.1. Martingale optimization problem . . . 213
1.2. Bass martingales and structure of stretched Brownian motion . . . 213
1.3. Main results . . . 215
1.4. Related literature . . . 216

2. Preliminaries 217
2.1. Dual viewpoint . . . 217
2.2. Static martingale optimal transport . . . 217
2.3. Structure of optimizers . . . 218

3. A variational characterization of Bass measures 220

4. An infinitesimal version of Theorem 1.5 226

5. Displacement convexity of the Bass functional 229

VIII. q-Bass martingales 233

1. Introduction 234
1.1. Martingale optimization problem 234
1.2. Dual formulation 235
1.3. q-Bass martingales 235
1.4. Related literature 238

2. Proof of the duality result 239
2.1. A first dual formulation 239
2.2. Convexity of the dual variables 240
2.3. Proof of Theorem 1.1 243

3. Existence and properties of q-Bass martingales 250
3.1. Proof of Theorem 1.5 252

Bibliography 256

Concluding discussion 267

Preamble

A probabilistic view of Wasserstein gradient flows

Diffusion processes that can be interpreted as Wasserstein gradient flows are well studied objects with applications in many different fields of mathematics. The first goal of this dissertation is to shed new light on this and related classical topics such as entropy dissipation, variational characterizations, and functional inequalities. For various models of diffusions, we set up a probabilistic framework in order to provide trajectorial interpretations of these topics. Our approach is based on stochastic analysis and time-reversal techniques. The gradient flow property is established via a perturbation analysis, where either the drift or volatility coefficient of the underlying diffusion is used as a control variable.

In Part I we revisit the variational characterization of conservative diffusion as entropic gradient flow. We provide a probabilistic interpretation via a perturbation analysis based on stochastic calculus. It was shown by Jordan, Kinderlehrer and Otto in the seminal work [JKO98] that, for diffusions of Langevin–Smoluchowski type, the Fokker–Planck probability density flow maximizes the rate of relative entropy dissipation with respect to the quadratic Wasserstein metric. Inspired by the work of Fontbona and Jourdain [FJ16] we obtain novel, stochastic-process versions of these features, valid along the trajectories of the diffusive motion in the backward direction of time, using a direct perturbation methodology. By averaging our trajectorial results with respect to the underlying measure on path space, we establish the maximal rate of entropy dissipation along the Fokker–Planck flow. We measure exactly the deviation from this maximum that corresponds to a given perturbation. As a further application of our trajectorial approach we derive the famous HWI inequality of Otto and Villani [OV00, Vil03, Vil09, CE02] relating relative entropy (H), Wasserstein distance (W) and relative Fisher information (I).

Part II builds directly on Part I and is formulated in the same setting. We show that the Langevin–Smoluchowski measure on path space is invariant under time-reversal followed by stochastic control of the drift with a novel entropic-type criterion. Repeated application of these forward-backward steps leads to a sequence of stochastic control problems, whose initial/terminal distributions converge to the Gibbs probability measure of the diffusion, and whose values decrease to zero along the relative entropy of the Langevin–Smoluchowski flow with respect to this Gibbs measure.

In Part III we formulate a trajectorial version of the relative entropy dissipation identity for McKean–Vlasov diffusions. We thus extend the results of Part I, which apply

to non-interacting diffusions, to nonlinear diffusions with interaction. Our stochastic analysis approach is based on time-reversal of diffusions and Lions' differential calculus [Lio07, CD18] over Wasserstein space. This approach 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, which was originally established in [CMV03, CMV06]. In this respect, we generalize the gradient flow formulation developed in Part I for the linear Fokker–Planck equation to the nonlinear case. Second, we show how the trajectorial approach leads to a new derivation of the HWBI inequality [AGK04, CEGH04, FK06], which relates relative entropy (H), Wasserstein distance (W), barycenter (B) and Fisher information (I).

In Part IV we consider a class of time-homogeneous diffusion processes on $\mathbb{R}^n$ with common invariant measure but varying volatility matrices. In Euclidean space, we show that the corresponding flow of time-marginal distributions admits an entropic gradient flow formulation in the quadratic Wasserstein space if the volatility matrix of the diffusion is the identity. Our approach is based on stochastic control of the diffusion coefficient and time-reversal techniques. Equipping $\mathbb{R}^n$ with a Riemannian metric, we prove that the diffusion process can be viewed as a gradient flow in the induced Wasserstein space if the volatility matrix is the inverse of the underlying metric tensor. In the Euclidean case, our probabilistic result corresponds to the gradient flow characterization of the Fokker–Planck equation, first discovered in [JKO98]. In this context, we also recover the gradient flow interpretation as established in Part I, but instead of the drift coefficient we now use the diffusion coefficient as control variable. In the Riemannian setting, the corresponding result on the level of partial differential equations was established by Lisini in [Lis09], building on the metric theory developed by Ambrosio, Gigli, and Savaré in the book [AGS08]. This setting was also used by Fathi in [Fat16] to develop a gradient flow approach to large deviations for diffusion processes.

The reversal of time plays a major role throughout Parts I – IV of this thesis. For the sake of clarity, we review in Part V the theory of time reversal for diffusion processes developed by Föllmer [Föl85, Föl86], Haussmann and Pardoux [HP86], and Pardoux [Par86]. This part can be read independently of the rest of the thesis; it does not present novel results.

In the following we present a prototypical example to illustrate what we mean by "a probabilistic view of Wasserstein gradient flows". We consider the stochastic differential equation of Langevin–Smoluchowski type

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T, \tag{1}$$

where B is standard n-dimensional Brownian motion and $V\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function. We denote by P_t the time-marginal distributions of the random variables X_t and by p_t the corresponding probability density functions, which satisfy the Fokker–

Planck equation

$$\partial_t p_t(x) = \operatorname{div}\Big(\big(\nabla \log p_t(x) + \nabla V(x)\big)\, p_t(x)\Big). \tag{2}$$

The invariant measure Q of the diffusion process $(X_t)_{0\leqslant t\leqslant T}$ has density $q(x) \coloneqq \mathrm{e}^{-V(x)}$. The *relative entropy* and the *relative Fisher information* of P_t with respect to Q are given by

$$H(P_t \,|\, \mathrm{Q}) = \mathbb{E}\big[\log \ell_t(X_t)\big] \qquad \text{and} \qquad I(P_t \,|\, \mathrm{Q}) = \mathbb{E}\big[|\nabla \log \ell_t(X_t)|^2\big], \tag{3}$$

respectively, where the *likelihood ratio function* $\ell_t(\,\cdot\,)$ is defined as $\ell_t(x) \coloneqq \frac{p_t(x)}{q(x)}$.

It was first established in [JKO98] that the "flow of probability measures" $(P_t)_{0\leqslant t\leqslant T}$ is the *gradient flow* of the relative entropy functional

$$\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n) \ni P \longmapsto H(P \,|\, \mathrm{Q}) = \int_{\mathbb{R}^n} \log\left(\frac{p(x)}{q(x)}\right) p(x)\,\mathrm{d}x \tag{4}$$

with respect to the quadratic Wasserstein distance

$$W_2(\mu,\nu) = \sqrt{\inf_{Y\sim\mu, Z\sim\nu} \mathbb{E}\big[|Y-Z|^2\big]}, \qquad \mu,\nu \in \mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n),$$

where $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$ denotes the collection of absolutely continuous probability measures with finite second moments. In the language of analysis in metric spaces, the gradient flow property of the Langevin–Smoluchowski diffusion $(X_t)_{0\leqslant t\leqslant T}$ can be formulated as follows [AGS08].

Proposition 1 ([AGS08])**.** *For every $(\tilde{P}_t)_{0\leqslant t\leqslant T}$ in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$ and $0 \leqslant t_0 \leqslant t \leqslant T$ we have*

$$\big|H(\tilde{P}_t \,|\, \mathrm{Q}) - H(\tilde{P}_{t_0} \,|\, \mathrm{Q})\big| \leqslant \int_{t_0}^{t} \sqrt{I(\tilde{P}_u \,|\, \mathrm{Q})}\, \lim_{h\to 0} \frac{W_2(\tilde{P}_{u+h}, \tilde{P}_u)}{|h|}\,\mathrm{d}u. \tag{5}$$

In the case of the Langevin–Smoluchowski diffusion $(P_t)_{0\leqslant t\leqslant T}$, we have the identity

$$H(P_{t_0} \,|\, \mathrm{Q}) - H(P_t \,|\, \mathrm{Q}) = \tfrac{1}{2}\int_{t_0}^{t} I(P_u \,|\, \mathrm{Q})\,\mathrm{d}u + \tfrac{1}{2}\int_{t_0}^{t} \left(\lim_{h\to 0} \frac{W_2(P_{u+h}, P_u)}{|h|}\right)^2 \mathrm{d}u. \tag{6}$$

The inequality (5) has the interpretation that the functional

$$P \longmapsto \sqrt{I(P \,|\, \mathrm{Q})} = \sqrt{\int_{\mathbb{R}^n} \left|\nabla \log\left(\frac{p(x)}{q(x)}\right)\right|^2 p(x)\,\mathrm{d}x} \tag{7}$$

is a *strong upper gradient* for the relative entropy functional (4). In the case of the Langevin–Smoluchowski diffusion, we obtain the so-called *energy identity* (6). This shows that $(P_t)_{0\leqslant t\leqslant T}$ is a *curve of maximal slope* for the relative entropy functional (4) with respect to its strong upper gradient (7) in the quadratic Wasserstein space $(\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n), W_2)$. As *curves of maximal slope coincide with gradient flows*, this verifies the stated gradient flow property of $(P_t)_{0\leqslant t\leqslant T}$.

The novelty of our approach is that it is probabilistic and based on stochastic analysis. Instead of the Fokker–Planck equation (2), our focus lies on the associated stochastic differential equation (1). Most of Proposition 1, which is a deterministic result, will then be a consequence of a trajectorial result, Theorem 1 below. Trajectorial approaches of a similar kind were initiated by Fontbona and Jourdain in [FJ16].

Having the above motivation in mind, we model relative entropy as a stochastic process, the *relative entropy process*

$$\log \ell_t(X_t) = \log\left(\frac{p_t(X_t)}{q(X_t)}\right), \qquad 0 \leqslant t \leqslant T. \tag{8}$$

Our first goal is to understand the time evolution and dissipation of entropy on a trajectorial level. Since taking the expectation in (8) leads to the relative entropy as in (3), the idea is to compute the stochastic differential of the relative entropy process. Motivated by results from [DPP89, Pav89], and most notably the work [FJ16], we expect that this stochastic differential will be most transparent when regarded in the backward direction of time. Thus we consider the time-reversed process

$$\overline{X}_s \coloneqq X_{T-s}, \qquad 0 \leqslant s \leqslant T.$$

The following Theorem 1 describes the semimartingale dynamics of the time-reversed relative entropy process

$$\log \overline{\ell}_s(\overline{X}_s) = \log \ell_{T-s}(X_{T-s}), \qquad 0 \leqslant s \leqslant T, \tag{9}$$

in terms of the time-reversed *cumulative relative Fisher information process*

$$\overline{F}_s \coloneqq \int_0^s |\nabla \log \overline{\ell}_u(\overline{X}_u)|^2 \,\mathrm{d}u, \qquad 0 \leqslant s \leqslant T. \tag{10}$$

In particular, we will see that the difference between the processes of (9) and (10) is a martingale.

Theorem 1 ([KST22])**.** *We have*

$$\mathbb{E}\big[\,\overline{F}_T\,\big] = \int_0^T \mathbb{E}\Big[|\nabla \log \ell_t(X_t)|^2\Big]\,\mathrm{d}t = \int_0^T I(P_t \,|\, \mathrm{Q})\,\mathrm{d}t,$$

and the process

$$\overline{M}_s \coloneqq \big(\log \overline{\ell}_s(\overline{X}_s) - \log \ell_T(X_T)\big) - \overline{F}_s, \qquad 0 \leqslant s \leqslant T,$$

is a martingale bounded in L^2. Moreover, $\overline{M}$ has the stochastic integral representation

$$\overline{M}_s = \int_0^s \Big\langle \nabla \log \overline{\ell}_u(\overline{X}_u)\,, \sqrt{2}\,\mathrm{d}\overline{B}_u \Big\rangle, \qquad 0 \leqslant s \leqslant T,$$

where the process $(\overline{B}_s)_{0 \leqslant s \leqslant T}$ is Brownian motion in the backward direction of time.

Taking expectations, we obtain the following deterministic result.

Corollary 1 ([AGS08]). *For all $t_0, t \in [0,T]$ we have the relative entropy identity*

$$H(P_{t_0} \,|\, \mathrm{Q}) - H(P_t \,|\, \mathrm{Q}) = \int_{t_0}^{t} \mathbb{E}\Big[|\nabla \log \ell_u(X_u)|^2\Big] \,\mathrm{d}u = \int_{t_0}^{t} I(P_u \,|\, \mathrm{Q}) \,\mathrm{d}u.$$

Furthermore, for Lebesgue-a.e. $t \in [0,T]$, the dissipation of relative entropy is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} H(P_t \,|\, \mathrm{Q}) = \mathbb{E}\Big[|\nabla \log \ell_t(X_t)|^2\Big] = I(P_t \,|\, \mathrm{Q}). \tag{11}$$

The relation (11) between entropy and Fisher information takes the familiar form of a *de Bruijn type identity* [Sta59] and is also known as an *entropy production equality*. Various authors have studied this identity in different settings [BÉ85, OV00, AMTU01, CMV03, Vil03, AGS08].

A further consequence of Theorem 1 is the trajectorial analogue of (11).

Corollary 2 ([KST22]). *For Lebesgue-a.e. $t \in [0,T]$, the one-sided trajectorial rates of relative entropy dissipation are given by*

$$\lim_{s \uparrow T-t} \frac{\log \overline{\ell}_s(\overline{X}_s) - \mathbb{E}\big[\log \ell_t(X_t) \,|\, \overline{\mathcal{F}}_s\big]}{T-t-s} = -|\nabla \log \ell_t(X_t)|^2,$$

$$\lim_{s \downarrow T-t} \frac{\log \ell_t(X_t) - \mathbb{E}\big[\log \overline{\ell}_s(\overline{X}_s) \,|\, \overline{\mathcal{F}}_{T-t}\big]}{s-(T-t)} = -|\nabla \log \ell_t(X_t)|^2,$$

where both limits exist in L^1.

To derive the inequality (5), the main idea is to introduce "perturbations", where either the drift or volatility coefficient of the stochastic differential equation (1) is used as a control variable. This gives rise to time-marginal distributions $(\tilde{P}_t)_{0 \leqslant t \leqslant T}$, for which the inequality (5) holds.

A structural description of the martingale Benamou–Brenier problem

In classical optimal transport, the contributions of Benamou, Brenier, and McCann [Bre87, Bre91, McC94, BB99] regarding the time-dependent version of the problem are cornerstones of the field and form the basis for a variety of applications in other mathematical areas [Vil03, Vil09, AG13, San15]. The continuous-time optimization problem

$$MT(\mu,\nu) = \inf_{\substack{M_0 \sim \mu,\, M_1 \sim \nu, \\ M_t = M_0 + \int_0^t \sigma_s \, dB_s}} \mathbb{E}\Big[\int_0^1 |\sigma_t - I_d|_{\mathrm{HS}}^2 \, dt\Big] \tag{MBB}$$

can be seen as the martingale counterpart of the classical formulation

$$T_2(\mu,\nu) = \inf_{\substack{X_0\sim\mu,\, X_1\sim\nu,\\ X_t = X_0 + \int_0^t b_s\, ds}} \mathbb{E}\Big[\int_0^1 |b_t|^2\, dt\Big]. \tag{BB}$$

Here, μ and ν are probability measures on $\mathbb{R}^d$ in convex order and with finite second moments, B is Brownian motion on $\mathbb{R}^d$, and $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. The continuous-time optimization problem (BB) from classical optimal transport was introduced by Benamou–Brenier in [BB99] — hence the abbreviation (MBB) for "Martingale Benamou–Brenier".

The problem (MBB) is equivalent to maximizing the covariance between the martingale M and Brownian motion subject to the marginal conditions $M_0 \sim \mu$ and $M_1 \sim \nu$, i.e. finding

$$P(\mu,\nu) = \sup_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s\, dB_s}} \mathbb{E}\Big[\int_0^1 \mathrm{tr}(\sigma_t)\, dt\Big]. \tag{P}$$

As shown in [BBHK20], the problems (MBB) and (P) admit the same strong Markov martingale $\hat{M}$ as the unique optimizer. This optimizer is called the *stretched Brownian motion* from μ to ν.

In Part VI of this thesis we provide a characterization of the stretched Brownian motion $\hat{M}$ in terms of gradients of convex functions, similar to the characterization of optimizers in the classical transport problem for quadratic distance cost. More precisely, we show that under the necessary and sufficient condition that (μ,ν) is *irreducible*, a martingale $\hat{M}$ is the stretched Brownian motion from μ to ν if and only if it is a *Bass martingale*, i.e. a martingale of the form

$$\hat{M}_t = \mathbb{E}[\nabla\hat{v}(B_1)\,|\,\sigma(B_s\colon s\leqslant t)] = \mathbb{E}[\nabla\hat{v}(B_1)\,|\,B_t], \qquad 0\leqslant t\leqslant 1. \tag{Bass}$$

Here, $\hat{v}\colon \mathbb{R}^d\to\mathbb{R}$ is a convex function and B is Brownian motion on $\mathbb{R}^d$ with possibly non-degenerate starting law $B_0\sim\hat{\alpha}$, which is referred to as the *Bass measure*. The irreducibility condition of the pair (μ,ν) states that, for all measurable sets $A,B\subseteq\mathbb{R}^d$ with $\mu(A)>0$ and $\nu(B)>0$, there exists a martingale $X=(X_t)_{0\leqslant t\leqslant 1}$ with $X_0\sim\mu$ and $X_1\sim\nu$ such that $\mathbb{P}(X_0\in A, X_1\in B)>0$.

We write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and $\hat{v}_t := \hat{v} * \gamma^{1-t}\colon \mathbb{R}^d\to\mathbb{R}$ for the convolution of the function $\hat{v}$ and the measure γ^{1-t}. In these terms, (Bass) amounts to

$$\hat{M}_t = \nabla\hat{v}_t(B_t), \qquad 0\leqslant t\leqslant 1,$$

which emphasizes that the Bass martingale is obtained as a monotone transformation of Brownian motion at each time point.

We also establish in Part VI that the value $P(\mu,\nu)$ of the primal problem (P) is equal to the value of the dual problem

$$D(\mu,\nu) = \inf_{\substack{\psi\in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big), \tag{D}$$

where γ is the standard Gaussian measure on $\mathbb{R}^d$, the symbol $*$ used as a superscript denotes the convex conjugate of a function, and

$$(\psi^* * \gamma)(y) = \int \psi^*(y-z)\,\gamma(dz), \qquad y \in \mathbb{R}^d$$

is the standard convolution of a function with a measure. Furthermore, we show in Part VI that the right-hand side of (D) is attained by a convex function $\hat{\psi}$ if and only if (μ,ν) is irreducible. In this case, the stretched Brownian motion from μ to ν is a Bass martingale induced by the convex function $\hat{v} = \hat{\psi}^*$ and with Bass measure $\hat{\alpha} = \nabla(\hat{v} * \gamma)^*(\mu)$.

In Part VII we introduce the *Bass functional*

$$\mathscr{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu),$$

where $\mathscr{P}_2(\mathbb{R}^d)$ is the collection of probability measures on $\mathbb{R}^d$ with finite second moments and $\mathrm{MCov}(\,\cdot\,,\,\cdot\,)$ denotes the *maximal covariance* between two probability measures, defined as

$$\mathrm{MCov}(p,\varrho) := \sup_{\tilde{\pi}\in \mathsf{Cpl}(\rho,\varrho)} \int \langle y, z\rangle \, \tilde{\pi}(dy,dz), \qquad p,\varrho \in \mathscr{P}_2(\mathbb{R}^d).$$

Here, $\mathsf{Cpl}(p,\varrho)$ is the collection of all couplings $\tilde{\pi}$ between p and ϱ, i.e. probability measures $\tilde{\pi}$ on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal p and second marginal ϱ. As a first result, we show for the value of the martingale optimization problem (P) that

$$P(\mu,\nu) = \inf_{\alpha\in\mathscr{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{$\mathcal{V}$}$$

Moreover, the right-hand side of ($\mathcal{V}$) is attained by $\hat{\alpha}$ if and only if there is a Bass martingale from μ to ν with Bass measure $\hat{\alpha}$.

We also establish in Part VII an infinitesimal version of the result above. Let $(M_t)_{0\leqslant t\leqslant 1}$ be an $\mathbb{R}^d$-valued martingale bounded in L^2, which is given by the stochastic integral

$$M_t = M_0 + \int_0^t \sigma_s \, dB_s, \qquad 0 \leqslant t \leqslant 1,$$

where $(\sigma_t)_{0\leqslant t\leqslant 1}$ is a progressively measurable process. Denote by μ_t the law of M_t. For Lebesgue-a.e. $t \in [0,1]$ we have, for each $\alpha \in \mathscr{P}_2(\mathbb{R}^d)$, the inequality

$$\mathbb{E}\big[\mathrm{tr}(\sigma_t)\big] \leqslant \liminf_{h\to 0} \tfrac{1}{h}\Big(\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h}) - \mathrm{MCov}(\alpha,\mu_t)\Big).$$

For a Bass martingale $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ of the form

$$d\hat{M}_t = \hat{\sigma}_t(\hat{M}_t)\, dB_t,$$

with associated Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ and diffusion function $\hat{\sigma}_t \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$, we have, for Lebesgue-a.e. $t \in [0,1]$, the equality

$$\mathbb{E}[\mathrm{tr}(\hat{\sigma}_t(\hat{M}_t))] = \frac{d}{dt}\,\mathrm{MCov}(\hat{\alpha} * \gamma^t, \hat{\mu}_t),$$

where $\hat{\mu}_t = \mathrm{Law}(\hat{M}_t)$.

Finally, we prove in Part VII the displacement convexity of the Bass functional $\mathcal{V}(\,\cdot\,)$ along certain generalized geodesics in the quadratic Wasserstein space.

Next, we recall from [BBHK20] that the continuous-time martingale optimization problem (P) has the same value as the discrete-time formulation

$$\tilde{P}(\mu,\nu) = \sup_{\pi\in\mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x,\gamma)\,\mu(dx). \tag{$\tilde{P}$}$$

Here, the set $\mathsf{MT}(\mu,\nu)$ denotes the collection of all martingale transports, i.e. couplings $\pi \in \mathsf{Cpl}(\mu,\nu)$ satisfying $\mathrm{bary}(\pi_x) \coloneqq \int y\,\pi_x(dy) = x$, for μ-a.e. $x \in \mathbb{R}^d$. The family of probability measures $\{\pi_x\}_{x\in\mathbb{R}^d}$ on $\mathbb{R}^d$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e. $\pi(dx,dy) = \pi_x(dy)\,\mu(dx)$. Furthermore, it was shown in [BBHK20] that there exists a unique optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ of $(\tilde{P})$, and if $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ is the stretched Brownian motion from μ to ν, then the law of $(\hat{M}_0,\hat{M}_1)$ is equal to $\hat{\pi}$.

In Part VIII we address an extension of the problem $(\tilde{P})$, in which the Gaussian distribution γ is replaced by an arbitrary reference measure q. More precisely, we consider the martingale optimization problem

$$P^q(\mu,\nu) = \sup_{\pi\in\mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x,q)\,\mu(dx). \tag{P^q}$$

We assume that the probability measure $q \in \mathscr{P}_2(\mathbb{R}^d)$ does not give mass to small sets, i.e. measurable sets with Hausdorff dimension at most $d-1$ in $\mathbb{R}^d$. This is the usual assumption of optimal transport under which Brenier's theorem (see, e.g. [Vil03, Theorem 2.12]) is formulated. As our first main result, we show that the primal problem (P^q) is uniquely attained and that its value $P^q(\mu,\nu)$ is equal to the value of the dual problem

$$D^q(\mu,\nu) = \inf_{\substack{\psi\in L^1(\nu),\\ \psi\ \text{convex}}} \Big(\int \psi\,d\nu - \int (\psi^\star \star q)^*\,d\mu \Big), \tag{D^q}$$

where

$$(\psi^\star \star q)(y) \coloneqq \int \psi^*(y+z)\,q(dz), \qquad y \in \mathbb{R}^d.$$

If q is symmetric (like, e.g. in the Gaussian case $q=\gamma$), the symbol $\star$ coincides with the standard convolution operator $*$.

The second main contribution of Part VIII is the introduction of q-Bass martingales in discrete time. This notion extends the concept of Bass martingales, as defined in (Bass) above, beyond the Gaussian case. A q-Bass martingale is a pair $(\hat{v}, \hat{\alpha})$ consisting of a convex function $\hat{v}$ and a probability measure $\hat{\alpha}$ satisfying the fundamental relations

$$(\nabla \hat{v} \star q)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * q) = \nu,$$

which we summarize in the following graphic.

$$\begin{array}{ccc} \hat{\alpha} * q & \xrightarrow{\nabla \hat{v}} & \nu \\ {\scriptstyle *}\uparrow & & \\ \hat{\alpha} & \xrightarrow{\nabla \hat{v} \star q} & \mu \end{array}$$

Finally, in Part VIII we also give sufficient conditions for the existence of q-Bass martingales. We raise some interesting open questions in the Concluding discussion.

Synopsis of the publications and manuscripts

This cumulative dissertation consists of the following publications and manuscripts:

Part	Reference	Status
Part I	[KST22]	published
Part II	[KT22]	published
Part III	[TY23]	published
Part IV	[Tsc23]	submitted
Part V	[KST20]	unpublished
Part VI	[BBST23]	submitted
Part VII	[BST23]	submitted
Part VIII	[Tsc24]	submitted

More precisely, Parts I, II and III of this thesis correspond, respectively, to the following peer-reviewed publications:

[KST22] 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*, **66**(4):668–707, 2022.

[KT22] I. Karatzas and B. Tschiderer. A variational characterization of Langevin–Smoluchowski diffusions. In G. Yin and T. Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization — A Commemorative Volume to Honor Mark H.A. Davis's Contributions*, pages 239–265. Springer, Cham, 2022.

[TY23] B. Tschiderer and L.C. Yeung. A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: gradient flows and HWBI inequalities. *Bernoulli*, **29**(1):725–756, 2023.

The content of Parts IV, VI, VII and VIII is available on arXiv:

[Tsc23] B. Tschiderer. Diffusion processes as Wasserstein gradient flows via stochastic control of the volatility matrix. arXiv:2310.18678, 2023.

[BBST23] J. Backhoff-Veraguas, M. Beiglböck, W. Schachermayer, and B. Tschiderer. The structure of martingale Benamou–Brenier in $\mathbb{R}^d$. arXiv:12306.11019, 2023.

[BST23] J. Backhoff-Veraguas, W. Schachermayer, and B. Tschiderer. The Bass functional of martingale transport. arXiv:2309.11181, 2023.

[Tsc24] B. Tschiderer. q-Bass martingales. arXiv:2402.05669, 2024.

These papers have been submitted to peer-reviewed journals and are currently under review.

The material of Part V stems from Appendix G of the unpublished manuscript

[KST20] I. Karatzas, W. Schachermayer, and B. Tschiderer. Trajectorial Otto calculus. arXiv:1811.08686, 2020.

and is not intended for publication.

Indication of the submitter's own contribution

For better readability, this section is written in the first-person perspective.

The initial impetus behind the publication [KST22] goes back to the time when I was a master student and my supervisor Walter Schachermayer suggested to work on this topic. We then decided to join forces with our friend and coauthor Ioannis Karatzas, which resulted in the publication [KST22]. All three of us contributed equally to the work.

The article [KT22] builds directly on [KST22]. Ioannis Karatzas and I wrote it together and both of us shared the work equally.

The publication [TY23] is joint work with my coauthor Lane Chun Yeung, who was at that time a PhD student under the supervision of Ioannis Karatzas and Daniel Lacker. We were both equally involved in this project.

The paper [BBST23] is joint work with Julio Backhoff-Veraguas and my two supervisors, Mathias Beiglböck and Walter Schachermayer; the paper [BST23] is joint work with Julio Backhoff-Veraguas and Walter Schachermayer. All authors contributed equally to both papers.

Finally, the manuscripts [Tsc23] and [Tsc24] are entirely my own work.

Part I.

A trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions

Joint work with Ioannis Karatzas and Walter Schachermayer

Abstract

We revisit the variational characterization of conservative diffusion as entropic gradient flow and provide for it a probabilistic interpretation via a perturbation analysis based on stochastic calculus. It was shown by Jordan, Kinderlehrer and Otto that, for diffusions of Langevin–Smoluchowski type, the Fokker–Planck probability density flow maximizes the rate of relative entropy dissipation, as measured by the distance traveled in the ambient space of probability measures with finite second moments, in terms of the quadratic Wasserstein metric. We obtain novel, stochastic-process versions of these features, valid along almost every trajectory of the diffusive motion in the backward direction of time, using a direct perturbation methodology. By averaging our trajectorial results with respect to the underlying measure on path space, we establish the maximal rate of entropy dissipation along the Fokker–Planck flow and measure exactly the deviation from this maximum that corresponds to a given perturbation. As a bonus of our trajectorial approach we derive the HWI inequality relating relative entropy (H), Wasserstein distance (W) and relative Fisher information (I).

1. Introduction

We provide a trajectorial interpretation of a seminal result by Jordan, Kinderlehrer and Otto [JKO98], and present a proof based on stochastic analysis. The basic theme of our approach could be described epigrammatically as "applying Itô calculus to Otto calculus". More precisely, we follow a stochastic analysis approach to the characterization of diffusions of Langevin–Smoluchowski type as entropic gradient flows in Wasserstein space, as in [JKO98]. We provide stronger, trajectorial versions of these results. For consistency and readability we adopt the setting and notation of [JKO98], and even copy some paragraphs of this paper almost verbatim in the remainder of this chapter.

Along the lines of [JKO98], we consider thus a Fokker–Planck or forward Kolmogorov [Kol31] equation of the form

$$\partial_t p(t,x) = \operatorname{div}\bigl(\nabla\Psi(x)\, p(t,x)\bigr) + \tfrac{1}{2}\Delta p(t,x), \qquad (t,x) \in (0,\infty)\times\mathbb{R}^n, \tag{1.1}$$

with initial condition

$$p(0,x) = p^0(x), \qquad x \in \mathbb{R}^n. \tag{1.2}$$

Here, $p(\,\cdot\,,\cdot\,)$ is a real-valued function defined for $(t,x) \in [0,\infty)\times\mathbb{R}^n$, the function $\Psi\colon \mathbb{R}^n \to [0,\infty)$ is smooth and plays the role of a potential, and $p^0(\,\cdot\,)$ is a probability density on $\mathbb{R}^n$. The solution $p(t,x)$ of (1.1) with initial condition (1.2) stays nonnegative and conserves its mass, which means that the spatial integral $\int_{\mathbb{R}^n} p(t,x)\,\mathrm{d}x$ is independent of the time parameter $t \geqslant 0$ and is thus equal to $\int_{\mathbb{R}^n} p^0(x)\,\mathrm{d}x = 1$. Therefore, $p(t,\,\cdot\,)$ must be a probability density on $\mathbb{R}^n$ for every fixed time $t \geqslant 0$.

As in [JKO98] we note that the Fokker–Planck equation (1.1) with initial condition (1.2) is inherently related (see [Fri75, Gar09, Ris96, Sch80]) to the stochastic differential equation of Langevin–Smoluchowski type

$$\mathrm{d}X(t) = -\nabla\Psi\bigl(X(t)\bigr)\,\mathrm{d}t + \mathrm{d}W(t), \qquad t \geqslant 0. \tag{1.3}$$

In the equation above, $(W(t))_{t\geqslant 0}$ is an n-dimensional Brownian motion started at the origin, and the $\mathbb{R}^n$-valued random variable $X(0)$ is independent of the process $(W(t))_{t\geqslant 0}$. The probability distribution of $X(0)$ has density $p^0(\,\cdot\,)$ and, unless specified otherwise, the reference measure will always be Lebesgue measure on $\mathbb{R}^n$. Then $p(t,\,\cdot\,)$, the solution of (1.1) with initial condition (1.2), gives at any given time $t \geqslant 0$ the probability density function of the random variable $X(t)$ from (1.3).

If the potential Ψ grows rapidly enough so that $\mathrm{e}^{-2\Psi} \in L^1(\mathbb{R}^n)$, then the partition constant

$$Z = \int_{\mathbb{R}^n} \mathrm{e}^{-2\Psi(x)}\,\mathrm{d}x \tag{1.4}$$

is finite and there exists a unique stationary solution of the Fokker–Planck equation (1.1); namely, the probability density $q_Z(\,\cdot\,)$ of the *Gibbs distribution* [Gar09, JK96, Ris96] given by

$$q_Z(x) = Z^{-1}\,\mathrm{e}^{-2\Psi(x)}, \qquad x \in \mathbb{R}^n. \tag{1.5}$$

When it exists, the probability measure on $\mathbb{R}^n$ with density function $q_Z(\,\cdot\,)$ is called Gibbs distribution, and it is the unique invariant measure for the Markov process $(X(t))_{t\geqslant 0}$ defined by the stochastic differential equation (1.3); see, e.g., [KS98, Exercise 5.6.18, p. 361].

In [JK96] it is shown that the stationary probability density $q_Z(\,\cdot\,)$ satisfies the following variational principle: it minimizes the *free energy functional*

$$\mathscr{F}(p) = \mathcal{E}(p) + \tfrac{1}{2}\,\mathcal{S}(p) \tag{1.6}$$

over all probability densities p on $\mathbb{R}^n$. Here, the functionals

$$\mathcal{E}(p) := \int_{\mathbb{R}^n} \Psi(x)\,p(x)\,\mathrm{d}x, \qquad\qquad \mathcal{S}(p) := \int_{\mathbb{R}^n} p(x)\log p(x)\,\mathrm{d}x$$

model the potential energy and the internal energy (given by the negative of the Gibbs–Boltzmann entropy functional), respectively.

1.1. Preview

We set up in Chapter 2 the model for the Langevin–Smoluchowski diffusion, and introduce its fundamental quantities: the current and the invariant distributions of particles, the resulting likelihood ratio process, the associated concepts of free energy, relative entropy and relative Fisher information. In Section 2.1 we discuss the regularity assumptions imposed in our probabilistic approach.

Chapters 3 and 4 present the basic results. Foremost among these is Theorem 3.1, which computes in terms of the relative Fisher information the rate of decay for the relative entropy, in the ambient Wasserstein space of probability density functions with finite second moment; and its "perturbed" counterpart, Theorem 3.3. We compute explicitly the difference between these perturbed and unperturbed rates, and show that it is always nonnegative — in fact strictly positive, unless the perturbation and the gradient of the log-likelihood ratio function are collinear. This way, the Langevin–Smoluchowski diffusion emerges as the *steepest descent* (or "gradient flow") of the relative entropy functional with respect to the Wasserstein metric — and the celebrated result of [JKO98] receives a crisp, direct probabilistic treatment via perturbation analysis. This is the main contribution of the present work.

Certain aspects of Theorems 3.1 and 3.3 are well known, and the special (quadratic) case $\Psi(x) = \frac{1}{2}|x|^2$ of Ornstein–Uhlenbeck dynamics goes back as far as the 1950's. Our novel contribution here, is that Theorems 3.1 and 3.3 are simple consequences of their stronger, trajectorial versions, Theorems 4.1 and 4.2, respectively. These results provide very detailed descriptions for the semimartingale dynamics of the relative entropy

process in both its "pure" and "perturbed" forms, and are most transparent when time is *reversed*. Theorems 3.1 and 3.3 then follow from Theorems 4.1 and 4.2, simply by taking expectations.

Several consequences and ramifications of Theorems 4.1 and 4.2 are developed in Sections 4.1 and 4.2, including a derivation of the famous HWI inequality of Otto and Villani [OV00, Vil03, Vil09, CE02] that relates relative entropy (H) to Wasserstein distance (W) and to relative Fisher information (I). Detailed arguments and proofs of the main results are collected in Chapter 5. The limiting behavior of the Wasserstein distance along the Langevin–Smoluchowski diffusion is analyzed in Section 5.5; here, most of the effort goes into showing that relative entropy and Wasserstein distance have exactly the same exceptional sets of zero Lebesgue measure, for their temporal rate of change. This, seemingly purely technical, point, is of paramount importance for the rigorous justification of the perturbation analysis deployed in Theorem 3.3; it turns out also to involve a rather delicate analysis.

Finally, in Chapter 6 we provide the proofs of auxiliary results and some technical lemmas.

2. The stochastic approach

In Chapter 1 we were mostly quoting from [JKO98]. We adopt now a more probabilistic point of view, and translate our setting into the language of stochastic processes and probability measures.

Let $P(0)$ be a probability measure on the Borel sets of $\mathbb{R}^n$ with probability density function $p^0(\,\cdot\,) = p(0,\,\cdot\,)$. This measure induces a probability measure $\mathbb{P}$ on path space $\Omega = \mathcal{C}(\mathbb{R}_+;\mathbb{R}^n)$ of $\mathbb{R}^n$-valued continuous functions on $\mathbb{R}_+ = [0,\infty)$, under which the canonical coordinate process $(X(t,\omega))_{t\geqslant 0} = (\omega(t))_{t\geqslant 0}$ satisfies the stochastic differential equation (1.3) with initial probability distribution $P(0)$. We shall denote by $P(t)$ the probability distribution of the random vector $X(t)$ under $\mathbb{P}$, and by $p(t,\,\cdot\,)$ the corresponding probability density function, at each time $t \geqslant 0$. This function solves the equation (1.1) with initial condition (1.2).

An important role will be played by the *Radon–Nikodým derivative*, or *likelihood ratio process*,

$$\frac{\mathrm{d}P(t)}{\mathrm{dQ}}(X(t)) = \ell(t,X(t)), \qquad \text{where} \quad \ell(t,x) := \frac{p(t,x)}{q(x)} = p(t,x)\,\mathrm{e}^{2\Psi(x)} \tag{2.1}$$

for $t \geqslant 0$ and $x \in \mathbb{R}^n$. Here and throughout, we denote by Q the σ-finite measure on the Borel sets of $\mathbb{R}^n$, whose density with respect to Lebesgue measure is

$$q(x) := \mathrm{e}^{-2\Psi(x)}, \qquad x \in \mathbb{R}^n. \tag{2.2}$$

The *relative entropy* and the *relative Fisher information* (see, e.g., [OV00, CT06]) of $P(t)$ with respect to this measure Q, are defined respectively as

$$H(P(t)\,|\,\mathrm{Q}) := \mathbb{E}_{\mathbb{P}}\big[\log \ell(t,X(t))\big] = \int_{\mathbb{R}^n} \log\left(\frac{p(t,x)}{q(x)}\right) p(t,x)\,\mathrm{d}x, \tag{2.3}$$

$$I(P(t)\,|\,\mathrm{Q}) := \mathbb{E}_{\mathbb{P}}\Big[\,\big|\nabla \log \ell(t,X(t))\big|^2\,\Big] = \int_{\mathbb{R}^n} \big|\nabla \log \ell(t,x)\big|^2\, p(t,x)\,\mathrm{d}x. \tag{2.4}$$

Following the approach of [Léo14b, Section 3], we show in Section 6.1 that the relative entropy $H(P\,|\,\mathrm{Q})$ is well-defined and takes values in $(-\infty,\infty]$, for every P in the collection $\mathscr{P}_2(\mathbb{R}^n)$ of probability measures on the Borel sets $\mathscr{B}(\mathbb{R}^n)$ of $\mathbb{R}^n$ with finite second moment.

Direct computation reveals that, along the curve of probability measures $(P(t))_{t\geqslant 0}$, the free energy functional (1.6) and the relative entropy (2.3) are related for each $t \geqslant 0$ through the equation

$$2\,\mathscr{F}(p(t,\,\cdot\,)) = H(P(t)\,|\,\mathrm{Q}). \tag{2.5}$$

This shows that studying the decay of the free energy $t \mapsto \mathscr{F}(p(t,\,\cdot\,))$ is equivalent to studying the decay of the relative entropy $t \mapsto H(P(t)\,|\,\mathrm{Q})$, a key aspect of thermodynamics. In light of condition (ii) in Assumptions 2.1 below, the quantity in (2.3) is finite for $t = 0$; thus, on account of (4.12) below, finite also for $t > 0$.

2.1. Regularity assumptions

In order to provide mathematically precise formulations of subsequent results, we have to specify convenient regularity assumptions. These issues are of a rather technical nature, and Section 2.1 might be skipped at a first reading.

By analogy with [JKO98, Theorem 5.1] we consider the following assumptions.

Assumptions 2.1.

(i) The potential $\Psi\colon \mathbb{R}^n \to [0,\infty)$ is of class $\mathcal{C}^\infty(\mathbb{R}^n;[0,\infty))$.

(ii) The probability distribution $P(0)$ of $X(0)$ in (1.3) has probability density function $p^0(\,\cdot\,) = p(0,\,\cdot\,)$ with respect to Lebesgue measure on $\mathbb{R}^n$, with finite second moment and finite free energy, i.e.,

$$\int_{\mathbb{R}^n} p^0(x)\,|x|^2\,\mathrm{d}x < \infty \quad \text{and} \quad \mathscr{F}(p^0) = \tfrac{1}{2}H\big(P(0)\,|\,\mathrm{Q}\big) \in (-\infty,\infty). \tag{2.6}$$

In [JKO98] it is also assumed that the potential Ψ satisfies, for some real constant $C > 0$, the bound $|\nabla\Psi| \leqslant C\,(\Psi + 1)$, which we do not need here. Instead of this requirement, we shall impose the following rather weak assumptions.

Assumptions 2.2 (Regularity assumptions for the trajectorial results)**.**
In addition to conditions (i) and (ii) of Assumptions 2.1, we also impose that:

(iii) The potential Ψ satisfies, for some real constants $c \geqslant 0$ and $R \geqslant 0$, the drift (or coercivity) condition

$$\forall\, x \in \mathbb{R}^n, |x| \geqslant R\colon \qquad \big\langle x\,,\nabla\Psi(x)\big\rangle \geqslant -c\,|x|^2. \tag{2.7}$$

(iv) The potential $\Psi\colon \mathbb{R}^n \to [0,\infty)$ is sufficiently well-behaved to guarantee that the solution $(t,x) \mapsto p(t,x)$ of (1.1) with initial condition (1.2) is continuous and strictly positive on $(0,\infty)\times\mathbb{R}^n$, differentiable with respect to the time variable t for each $x \in \mathbb{R}^n$, and smooth in the space variable x for each $t > 0$. We also assume that the logarithmic derivative $(t,x) \mapsto \nabla \log p(t,x)$ is continuous on $(0,\infty)\times\mathbb{R}^n$. For example, by requiring that all derivatives of Ψ grow at most exponentially as $|x|$ tends to infinity, one may adapt the arguments from [Rog85] showing that this is indeed the case.

For the formulation of Theorem 3.3 we will need a vector field $\beta\colon \mathbb{R}^n \to \mathbb{R}^n$ which is the gradient of a potential $B\colon \mathbb{R}^n \to \mathbb{R}$ satisfying the following regularity assumption:

(v) The potential $B\colon \mathbb{R}^n \to \mathbb{R}$ is of class $\mathcal{C}^\infty(\mathbb{R}^n;\mathbb{R})$ and has compact support. Consequently, its gradient $\beta := \nabla B\colon \mathbb{R}^n \to \mathbb{R}^n$ is of class $\mathcal{C}^\infty(\mathbb{R}^n;\mathbb{R}^n)$ and again compactly supported. We also assume that, for every such β, the perturbed potential $\Psi + B$ satisfies condition (iv).

The Assumptions 2.2 are satisfied by typical convex potentials Ψ. They also accommodate examples such as double-well potentials of the form $\Psi(x) = (x^2 - \alpha^2)^2$ on the real line, for real constants $\alpha > 0$. It is important to point out, that these assumptions do not rule out the case when the constant Z in (1.4) is infinite; thus, they allow for cases (such as $\Psi \equiv 0$) in which the stationary probability density function $q_Z(\,\cdot\,)$ in (1.5) does not exist. In fact, in [JKO98] the authors point out explicitly that, even when the stationary probability density $q_Z(\,\cdot\,)$ is not defined, the free energy (1.6) of a density $p(t,x)$ satisfying the Fokker–Planck equation (1.1) with initial condition (1.2) can be defined, provided that the free energy $\mathscr{F}(p^0)$ is finite. Furthermore, we note that the Assumptions 2.2 are designed in such a way that they are invariant when passing from the potential Ψ to $\Psi + B$ if B satisfies condition (v).

Lemma 2.3. *Under the Assumptions 2.2, the Langevin–Smoluchowski diffusion equation (1.3) with initial distribution $P(0)$ admits a pathwise unique, strong solution, which satisfies $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$ for all $t \geqslant 0$.*

The proof of Lemma 2.3 follows from the fact that the drift condition (2.7) guarantees that the second-moment condition in (2.6) propagates in time, i.e.,

$$\forall\, t \geqslant 0\colon \qquad \int_{\mathbb{R}^n} p(t,x)\,|x|^2\,\mathrm{d}x < \infty; \tag{2.8}$$

see [Fri75, Theorem 2.2, p. 104] and the first problem on p. 125 of [Fri75]. We give a proof of (2.8) in Section 6.2. This property also holds when the potential Ψ is replaced by $\Psi + B$ as in condition (v) of Assumptions 2.2; see Lemma 3.2 below.

Assumptions 2.4 (Regularity assumptions regarding the Wasserstein distance)**.**
In addition to conditions (i) – (v) of Assumptions 2.2, we require that:

(vi) For every $t \geqslant 0$, there exists a sequence of functions $(\varphi_m(t,\,\cdot\,))_{m\geqslant 1} \subseteq \mathcal{C}_c^\infty(\mathbb{R}^n;\mathbb{R})$, whose gradients $(\nabla\varphi_m(t,\,\cdot\,))_{m\geqslant 1}$ converge in the space $L^2(P(t))$ to the velocity field $v(t,\,\cdot\,) = \nabla\varphi(t,\,\cdot\,)$ of gradient type as in (5.47) with $\varphi(t,x) = -\Psi(x) - \frac{1}{2}\log p(t,x)$, as $m \to \infty$.

This last requirement guarantees, for every $t \geqslant 0$, that the velocity field $v(t,\,\cdot\,)$ is an element of the tangent space of $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$ in the sense of [AGS08, Definition 8.4.1]. For the details we refer to Section 5.5 below, in particular, the display (5.48).

The condition (vi) in Assumptions 2.4 is satisfied by simple potentials such as for example $\Psi \equiv 0$ or $\Psi(x) = \frac{1}{2}|x|^2$. More generally, potentials with a curvature lower bound $\mathrm{Hess}(\Psi) \geqslant \kappa\, I_n$, for some $\kappa \in \mathbb{R}$ (as in (4.41) below), for instance the double-well potential $\Psi(x) = (x^2 - \alpha^2)^2$ on the real line, satisfy this condition; more on this theme can be found in [AGS08, Theorem 10.4.13].

3. The main theorems in aggregate form

In light of (2.5), the goal of [JKO98] is to relate the decay of the *relative entropy functional*

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P \,|\, \mathrm{Q}) \in (-\infty, \infty] \tag{3.1}$$

along the curve $(P(t))_{t\geqslant 0}$, to the quadratic Wasserstein distance

$$W_2(\mu,\nu) := \Big(\inf_{Y\sim\mu, Z\sim\nu} \mathbb{E}\big[|Y-Z|^2\big] \Big)^{1/2}, \qquad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^n)$$

on $\mathscr{P}_2(\mathbb{R}^n)$ (cf. [Vil03, AGS08, AG13]). We resume the remarkable relation between these two quantities in the following two theorems; these quantify the relationship between displacement in the ambient space (the denominator in (3.4)) and fluctuations of the free energy, or equivalently of the relative entropy (the numerator in (3.4)). The proofs will be given in Section 4.1 below.

Theorem 3.1. *Under the Assumptions 2.4, the relative Fisher information $I(P(t_0) \,|\, \mathrm{Q})$ is finite for Lebesgue-a.e. $t_0 \geqslant 0$, and we have the* generalized de Bruijn identity

$$\lim_{t\to t_0} \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2}\, I(P(t_0) \,|\, \mathrm{Q}), \tag{3.2}$$

as well as the limiting behavior of the quadratic Wasserstein distance

$$\lim_{t\to t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \tfrac{1}{2} \sqrt{I(P(t_0) \,|\, \mathrm{Q})}, \tag{3.3}$$

so that

$$\lim_{t\to t_0} \left(\operatorname{sgn}(t - t_0) \cdot \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{W_2(P(t), P(t_0))} \right) = -\sqrt{I(P(t_0) \,|\, \mathrm{Q})}. \tag{3.4}$$

Furthermore, if $t_0 \geqslant 0$ is chosen so that the generalized de Bruijn identity (3.2) does hold, then the limiting assertions (3.3) and (3.4) are also valid.

The ratio on the left-hand side of (3.4) can be interpreted as the rate of decay for the relative entropy functional (3.1) at $P = P(t_0)$ along the curve $(P(t))_{t\geqslant 0}$, if distances in the ambient space $\mathscr{P}_2(\mathbb{R}^n)$ are measured by the quadratic Wasserstein distance W_2. The quantity appearing on the right-hand side of (3.4) is the square root of the relative Fisher information in (2.4), written more explicitly in terms of the "score function" $\nabla\ell(t,\,\cdot\,)/\ell(t,\,\cdot\,)$ as

$$I(P(t) \,|\, \mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\left[\frac{\big|\nabla \ell(t, X(t))\big|^2}{\ell(t, X(t))^2} \right] = \int_{\mathbb{R}^n} \left| \frac{\nabla p(t,x)}{p(t,x)} + 2\,\nabla\Psi(x) \right|^2 p(t,x)\,\mathrm{d}x. \tag{3.5}$$

For future reference, we denote by N the set of exceptional points $t_0 \geqslant 0$ for which the right-sided version of the limit in (3.2), i.e., the limiting assertion

$$\lim_{t \downarrow t_0} \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2}\, I(P(t_0) \,|\, \mathrm{Q}), \tag{3.6}$$

fails. According to Theorem 3.1, this exceptional set N has zero Lebesgue measure.

The remarkable insight of [JKO98] states that the rate of entropy decay (3.4) along the curve $(P(t))_{t \geqslant 0}$ is, in fact, the slope of *steepest descent* for the relative entropy functional (3.1) with respect to the Wasserstein distance W_2 at the point $P = P(t_0)$ on the curve. To formalize this assertion, we fix a time $t_0 \geqslant 0$ and let the vector field $\beta = \nabla B \colon \mathbb{R}^n \to \mathbb{R}^n$ be the gradient of a potential B, as in condition (v) of Assumptions 2.2. This gradient vector field β will serve as a perturbation in the thus perturbed Fokker–Planck equation

$$\partial_t p^\beta(t,x) = \operatorname{div}\Big(\big(\nabla \Psi(x) + \beta(x)\big)\, p^\beta(t,x) \Big) + \tfrac{1}{2} \Delta p^\beta(t,x), \tag{3.7}$$

for $(t,x) \in (t_0, \infty) \times \mathbb{R}^n$, with initial condition

$$p^\beta(t_0, x) = p(t_0, x), \qquad x \in \mathbb{R}^n. \tag{3.8}$$

We denote by $\mathbb{P}^\beta$ the probability measure on path space $\Omega = \mathcal{C}([t_0, \infty); \mathbb{R}^n)$, under which the canonical coordinate process $(X(t))_{t \geqslant t_0}$ satisfies the stochastic differential equation

$$\mathrm{d}X(t) = -\Big(\nabla \Psi(X(t)) + \beta(X(t)) \Big)\, \mathrm{d}t + \mathrm{d}W^\beta(t), \qquad t \geqslant t_0 \tag{3.9}$$

with initial probability distribution $P(t_0)$. Here, the process $(W^\beta(t))_{t \geqslant t_0}$ is Brownian motion under $\mathbb{P}^\beta$. The probability distribution of $X(t)$ under $\mathbb{P}^\beta$ on $\mathbb{R}^n$ will be denoted by $P^\beta(t)$, for $t \geqslant t_0$; as before, the corresponding probability density function $p^\beta(t, \,\cdot\,)$ solves the equation (3.7) subject to the initial condition (3.8).

In the following analogue of Lemma 2.3, we state that the perturbed probability density functions $p^\beta(t, \,\cdot\,)$ of (3.7), (3.8) also admit finite second moments at all times $t \geqslant t_0$. For the proof we refer again to Section 6.2.

Lemma 3.2. *Under the Assumptions 2.2, let $t_0 \geqslant 0$. Then the perturbed diffusion equation (3.9) with initial distribution $P^\beta(t_0) = P(t_0)$ admits a pathwise unique, strong solution, which satisfies $P^\beta(t) \in \mathscr{P}_2(\mathbb{R}^n)$ for all $t \geqslant t_0$.*

After these preparations we can state the result formalizing the gradient flow, or *steepest descent*, property of the curve $(P(t))_{t \geqslant 0}$ generated by the Langevin–Smoluchowski diffusion (1.3) in the ambient space of probability measures $\mathscr{P}_2(\mathbb{R}^n)$ endowed with the quadratic Wasserstein metric.

Theorem 3.3. *Under the Assumptions 2.4, the following assertions hold for every point $t_0 \in \mathbb{R}_+ \setminus N$ (at which the right-sided limiting identity (3.6) is valid):*

The $\mathbb{R}^n$-valued random vectors

$$\begin{aligned} a &:= \nabla \log \ell(t_0, X(t_0)) = \nabla \log p(t_0, X(t_0)) + 2\,\nabla\Psi(X(t_0)), \\ b &:= \beta(X(t_0)) \end{aligned} \tag{3.10}$$

are elements of the Hilbert space $L^2(\mathbb{P})$, and the perturbed version of the generalized de Bruijn identity (3.2) reads

$$\begin{aligned} \lim_{t \downarrow t_0} \frac{H(P^\beta(t) \,|\, \mathrm{Q}) - H(P^\beta(t_0) \,|\, \mathrm{Q})}{t - t_0} &= -\tfrac{1}{2}\, I(P(t_0) \,|\, \mathrm{Q}) - \langle a, b \rangle_{L^2(\mathbb{P})} \\ &= -\tfrac{1}{2}\, \langle a, a + 2b \rangle_{L^2(\mathbb{P})}. \end{aligned} \tag{3.11}$$

The limiting behavior of the quadratic Wasserstein distance (3.3) in this perturbed context is given by

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \tfrac{1}{2}\, \|a + 2b\|_{L^2(\mathbb{P})}. \tag{3.12}$$

Combining (3.11) with (3.12), and assuming $a + 2b \neq 0$, we have

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) \,|\, \mathrm{Q}) - H(P^\beta(t_0) \,|\, \mathrm{Q})}{W_2(P^\beta(t), P^\beta(t_0))} = -\left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}, \tag{3.13}$$

and therefore

$$\lim_{t \downarrow t_0} \left(\frac{H(P^\beta(t) \,|\, \mathrm{Q}) - H(P^\beta(t_0) \,|\, \mathrm{Q})}{W_2(P^\beta(t), P^\beta(t_0))} - \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{W_2(P(t), P(t_0))} \right) \tag{3.14}$$

$$= \|a\|_{L^2(\mathbb{P})} - \left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \tag{3.15}$$

On the strength of the Cauchy–Schwarz inequality, the expression in (3.15) is nonnegative, and vanishes if and only if $a + 2b$ is a positive multiple of a. Consequently, when the vector field β is not a scalar multiple of $\nabla \log \ell(t_0, \,\cdot\,)$, the difference of the two slopes in (3.14) is strictly positive. In other words, the slope quantified by the first term of the difference (3.14), is then strictly bigger than the (negative) slope expressed by the second term of (3.14).

Theorems 3.1 and 3.3 are essentially well known. They build upon a vast amount of previous work. In the quadratic case $\Psi(x) = \frac{1}{2}|x|^2$, i.e., when the process $(X(t))_{t \geqslant 0}$ in (1.3) is Ornstein–Uhlenbeck with invariant measure in (1.5) being standard Gaussian, the relation

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, H(P(t) \,|\, \mathrm{Q}) = -\tfrac{1}{2}\, I(P(t) \,|\, \mathrm{Q}) \tag{3.16}$$

has been known since [Sta59] as *de Bruijn's identity*. This relationship between the two fundamental information measures, due to Shannon and Fisher, respectively, is a dominant theme in many aspects of information theory and probability. We refer to the book [CT06] by Cover and Thomas for an account of the results by Barron, Blachman, Brown, Linnik, Rényi, Shannon, Stam and many others; in a similar vein, see also the seminal work [BÉ85] by Bakry and Émery, as well as the paper [MV00] by Markowich and Villani, and the book [Vil03] by Villani. Consult also Carlen and Soffer [CS91] and Johnson [Joh04] on the relation of (3.16) to the central limit theorem. For the connections with large deviations we refer to [ADPZ13] and [Fat16]. In [Föl88] related pathwise results are obtained for the case $\Psi \equiv 0$, i.e., when the process $(X(t))_{t\geqslant 0}$ is Brownian motion.

The paper [JKO98] broke new ground in this respect, as it considered a general potential Ψ and established the relation to the quadratic Wasserstein distance, culminating with the characterization of the curve $(P(t))_{t\geqslant 0}$ as a gradient flow. This relation was further investigated by Otto in the paper [Ott01], where the theory now known as *"Otto calculus"* was developed. For a recent application of Otto calculus to the Schrödinger problem, see [GLR20].

The statements of our Theorems 3.1 and 3.3 complement the existing results in some details, e.g., the precise form (3.15), measuring the difference of the two slopes appearing in (3.14). The main novelty of our approach, however, will only become apparent below with the formulation of Theorems 4.1 and 4.2, the trajectorial versions of Theorems 3.1 and 3.3.

4. The main theorems in trajectorial form

Our main goal is to investigate Theorems 3.1 and 3.3 in a trajectorial fashion, by considering the *relative entropy process*

$$\log \ell(t, X(t)) = \log \left(\frac{p(t, X(t))}{q(X(t))} \right) = \log p(t, X(t)) + 2\Psi(X(t)), \qquad t \geqslant 0 \tag{4.1}$$

along each trajectory of the canonical coordinate process $(X(t))_{t \geqslant 0}$, and calculating its dynamics (stochastic differential) under the probability measure $\mathbb{P}$. The $\mathbb{P}$-expectation of this quantity is, of course, the relative entropy in (2.3). A decisive tool in the analysis of the relative entropy process (4.1) is to *reverse time*, and use a remarkable insight due to Pavon [Pav89], and Fontbona and Jourdain [FJ16]. These authors consider the canonical coordinate process $(X(t))_{0 \leqslant t \leqslant T}$ on path space $\Omega = \mathcal{C}([0,T]; \mathbb{R}^n)$ in the reverse direction of time, i.e., they work with the time-reversed process

$$\big(X(T-s)\big)_{0 \leqslant s \leqslant T}.$$

It is then notationally convenient to consider a finite time interval $[0,T]$, rather than $\mathbb{R}_+$. For another application of time reversal in a similar context, see [Léo17].

At this stage it becomes important to specify the relevant filtrations: We denote by $(\mathcal{F}(t))_{t \geqslant 0}$ the smallest continuous filtration to which the canonical coordinate process $(X(t))_{t \geqslant 0}$ is adapted. That is, modulo $\mathbb{P}$-augmentation, we have

$$\mathcal{F}(t) = \sigma\big(X(u) \colon\ 0 \leqslant u \leqslant t\big), \qquad t \geqslant 0;$$

and we call $(\mathcal{F}(t))_{t \geqslant 0}$ the "filtration generated by $(X(t))_{t \geqslant 0}$". Likewise, we let

$$\mathbb{G} := \big(\mathcal{G}(T-s)\big)_{0 \leqslant s \leqslant T}$$

be the "filtration generated by the time-reversed process $(X(T-s))_{0 \leqslant s \leqslant T}$" in the same sense as before. In other words,

$$\mathcal{G}(T-s) = \sigma\big(X(T-u) \colon\ 0 \leqslant u \leqslant s\big), \qquad 0 \leqslant s \leqslant T, \tag{4.2}$$

modulo $\mathbb{P}$-augmentation. For the necessary measure-theoretic operations that ensure the continuity (from both left and right) of filtrations associated with continuous processes, the reader may consult Section 2.7 in [KS98]; in particular, Problems 7.1 – 7.6 and Proposition 7.7.

The following two Theorems 4.1 and 4.2 are the main new results of this work. They can be regarded as trajectorial versions of Theorems 3.1 and 3.3, whose proofs will

follow from Theorems 4.1 and 4.2 simply by taking expectations. Similar trajectorial approaches have already been applied to the temporal dissipation of relative entropy and Fisher information [DPP89, Pav89, FJ16], to the theory of optimal stopping [DK94], to Doob's martingale inequalities [ABP+13], and to the Burkholder–Davis–Gundy inequality [BS15]. An analogue of Theorem 4.1 in a different setting is provided by Theorem 1.4 in [FJ16]: it is formulated under a probability measure $\mathbb{Q}$ on path space, induced by a *probability measure* Q on $\mathbb{R}^n$. In our context, it is crucial to allow for an invariant Gibbs measure Q which has possibly infinite mass. Our emphasis then lies in establishing the square-integrability of the process (4.4), and in proving (4.6) under the mild assumption that the initial relative entropy $H(P(0)\,|\,\mathrm{Q})$ is finite.

The significance of Theorem 4.1 right below, is that the trade-off between the temporal decay of relative entropy, and the temporal growth of the quadratic Wasserstein distance along the curve of probability measures $(P(t))_{t\geqslant 0}$, both of which are characterized in terms of the cumulative relative Fisher information process, is valid not only in expectation, but also along (almost) every trajectory, provided we run time in the reverse direction. The implicit Euler scheme used in [JKO98] also reflects the idea of going back in time at each step of the discretization.

Theorem 4.1. *Under the Assumptions 2.2, we fix $T \in (0,\infty)$ and define the cumulative relative Fisher information process, accumulated from the right, as*

$$\begin{aligned} F(T-s) &:= \int_0^s \frac{1}{2}\frac{\big|\nabla \ell(T-u, X(T-u))\big|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u \\ &= \int_0^s \frac{1}{2}\Big|\frac{\nabla p(T-u, X(T-u))}{p(T-u, X(T-u))} + 2\,\nabla\Psi\big(X(T-u)\big)\Big|^2\,\mathrm{d}u \end{aligned} \tag{4.3}$$

for $0 \leqslant s \leqslant T$. Then the process $(M(T-s))_{0\leqslant s\leqslant T}$ given by

$$M(T-s) := \Big(\log \ell(T-s, X(T-s)) - \log \ell(T, X(T))\Big) - F(T-s) \tag{4.4}$$

is a square-integrable martingale of the backwards filtration $\mathbb{G}$ under the probability measure $\mathbb{P}$. More explicitly, the martingale of (4.4) can be represented as

$$M(T-s) = \int_0^s \left\langle \frac{\nabla \ell(T-u, X(T-u))}{\ell(T-u, X(T-u))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u)\right\rangle \tag{4.5}$$

for a $\mathbb{P}$-Brownian motion $\big(\overline{W}^{\mathbb{P}}(T-s)\big)_{0\leqslant s\leqslant T}$ of the backwards filtration $\mathbb{G}$. In particular, the quadratic variation of the martingale of (4.4) is given by the nondecreasing process in (4.3), up to the multiplicative factor of $1/2$, and we have

$$H(P(0)\,|\,\mathrm{Q}) - H(P(T)\,|\,\mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\big[F(0)\big] = \tfrac{1}{2}\int_0^T I(P(t)\,|\,\mathrm{Q})\,\mathrm{d}t < \infty. \tag{4.6}$$

Next, we state the trajectorial version of Theorem 3.3 — or equivalently, the "perturbed" analogue of Theorem 4.1. As we did in Theorem 3.3, in particular in the preceding equations (3.7) – (3.9), we consider the perturbation $\beta\colon \mathbb{R}^n \to \mathbb{R}^n$ and denote the *perturbed likelihood ratio function* by

$$\ell^\beta(t,x) := \frac{p^\beta(t,x)}{q(x)} = p^\beta(t,x)\,\mathrm{e}^{2\Psi(x)}, \qquad (t,x) \in [t_0,\infty)\times\mathbb{R}^n. \tag{4.7}$$

The stochastic analogue of this quantity is the *perturbed likelihood ratio process*

$$\ell^\beta(t,X(t)), \qquad t \geqslant t_0.$$

The logarithm of this process is the *perturbed relative entropy process*

$$\log \ell^\beta(t,X(t)) = \log\left(\frac{p^\beta(t,X(t))}{q(X(t))}\right) = \log p^\beta(t,X(t)) + 2\,\Psi(X(t)), \qquad t \geqslant t_0. \tag{4.8}$$

Theorem 4.2. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. We define the perturbed cumulative relative Fisher information process, accumulated from the right, as*

$$F^\beta(T-s) := \int_0^s \left(\frac{1}{2}\frac{\left|\nabla\ell^\beta(T-u,X(T-u))\right|^2}{\ell^\beta(T-u,X(T-u))^2} + \Big(\langle\beta\,,\,2\,\nabla\Psi\rangle - \operatorname{div}\beta\Big)(X(T-u))\right)\mathrm{d}u \tag{4.9}$$

for $0 \leqslant s \leqslant T-t_0$. Then

$$\mathbb{E}_{\mathbb{P}^\beta}\left[F^\beta(t_0)\right] < \infty,$$

and the process $(M^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ given by

$$M^\beta(T-s) := \Big(\log\ell^\beta(T-s,X(T-s)) - \log\ell^\beta(T,X(T))\Big) - F^\beta(T-s) \tag{4.10}$$

is a square-integrable martingale of the backwards filtration

$$\widetilde{\mathbb{G}} := \left(\mathcal{G}(T-s)\right)_{0\leqslant s\leqslant T-t_0}$$

under the probability measure $\mathbb{P}^\beta$. More explicitly, the martingale (4.10) can be represented as

$$M^\beta(T-s) = \int_0^s \left\langle \frac{\nabla\ell^\beta(T-u,X(T-u))}{\ell^\beta(T-u,X(T-u))}\,,\,\mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-u)\right\rangle \tag{4.11}$$

for a $\mathbb{P}^\beta$-Brownian motion $\left(\overline{W}^{\mathbb{P}^\beta}(T-s)\right)_{0\leqslant s\leqslant T-t_0}$ of the backwards filtration $\widetilde{\mathbb{G}}$.

4.1. Consequences of the trajectorial results

Before tackling the proofs of Theorems 4.1 and 4.2, we state several important consequences of these two basic results. In particular, we indicate how the corresponding assertions in the earlier Theorems 3.1 and 3.3 follow directly from these results by taking expectations.

Corollary 4.3 (Dissipation of relative entropy)**.**
Under the Assumptions 2.4, we have for all $t, t_0 \geqslant 0$ the relative entropy identity

$$\begin{aligned} H(P(t)\,|\,\mathrm{Q}) - H(P(t_0)\,|\,\mathrm{Q}) &= \mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{\ell(t,X(t))}{\ell(t_0,X(t_0))}\right)\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\int_{t_0}^{t}\left(-\frac{1}{2}\frac{|\nabla\ell(u,X(u))|^2}{\ell(u,X(u))^2}\right)\mathrm{d}u\right]. \end{aligned} \tag{4.12}$$

Furthermore, we have for Lebesgue-a.e. $t_0 \geqslant 0$ the generalized de Bruijn identity

$$\lim_{t\to t_0}\frac{H(P(t)\,|\,\mathrm{Q}) - H(P(t_0)\,|\,\mathrm{Q})}{t-t_0} = -\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\left[\frac{|\nabla\ell(t_0,X(t_0))|^2}{\ell(t_0,X(t_0))^2}\right], \tag{4.13}$$

as well as the limiting behavior of the quadratic Wasserstein distance

$$\lim_{t\to t_0}\frac{W_2(P(t),P(t_0))}{|t-t_0|} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\frac{|\nabla\ell(t_0,X(t_0))|^2}{\ell(t_0,X(t_0))^2}\right]\right)^{1/2}. \tag{4.14}$$

If $t_0 \geqslant 0$ is chosen so that the generalized de Bruijn identity (4.13) holds, then the limiting assertion (4.14) pertaining to the Wasserstein distance is also valid.

Proof of Corollary 4.3 from Theorem 4.1:
The identity (4.12) follows by taking expectations in (4.5) with respect to the probability measure $\mathbb{P}$, recalling the definitions (4.3), (4.4), and invoking the martingale property of the process in (4.4) for $T \geqslant \max\{t_0, t\}$. In particular, (4.12) shows that the relative entropy function $t \mapsto H(P(t)\,|\,\mathrm{Q})$ from (2.3), thus also the free energy function $t \mapsto \mathscr{F}(p(t,\,\cdot\,))$ from (2.5), are strictly decreasing provided $\ell(t,\,\cdot\,)$ is not constant.

By the Lebesgue differentiation theorem, the monotone function $t \mapsto H(P(t)\,|\,\mathrm{Q})$ is differentiable for Lebesgue-a.e. $t_0 \geqslant 0$, in which case (4.12) leads to the identity (4.13).

The limiting behavior (4.14) of the Wasserstein distance, for Lebesgue-a.e. $t_0 \geqslant 0$, is well known and worked out in [AGS08]; Section 5.5 below provides details. Theorem 5.10 establishes the important, novel aspect of Corollary 4.3; namely, its last assertion, that the validity of (4.13) for some $t_0 \geqslant 0$ implies that the limiting assertion (4.14) also holds for the same point t_0. This seemingly harmless issue is actually quite delicate, and will be of crucial importance for our gradient flow analysis; it is here that we shall have to rely on condition (vi) of Assumptions 2.4. □

Proof of Theorem 3.1 from Theorem 4.1:
The assertions of Theorem 3.1 are direct consequences of Corollary 4.3. □

In a manner similar to the derivation of Corollary 4.3 from Theorem 4.1, we deduce now from Theorem 4.2 the following Corollary 4.4. Its first identity (4.15) shows, in particular, that the relative entropy $H(P^\beta(t)\,|\,\mathrm{Q})$ is finite for all $t \geqslant t_0$.

Corollary 4.4 (Dissipation of relative entropy under perturbations).
Under the Assumptions 2.4, we have, for all $t \geqslant t_0 \geqslant 0$, the relative entropy identity

$$\begin{aligned} H(P^\beta(t)\,|\,\mathrm{Q}) - H(P^\beta(t_0)\,|\,\mathrm{Q}) &= \mathbb{E}_{\mathbb{P}^\beta}\left[\log\left(\frac{\ell^\beta(t,X(t))}{\ell^\beta(t_0,X(t_0))}\right)\right] \\ &= \mathbb{E}_{\mathbb{P}^\beta}\left[\int_{t_0}^t\left(-\frac{1}{2}\frac{\left|\nabla\ell^\beta(u,X(u))\right|^2}{\ell^\beta(u,X(u))^2} + \Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle\Big)(X(u))\right)\mathrm{d}u\right]. \end{aligned} \tag{4.15}$$

Furthermore, for every point $t_0 \in \mathbb{R}_+ \setminus N$ (at which the right-sided limiting assertion (3.6) is valid), we have also the limiting identities

$$\begin{aligned} &\lim_{t\downarrow t_0}\frac{H(P^\beta(t)\,|\,\mathrm{Q}) - H(P^\beta(t_0)\,|\,\mathrm{Q})}{t-t_0} \\ &\qquad = \mathbb{E}_{\mathbb{P}}\left[-\frac{1}{2}\frac{\left|\nabla\ell(t_0,X(t_0))\right|^2}{\ell(t_0,X(t_0))^2} + \Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle\Big)(X(t_0))\right] \end{aligned} \tag{4.16}$$

and

$$\lim_{t\downarrow t_0}\frac{W_2(P^\beta(t),P^\beta(t_0))}{t-t_0} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\left|\frac{\nabla\ell(t_0,X(t_0))}{\ell(t_0,X(t_0))} + 2\,\beta(X(t_0))\right|^2\right]\right)^{1/2}. \tag{4.17}$$

Proof of Corollary 4.4 from Theorem 4.2:
Taking expectations in (4.11) under the probability measure $\mathbb{P}^\beta$, recalling the definitions (4.9), (4.10), and using the martingale property of the process in (4.10) for $T \geqslant t \geqslant t_0$, leads to the identity (4.15). In order to derive from (4.15) the limiting identity (4.16), extra care is needed to show that (4.16) is valid for every time $t_0 \in \mathbb{R}_+ \setminus N$.

We shall verify in Lemma 5.8 of Section 5.3 below the following estimates on the ratio between the probability density function $p(t,\,\cdot\,)$ and its perturbed version $p^\beta(t,\,\cdot\,)$: For every $t_0 \geqslant 0$ and $T > t_0$ there is a constant $C > 0$ such that

$$\left|\frac{\ell^\beta(t,x)}{\ell(t,x)} - 1\right| = \left|\frac{p^\beta(t,x)}{p(t,x)} - 1\right| \leqslant C\,(t-t_0)\,, \qquad (t,x) \in [t_0,T]\times\mathbb{R}^n \tag{4.18}$$

as well as

$$\mathbb{E}_{\mathbb{P}}\left[\int_{t_0}^t\left|\nabla\log\left(\frac{\ell^\beta(u,X(u))}{\ell(u,X(u))}\right)\right|^2\mathrm{d}u\right] \leqslant C\,(t-t_0)^2\,, \qquad t_0 \leqslant t \leqslant T. \tag{4.19}$$

We turn now to the derivation of (4.16) from (4.15). First, since the perturbation β is smooth and compactly supported, and the paths of the canonical coordinate process $(X(t))_{t\geqslant 0}$ are continuous, we have

$$\lim_{t\downarrow t_0}\frac{1}{t-t_0}\,\mathbb{E}_{\mathbb{P}^\beta}\left[\int_{t_0}^t\Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle\Big)(X(u))\,\mathrm{d}u\right] = \mathbb{E}_{\mathbb{P}^\beta}\left[\Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle\Big)(X(t_0))\right] \tag{4.20}$$

for every $t_0 \geqslant 0$. Secondly, the random variable $X(t_0)$ has the same distribution under $\mathbb{P}$, as it does under $\mathbb{P}^\beta$, so it is immaterial whether we express the expectation on the right-hand side of (4.20) with respect to the probability measure $\mathbb{P}$ or $\mathbb{P}^\beta$. Hence this expression equals the corresponding term on the right-hand side of (4.16).

Regarding the remaining term on the right-hand side of (4.16), the equality

$$\begin{aligned} &\lim_{t \downarrow t_0} \frac{1}{t - t_0} \, \mathbb{E}_{\mathbb{P}^\beta} \left[\int_{t_0}^t \left(- \frac{1}{2} \frac{|\nabla \ell^\beta(u, X(u))|^2}{\ell^\beta(u, X(u))^2} \right) \mathrm{d}u \right] \\ &\qquad = \lim_{t \downarrow t_0} \frac{1}{t - t_0} \, \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(- \frac{1}{2} \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} \right) \mathrm{d}u \right] \end{aligned} \tag{4.21}$$

holds as long as $t_0 \geqslant 0$ is chosen so that one of the limits exists. Indeed, the equality

$$\begin{aligned} &\lim_{t \downarrow t_0} \frac{1}{t - t_0} \, \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(- \frac{1}{2} \frac{|\nabla \ell^\beta(u, X(u))|^2}{\ell^\beta(u, X(u))^2} \right) \mathrm{d}u \right] \\ &\qquad = \lim_{t \downarrow t_0} \frac{1}{t - t_0} \, \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(- \frac{1}{2} \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} \right) \mathrm{d}u \right] \end{aligned} \tag{4.22}$$

follows from (4.19), and (4.18) implies that it is immaterial whether we take expectations with respect to $\mathbb{P}$ or $\mathbb{P}^\beta$ in the two limits appearing in (4.22). Summing up, existence and equality of the limits in (4.21) are guaranteed if and only if $t_0 \in \mathbb{R}_+ \setminus N$. It develops that both limits in (4.21) exist if $t_0 \geqslant 0$ is not in the exceptional set N of zero Lebesgue measure, and their common value is

$$-\tfrac{1}{2} \, I\big(P(t_0) \,|\, \mathrm{Q}\big) = -\tfrac{1}{2} \, \mathbb{E}_{\mathbb{P}} \left[\, \frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \, \right]. \tag{4.23}$$

In conjunction with (4.20), which is valid for every $t_0 \geqslant 0$, this establishes the limiting identity (4.16) for every $t_0 \in \mathbb{R}_+ \setminus N$. Therefore, the right-sided limiting assertion (3.6), and the similar perturbed limiting assertion in (4.16), fail on precisely the same set of exceptional points N.

As regards the final assertion we note that, by analogy with (4.14), the limiting behavior of the Wasserstein distance (4.17), for Lebesgue-a.e. $t_0 \geqslant 0$, is well known [AGS08]; details are in Section 5.5 below. More precisely, Theorem 5.11 establishes (with the choice $\beta \equiv 0$) the novel and very crucial aspect, that the limiting assertion

$$\lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0} = \tfrac{1}{2} \sqrt{I\big(P(t_0) \,|\, \mathrm{Q}\big)} \tag{4.24}$$

is valid for every $t_0 \in \mathbb{R}_+ \setminus N$. Once again, concerning the relation between the limits in (4.24) and (4.17) pertaining to the Wasserstein distance, we discern a similar pattern as in the case of the generalized de Bruijn identity. In fact, Theorem 5.11 will tell us that the perturbed Wasserstein limit (4.17) also holds for every $t_0 \in \mathbb{R}_+ \setminus N$. □

Proof of Theorem 3.3 from Theorems 4.1, 4.2 and Corollaries 4.3, 4.4:

Let $t_0 \in \mathbb{R}_+ \setminus N$, so that the limiting identities (4.16) and (4.17) from Corollary 4.4 are valid. Recalling the abbreviations in (3.10), we claim that the following limiting identities hold:

$$\lim_{t \downarrow t_0} \frac{H\big(P(t) \,|\, \mathrm{Q}\big) - H\big(P(t_0) \,|\, \mathrm{Q}\big)}{t - t_0} = -\tfrac{1}{2} \, \|a\|^2_{L^2(\mathbb{P})}, \tag{4.25}$$

$$\lim_{t \downarrow t_0} \frac{W_2\big(P(t), P(t_0)\big)}{t - t_0} = \tfrac{1}{2} \, \|a\|_{L^2(\mathbb{P})}, \tag{4.26}$$

$$\lim_{t \downarrow t_0} \frac{H\big(P^\beta(t) \,|\, \mathrm{Q}\big) - H\big(P^\beta(t_0) \,|\, \mathrm{Q}\big)}{t - t_0} = -\tfrac{1}{2} \, \big\langle a, a + 2b \big\rangle_{L^2(\mathbb{P})}, \tag{4.27}$$

$$\lim_{t \downarrow t_0} \frac{W_2\big(P^\beta(t), P^\beta(t_0)\big)}{t - t_0} = \tfrac{1}{2} \, \|a + 2b\|_{L^2(\mathbb{P})}. \tag{4.28}$$

Indeed, the equations (4.25), (4.26), and (4.28) correspond to (3.6), (4.24), and (4.17), respectively. As for (4.27), we note that, according to equation (4.16) of Corollary 4.4, the limit in (4.27) equals

$$-\tfrac{1}{2} \, \|a\|^2_{L^2(\mathbb{P})} + \mathbb{E}_{\mathbb{P}} \Big[\Big(\operatorname{div} \beta - 2 \, \langle \beta, \nabla \Psi \rangle \Big) \big(X(t_0) \big) \Big].$$

Therefore, in view of the right-hand side of (4.27), we have to show the identity

$$\mathbb{E}_{\mathbb{P}} \Big[\Big(\operatorname{div} \beta - \langle \beta \, , \, 2 \, \nabla \Psi \rangle \Big) \big(X(t_0) \big) \Big] = -\langle a, b \rangle_{L^2(\mathbb{P})}. \tag{4.29}$$

In order to do this, we write the left-hand side of (4.29) as

$$\int_{\mathbb{R}^n} \Big(\operatorname{div} \beta(x) - \big\langle \beta(x) \, , \, 2 \, \nabla \Psi(x) \big\rangle \Big) \, p(t_0, x) \, \mathrm{d}x. \tag{4.30}$$

Using — for the first time, and only in order to show the identity (4.29) — integration by parts, and the fact that the perturbation β is assumed to be smooth and compactly supported, we see that the expression (4.30) becomes

$$- \int_{\mathbb{R}^n} \Big\langle \beta(x) \, , \, \nabla \log p(t_0, x) + 2 \, \nabla \Psi(x) \Big\rangle \, p(t_0, x) \, \mathrm{d}x,$$

which is the same as

$$- \Big\langle \beta\big(X(t_0)\big) \, , \, \nabla \log \ell\big(t_0, X(t_0)\big) \Big\rangle_{L^2(\mathbb{P})} = -\langle b, a \rangle_{L^2(\mathbb{P})}.$$

The limiting identities (4.25) – (4.28) now imply the assertions of Theorem 3.3. □

The following Propositions 4.5 and 4.7 are trajectorial versions of Corollaries 4.3 and 4.4, respectively. They compute the rate of temporal change of relative entropy for the equation (1.3) and for its perturbed version (3.9), respectively, in the more precise trajectorial manner of Theorems 4.1, 4.2.

Proposition 4.5 (Trajectorial rate of relative entropy dissipation)**.**
Under the Assumptions 2.2, we let $t_0 \in \mathbb{R}_+ \setminus N$ and $T > t_0$. Then the relative entropy process (4.1) satisfies the trajectorial relation

$$\begin{aligned} \lim_{s \uparrow T - t_0} & \frac{\mathbb{E}_{\mathbb{P}}\Big[\log \ell(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell(T-s, X(T-s))}{T - t_0 - s} \\ &= \frac{1}{2} \frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2}, \end{aligned} \tag{4.31}$$

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

Remark 4.6. The limiting assertion (4.31) of Proposition 4.5 is the conditional trajectorial version of the generalized de Bruijn identity (4.13). ⋄

Proof of Proposition 4.5 from Theorem 4.1:
Let $t_0 \in \mathbb{R}_+ \setminus N$, i.e., so that the right-sided limiting assertion (3.6) is valid, and select $T > t_0$. The martingale property of the process in (4.4) allows us to write the numerator on the first line of (4.31) as

$$\mathbb{E}_{\mathbb{P}}\Big[F(t_0) - F(T-s) \mid \mathcal{G}(T-s)\Big], \qquad 0 \leqslant s \leqslant T - t_0$$

in the notation of (4.3), which expresses the Fisher information process $(F(T-s))_{0 \leqslant s \leqslant T}$ as the primitive of

$$B(u) = \frac{1}{2} \frac{\big|\nabla \ell(T-u, X(T-u))\big|^2}{\ell(T-u, X(T-u))^2}, \qquad 0 \leqslant u \leqslant T. \tag{4.32}$$

In other words,

$$F(T-s) = \int_0^s B(u)\, \mathrm{d}u, \qquad 0 \leqslant s \leqslant T - t_0.$$

By analogy with the derivation of (4.13) from (4.12), where we calculated real-valued expectations, we rely on the Lebesgue differentiation theorem to obtain the corresponding result (4.31) for conditional expectations. Using the left-continuity of the backwards filtration

$$\mathbb{G} = \big(\mathcal{G}(T-s)\big)_{0 \leqslant s \leqslant T},$$

we can invoke Proposition 6.3 of Section 6.3, with the choice of the process B as in (4.32) and $C \equiv 0$. This establishes the claim (4.31) pertaining to conditional expectations. □

Next we discuss the corresponding result in the perturbed setting: the limiting assertion (4.33) below is the perturbed analogue of (4.31). In (4.34) we consider the same limit, but under the probability measure $\mathbb{P}^\beta$ instead of $\mathbb{P}$. Additionally, we directly compare the relative entropy process (4.1) with its perturbed counterpart (4.8) in (4.35).

Proposition 4.7 (Trajectorial rate of relative entropy dissipation under perturbations)**.**
Under the Assumptions 2.2, we let $t_0 \in \mathbb{R}_+ \setminus N$ and $T > t_0$. Then the perturbed relative entropy process (4.8) satisfies the trajectorial relations

$$\begin{aligned}\lim_{s\uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}^\beta}\Big[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ = \frac{1}{2}\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2} - \operatorname{div}\beta(X(t_0)) + \Big\langle \beta(X(t_0))\,,\, 2\,\nabla\Psi(X(t_0))\Big\rangle,\end{aligned} \tag{4.33}$$

$$\begin{aligned}\lim_{s\uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}}\Big[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ = \frac{1}{2}\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2} - \operatorname{div}\beta(X(t_0)) - \Big\langle \beta(X(t_0))\,,\, \nabla \log p(t_0, X(t_0))\Big\rangle,\end{aligned} \tag{4.34}$$

$$\begin{aligned}\lim_{s\uparrow T-t_0} \frac{\log \ell^\beta(T-s, X(T-s)) - \log \ell(T-s, X(T-s))}{T-t_0-s} \\ = \operatorname{div}\beta(X(t_0)) + \Big\langle \beta(X(t_0))\,,\, \nabla \log p(t_0, X(t_0))\Big\rangle,\end{aligned} \tag{4.35}$$

where the limits in (4.33) – (4.35) exist in both $L^1(\mathbb{P})$ and $L^1(\mathbb{P}^\beta)$.

Remark 4.8. It is noteworthy that the three limiting expressions in (4.33), (4.34) and (4.35) are quite different from each other. The first limiting assertion (4.33) of Proposition 4.7 is the conditional trajectorial version of the perturbed de Bruijn identity (4.16). We also note that in fact the third limiting assertion (4.35) is valid for all $t_0 > 0$. ⋄

Proof of (4.33) from Theorem 4.2:
Let $t_0 \in \mathbb{R}_+ \setminus N$, i.e., so that the right-sided limiting assertion (3.6) is valid, and select $T > t_0$. In (4.21) from Corollary 4.4 of Theorem 4.2 we have seen that the limits in (3.6) and (4.16) have the same exceptional sets, hence the limiting identity (4.16) also holds. Now, for such $t_0 \in \mathbb{R}_+ \setminus N$, we show the limiting assertion (4.33) in the same way as the assertion (4.31) in the proof of Proposition 4.5 above. Indeed, this time we invoke the $\mathbb{P}^\beta$-martingale property of the process in (4.10), and write the numerator on the first line of (4.33) as

$$\mathbb{E}_{\mathbb{P}^\beta}\Big[F^\beta(t_0) - F^\beta(T-s) \mid \mathcal{G}(T-s)\Big], \qquad 0 \leqslant s \leqslant T-t_0$$

in the notation of (4.9), which expresses the process $(F^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ as the primitive

of $(B(u) + C(u))_{0 \leqslant s \leqslant T - t_0}$, with

$$\begin{aligned} B(u) &= \frac{1}{2} \frac{\left| \nabla \ell^\beta (T-u, X(T-u)) \right|^2}{\ell^\beta (T-u, X(T-u))^2}, \\ C(u) &= \Big(\langle \beta \,,\, 2 \nabla \Psi \rangle - \operatorname{div} \beta \Big) (X(T-u)). \end{aligned} \tag{4.36}$$

Applying Proposition 6.3 of Section 6.3 in this situation proves the limiting identity (4.33) in $L^1(\mathbb{P}^\beta)$. As we shall see in Lemma 5.7 of Section 5.3 below, the probability measures $\mathbb{P}$ and $\mathbb{P}^\beta$ are equivalent, and the mutual Radon–Nikodým derivatives $\frac{\mathrm{d}\mathbb{P}^\beta}{\mathrm{d}\mathbb{P}}$ and $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^\beta}$ are bounded on the σ-algebra $\mathcal{F}(T) = \mathcal{G}(0)$ (recall, in this vein, the claims of (4.18)). Hence, convergence in $L^1(\mathbb{P})$ is equivalent to convergence in $L^1(\mathbb{P}^\beta)$. This readily proves assertion (4.33). □

The proofs of the limiting assertions (4.34) and (4.35) are postponed to Section 5.4.

4.2. A trajectorial proof of the HWI inequality

The aim of this section is to provide a proof of the celebrated *HWI inequality* due to Otto and Villani [OV00] by applying trajectorial arguments similar to those in Theorem 4.1, in fact quite easier. We thus obtain an intuitive geometric picture and deduce the sharpened form of the HWI inequality; see also [CE02], [OV00] and [Vil09, p. 650]).

The goal is to compare the relative entropies $H(P_0 \,|\, \mathrm{Q})$ and $H(P_1 \,|\, \mathrm{Q})$ for arbitrary probability measures $P_0, P_1 \in \mathscr{P}_2(\mathbb{R}^n)$. Using *Brenier's theorem* [Bre91], we first define the constant speed geodesic $(P_t)_{0 \leqslant t \leqslant 1}$ between P_0 and P_1 with respect to the Wasserstein distance W_2 (details are given below). We remark, that we have chosen the subscript notation for P_t in order to avoid confusion with the probability measure $P(t)$ from our Chapter 2 here. With $p_t(\,\cdot\,)$ the density function of the probability measure P_t, we define the likelihood ratio function

$$\ell_t(x) := \frac{p_t(x)}{q(x)}, \qquad (t,x) \in [0,1] \times \mathbb{R}^n. \tag{4.37}$$

We shall investigate the behavior of the relative entropy function $t \mapsto f(t) := H(P_t \,|\, \mathrm{Q})$ along the constant speed geodesic $(P_t)_{0 \leqslant t \leqslant 1}$ by estimating two quantities: First, we want a lower bound on the first derivative $f'(0^+)$. Secondly, we want a lower bound on the second derivative $(f''(t))_{0 \leqslant t \leqslant 1}$. It should be geometrically obvious (and will be spelled out in the proof of Theorem 4.11 below) that information on these two lower bounds leads to a lower bound on $f(1) - f(0)$. The latter is the content of the HWI inequality. As regards the second derivative $(f''(t))_{0 \leqslant t \leqslant 1}$, we shall rely on a fundamental result on displacement convexity due to McCann [McC97] and have no novel contribution. As regards $f'(0^+)$, however, we shall obtain a sharp estimate for this quantity by applying a trajectorial reasoning similar to that deployed in the proof of Theorem 4.1.

We will define an $\mathbb{R}^n$-valued stochastic process $(X_t)_{0\leqslant t\leqslant 1}$, with marginal distributions $(P_t)_{0\leqslant t\leqslant 1}$ moving along straight lines in $\mathbb{R}^n$, and calculate the relevant quantities of this finite variation process along every trajectory, by analogy with the proof of Theorem 4.1. This gives the desired bound (and actually an equality) for the derivative $f'(0^+)$.

We now cast these ideas into formal terms. The first step is to calculate the decay of the relative entropy function $t \mapsto H(P_t \,|\, \mathrm{Q})$ along the "straight line" $(P_t)_{0\leqslant t\leqslant 1}$ joining the elements P_0 and P_1 in $\mathscr{P}_2(\mathbb{R}^n)$. To this end, we impose temporarily the following strong regularity conditions. In the proof of Theorem 4.11 we shall see that these will not restrict the generality of the argument.

Assumptions 4.9 (Regularity assumptions of Lemma 4.10)**.**
We impose that P_0 and P_1 are probability measures in $\mathscr{P}_2(\mathbb{R}^n)$ with smooth densities, which are compactly supported and strictly positive in the interior of their respective supports. Hence there exists a map $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$\gamma(x) = \nabla\big(G(x) - \tfrac{|x|^2}{2}\big)$$

for some convex function $G\colon \mathbb{R}^n \to \mathbb{R}$, uniquely defined on and supported by the support of P_0, and smooth in the interior of this set, such that γ induces the optimal quadratic Wasserstein transport from P_0 to P_1 via

$$\begin{aligned} T_t^\gamma(x) &:= x + t\,\gamma(x) = (1-t)\,x + t\,\nabla G(x), \\ P_t &:= (T_t^\gamma)_\#(P_0) = P_0 \circ (T_t^\gamma)^{-1} \end{aligned} \tag{4.38}$$

for $0 \leqslant t \leqslant 1$; to wit, the curve $(P_t)_{0\leqslant t\leqslant 1}$ is the displacement interpolation (constant speed geodesic) between P_0 and P_1, and we have along it the linear growth of the quadratic Wasserstein distance

$$W_2(P_0, P_t) = t\sqrt{\int_{\mathbb{R}^n} |x - \nabla G(x)|^2 \,\mathrm{d}P_0(x)} = t\,\|\gamma\|_{L^2(P_0)}, \qquad 0 \leqslant t \leqslant 1. \tag{4.39}$$

For existence and uniqueness of the optimal transport map γ we refer to [Vil03, Theorem 2.12], and for its smoothness to [Vil03, Theorem 4.14] as well as [Vil03, Remarks 4.15]. These results are known collectively under the rubric of Brenier's theorem [Bre91].

Next we compute the slope of the relative entropy function $t \mapsto H(P_t \,|\, \mathrm{Q})$ along the straight line $(P_t)_{0\leqslant t\leqslant 1}$.

Lemma 4.10. *Under the Assumptions 4.9, let $X_0\colon S \to \mathbb{R}^n$ be a random variable with probability distribution $P_0 \in \mathscr{P}_2(\mathbb{R}^n)$, defined on some probability space $(S, \mathcal{S}, \nu)$. Then we have*

$$\lim_{t\downarrow 0} \frac{H(P_t \,|\, \mathrm{Q}) - H(P_0 \,|\, \mathrm{Q})}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)}. \tag{4.40}$$

We relegate the proof of Lemma 4.10 to Section 6.4, which follows a similar (but considerably simpler) trajectorial line of reasoning as the proof of Theorem 3.3. Combining

Lemma 4.10 with well-known arguments, in particular, with a fundamental result on displacement convexity due to McCann [McC97], we derive now the HWI inequality of Otto and Villani [OV00].

Theorem 4.11 (HWI inequality [OV00]). *We fix probability measures $P_0, P_1 \in \mathscr{P}_2(\mathbb{R}^n)$ and assume that the relative entropy $H(P_1 \,|\, \mathrm{Q})$ is finite. We suppose in addition that the potential $\Psi \in \mathcal{C}^\infty(\mathbb{R}^n; [0,\infty))$ satisfies a curvature lower bound*

$$\mathrm{Hess}(\Psi) \geqslant \kappa\, I_n, \tag{4.41}$$

for some $\kappa \in \mathbb{R}$. Then we have

$$H(P_0 \,|\, \mathrm{Q}) - H(P_1 \,|\, \mathrm{Q}) \leqslant -\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1), \tag{4.42}$$

where the likelihood ratio function ℓ_0, the random variable X_0, the optimal transport map γ, and the probability measure ν, are as in Lemma 4.10.

We stress that Theorem 4.11 does not require the reference measure Q with density $q(x) = \mathrm{e}^{-2\Psi(x)}$ to be a finite measure in the formulation of the HWI inequality (4.42).

On the strength of the Cauchy–Schwarz inequality, we have

$$-\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} \leqslant \|\nabla \log \ell_0(X_0)\|_{L^2(\nu)}\; \|\gamma(X_0)\|_{L^2(\nu)},$$

with equality if and only if the functions $\nabla \log \ell_0(\,\cdot\,)$ and $\gamma(\,\cdot\,)$ are negatively collinear. The relative Fisher information of P_0 with respect to Q equals

$$I(P_0 \,|\, \mathrm{Q}) = \mathbb{E}_\nu\Big[|\nabla \log \ell_0(X_0)|^2 \Big] = \|\nabla \log \ell_0(X_0)\|^2_{L^2(\nu)},$$

and by Brenier's theorem [Vil03, Theorem 2.12] we deduce

$$\|\gamma(X_0)\|_{L^2(\nu)} = W_2(P_0, P_1)$$

as in (4.39), along with the inequality

$$-\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} \leqslant \sqrt{I(P_0 \,|\, \mathrm{Q})}\; W_2(P_0, P_1). \tag{4.43}$$

Inserting (4.43) into (4.42) we obtain the usual form of the HWI inequality

$$H(P_0 \,|\, \mathrm{Q}) - H(P_1 \,|\, \mathrm{Q}) \leqslant W_2(P_0, P_1)\, \sqrt{I(P_0 \,|\, \mathrm{Q})} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1). \tag{4.44}$$

When there is a nontrivial angle between $-\nabla \log \ell_0(X_0)$ and $\gamma(X_0)$ in $L^2(\nu)$, the inequality (4.42) gives a sharper bound than (4.44). We refer to the original paper [OV00], as well as to [CE02], [Vil03, Chapter 5], [Vil09, p. 650] and the recent papers [GLRT20, KMS21] for detailed discussions of the HWI inequality in several contexts. For a good survey on transport inequalities, see [GL10].

Proof of Theorem 4.11. As elaborated in [Vil03, Section 9.4] we may assume without loss of generality that P_0 and P_1 satisfy the strong regularity Assumptions 4.9, guaranteeing existence and smoothness of the optimal transport map γ.

We consider now the relative entropy with respect to Q along the constant-speed geodesic $(P_t)_{0\leqslant t\leqslant 1}$, namely, the function $f(t) := H(P_t \,|\, \mathrm{Q})$, for $0 \leqslant t \leqslant 1$. The displacement convexity results of McCann [McC97] imply

$$f''(t) \geqslant \kappa\, W_2^2(P_0, P_1), \qquad 0 \leqslant t \leqslant 1. \tag{4.45}$$

Indeed, under the condition (4.41), the potential Ψ is κ-uniformly convex. Consequently, by items (i) and (ii) of [Vil03, Theorem 5.15], the internal and potential energies

$$g(t) := \int_{\mathbb{R}^n} p_t(x) \log p_t(x)\, \mathrm{d}x, \qquad h(t) := 2\int_{\mathbb{R}^n} \Psi(x)\, p_t(x)\, \mathrm{d}x, \qquad 0 \leqslant t \leqslant 1,$$

are displacement convex and κ-uniformly displacement convex, respectively; i.e.,

$$g''(t) \geqslant 0, \qquad h''(t) \geqslant \kappa\, W_2^2(P_0, P_1), \qquad 0 \leqslant t \leqslant 1.$$

As we have $f = g + h$, we conclude that the relative entropy function f is κ-uniformly displacement convex, i.e., its second derivative satisfies (4.45). We appeal now to Lemma 4.10, according to which

$$f'(0^+) = \lim_{t\downarrow 0} \frac{f(t) - f(0)}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)}. \tag{4.46}$$

In conjunction with (4.45) and (4.46), the Taylor formula

$$f(1) = f(0) + f'(0^+) + \int_0^1 (1-t) f''(t)\, \mathrm{d}t$$

now yields (4.42). □

5. Details and proofs of the main results

In this chapter we complete the proofs of Corollary 4.4 and Proposition 4.7, and provide the proofs of the crucial results, Theorems 4.1 and 4.2. For this, we apply Itô's formula so as to calculate the dynamics, i.e., the stochastic differentials, of the "pure" and "perturbed" relative entropy processes of (4.1) and (4.8) under the measures $\mathbb{P}$ and $\mathbb{P}^\beta$, respectively. As already discussed, we shall do this in the backward direction of time. Finally, the limiting behavior of the quadratic Wasserstein distance along the Langevin–Smoluchowski diffusion is analyzed in Section 5.5.

5.1. The proof of Theorem 4.1

We start by calculating the stochastic differential of the time-reversed canonical coordinate process $(X(T-s))_{0\leqslant s\leqslant T}$ under the probability measure $\mathbb{P}$, a well-known and classical theme; see e.g. [Föl85, Föl86], [HP86], [Mey94], [Nel01], and [Par86]. The reader may consult Part V of this thesis for an extensive presentation of the relevant facts regarding the theory of time reversal for diffusion processes. The idea of time reversal goes back to Boltzmann [Bol96, Bol98a, Bol98b] and Schrödinger [Sch31, Sch32], as well as Kolmogorov [Kol37]. In fact, the relation between time reversal of a Brownian motion and the quadratic Wasserstein distance may *in nuce* be traced back to an insight of Bachelier in his thesis [Bac00, Bac06] from 1900. This theme is discussed in Appendix A of [KST20].

Recall that the probability measure $\mathbb{P}$ was defined on path space $\Omega = \mathcal{C}(\mathbb{R}_+;\mathbb{R}^n)$ so that the canonical coordinate process $(X(t,\omega))_{t\geqslant 0} = (\omega(t))_{t\geqslant 0}$ satisfies the stochastic differential equation (1.3) with initial probability distribution $P(0)$ for $X(0)$ under $\mathbb{P}$. In other words, the process

$$W(t) = X(t) - X(0) + \int_0^t \nabla\Psi(X(u))\,\mathrm{d}u, \qquad t \geqslant 0 \tag{5.1}$$

is a Brownian motion of the forward filtration $(\mathcal{F}(t))_{t\geqslant 0}$ under the probability measure $\mathbb{P}$. Passing to the reverse direction of time, the following classical result is well known to hold under the present assumptions. For proof and references we refer to Theorems 4.1 and 5.1 of Part V.

Proposition 5.1. *Under the Assumptions 2.2, fix $T > 0$. The process*

$$\overline{W}^{\mathbb{P}}(T-s) := W(T-s) - W(T) - \int_0^s \nabla \log p(T-u, X(T-u))\,\mathrm{d}u, \tag{5.2}$$

for $0 \leqslant s \leqslant T$, is a Brownian motion of the backwards filtration

$$\mathbb{G} = (\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$$

under the probability measure $\mathbb{P}$*. Moreover, the time-reversed canonical coordinate process* $(X(T-s))_{0\leqslant s\leqslant T}$ *satisfies the stochastic differential equation*

$$\mathrm{d}X(T-s) = \Big(\nabla \log p(T-s, X(T-s)) + \nabla\Psi(X(T-s))\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) \tag{5.3}$$

$$= \Big(\nabla \log \ell(T-s, X(T-s)) - \nabla\Psi(X(T-s))\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}}(T-s), \tag{5.4}$$

for $0 \leqslant s \leqslant T$*, with respect to the backwards filtration* $\mathbb{G}$*.*

The following result computes the forward dynamics of the likelihood ratio process $(\ell(t, X(t)))_{t\geqslant 0}$ of (2.1) and compares it with the stochastic differential of the time-reversed likelihood ratio process

$$\ell(T-s, X(T-s)) = \frac{p(T-s, X(T-s))}{q(X(T-s))}, \qquad 0 \leqslant s \leqslant T, \tag{5.5}$$

as well as its logarithmic differential.

Proposition 5.2. *Under the* Assumptions 2.2*, the likelihood ratio process* (2.1) *is a continuous semimartingale with respect to the forward filtration* $(\mathcal{F}(t))_{t\geqslant 0}$ *and satisfies, for* $t \geqslant 0$*, the stochastic differential equation*

$$\mathrm{d}\ell(t, X(t)) = \Big\langle \nabla\ell(t, X(t))\,,\, \mathrm{d}W(t)\Big\rangle + \Big(\Delta\ell(t, X(t)) - \Big\langle \nabla\ell(t, X(t))\,,\, 2\,\nabla\Psi(X(t))\Big\rangle\Big)\,\mathrm{d}t. \tag{5.6}$$

Furthermore, the time-reversed likelihood ratio process (5.5) *is a continuous semimartingale with respect to the backwards filtration* $\mathbb{G}$ *and satisfies, for* $0 \leqslant s \leqslant T$*, the stochastic differential equations*

$$\begin{aligned}\mathrm{d}\ell(T-s, X(T-s)) &= \Big\langle \nabla\ell(T-s, X(T-s))\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s)\Big\rangle \\ &\quad + \frac{\big|\nabla\ell(T-s, X(T-s))\big|^2}{\ell(T-s, X(T-s))}\,\mathrm{d}s,\end{aligned} \tag{5.7}$$

$$\begin{aligned}\frac{\mathrm{d}\ell(T-s, X(T-s))}{\ell(T-s, X(T-s))} &= \left\langle \frac{\nabla\ell(T-s, X(T-s))}{\ell(T-s, X(T-s))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s)\right\rangle \\ &\quad + \frac{\big|\nabla\ell(T-s, X(T-s))\big|^2}{\ell(T-s, X(T-s))^2}\,\mathrm{d}s,\end{aligned} \tag{5.8}$$

$$\begin{aligned}\mathrm{d}\log\ell(T-s, X(T-s)) &= \left\langle \frac{\nabla\ell(T-s, X(T-s))}{\ell(T-s, X(T-s))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s)\right\rangle \\ &\quad + \frac{1}{2}\frac{\big|\nabla\ell(T-s, X(T-s))\big|^2}{\ell(T-s, X(T-s))^2}\,\mathrm{d}s.\end{aligned} \tag{5.9}$$

Proof. We start with the following observation. Writing the Fokker–Planck equation (1.1) as

$$\partial_t p(t,x) = \tfrac{1}{2}\Delta p(t,x) + \langle \nabla p(t,x)\,, \nabla\Psi(x)\rangle + p(t,x)\,\Delta\Psi(x), \qquad t>0$$

and substituting the expression

$$p(t,x) = \ell(t,x)\,q(x) = \ell(t,x)\,\mathrm{e}^{-2\Psi(x)}, \qquad t \geqslant 0$$

into this equation, we find that the likelihood ratio function $(t,x) \mapsto \ell(t,x)$ solves the *backwards* Kolmogorov equation

$$\partial_t \ell(t,x) = \tfrac{1}{2}\Delta\ell(t,x) - \langle \nabla\ell(t,x)\,, \nabla\Psi(x)\rangle, \qquad t>0. \tag{5.10}$$

Now we turn to the proofs of (5.6) – (5.9). By Assumptions 2.2, the likelihood ratio function $(t,x) \mapsto \ell(t,x)$ is sufficiently smooth to allow an application of Itô's formula. Together with the Langevin–Smoluchowski dynamics (1.3) and the backwards Kolmogorov equation (5.10), we obtain (5.6) by direct calculation. A similar calculation, this time relying on the backwards dynamics (5.4), shows (5.7). Finally, the equations (5.8) and (5.9) follow from (5.7) and Itô's formula. □

The crucial feature of the stochastic differentials (5.6) – (5.9) is that, after passing to time reversal, the finite-variation term $\Delta\ell - \langle\nabla\ell\,,\,2\,\nabla\Psi\rangle$ in (5.6), involving the Laplacian $\Delta\ell$, gets replaced by a term involving only the likelihood ratio function ℓ and its gradient $\nabla\ell$.

Proof of Theorem 4.1. On a formal level, the expressions (4.3) – (4.5) are just integral versions of the Itô differential (5.9). What remains to check is that the integrals in (4.3) and (4.5) indeed make rigorous sense and satisfy the claimed integrability conditions.

By condition (iv) of Assumptions 2.2 the function $(t,x) \mapsto \nabla \log \ell(t,x)$ is continuous. Together with the continuity of the paths of the canonical coordinate process $(X(t))_{t\geqslant 0}$, this implies

$$\int_0^{T-\varepsilon} \frac{\big|\nabla\ell\big(T-u, X(T-u)\big)\big|^2}{\ell\big(T-u, X(T-u)\big)^2}\,\mathrm{d}u < \infty, \qquad \mathbb{P}\text{-a.s.} \tag{5.11}$$

for every $0<\varepsilon\leqslant T$. On account of (5.11), the sequence of stopping times (with respect to the backwards filtration $\mathbb{G}$)

$$\tau_n := \inf\left\{ t\geqslant 0 \colon \int_0^t \frac{\big|\nabla\ell\big(T-u, X(T-u)\big)\big|^2}{\ell\big(T-u, X(T-u)\big)^2}\,\mathrm{d}u \geqslant n \right\} \wedge T, \qquad n\in\mathbb{N}_0$$

is nondecreasing and converges $\mathbb{P}$-a.s. to T. Defining M via (4.4), each stopped process M^{τ_n} is bounded in $L^2(\mathbb{P})$ and satisfies the stopped version of (4.5), i.e.,

$$M^{\tau_n}(T-s) = M\big(T-(s\wedge\tau_n)\big) = \int_0^{s\wedge\tau_n} \left\langle \frac{\nabla\ell\big(T-u, X(T-u)\big)}{\ell\big(T-u, X(T-u)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u)\right\rangle,$$

for $0 \leqslant s \leqslant T$. To show that, in fact, the process M is a true $\mathbb{P}$-martingale, we have to rely on condition (2.6), which asserts that the initial relative entropy $H(P(0) \,|\, \mathrm{Q})$ is finite.

We consider the process

$$\ell^{-1}(T-s, X(T-s)) = \frac{q(X(T-s))}{p(T-s, X(T-s))}, \qquad 0 \leqslant s \leqslant T, \tag{5.12}$$

where $\ell^{-1}(t, \,\cdot\,) = \frac{1}{\ell(t,\,\cdot\,)}$ is the likelihood ratio function of $\frac{\mathrm{dQ}}{\mathrm{d}P(t)}(\,\cdot\,)$. Applying Itô's formula and using (5.7), we find the stochastic differential

$$\mathrm{d}\ell^{-1}(T-s, X(T-s)) = -\left\langle \frac{\nabla \ell(T-s, X(T-s))}{\ell(T-s, X(T-s))^2}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) \right\rangle, \tag{5.13}$$

revealing that the locally bounded process (5.12) is a local martingale under $\mathbb{P}$. In fact, this result does not come as a surprise: it is a consequence of an eye-opening result by Pavon [Pav89], and Fontbona and Jourdain [FJ16], at least when Q is a finite measure. We refer to Subsection 4.2 of [KST20] for more information on this theme, and for a more direct proof of Theorem 4.1 in the case when Q is a finite measure on $\mathbb{R}^n$; see also Theorem 4.2 and Appendix E in [KST20] for an extensive discussion and a proof of the Pavon–Fontbona–Jourdain theorem.

From (5.13), we deduce the stochastic differential of the logarithm of the process (5.12) and obtain in accordance with (5.9) its form

$$\begin{aligned}\mathrm{d}\log \ell^{-1}(T-s, X(T-s)) = &-\left\langle \frac{\nabla \ell(T-s, X(T-s))}{\ell(T-s, X(T-s))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) \right\rangle \\ &- \frac{1}{2}\frac{\big|\nabla \ell(T-s, X(T-s))\big|^2}{\ell(T-s, X(T-s))^2}\,\mathrm{d}s.\end{aligned}$$

We know that the terminal value $\log \ell^{-1}(0, X(0))$ is $\mathbb{P}$-integrable, with

$$\mathbb{E}_{\mathbb{P}}\big[\log \ell^{-1}(0, X(0))\big] = -H(P(0) \,|\, \mathrm{Q}) \in (-\infty, \infty).$$

On the other hand, the initial value

$$\mathbb{E}_{\mathbb{P}}\big[\log \ell^{-1}(T, X(T))\big] = -H(P(T) \,|\, \mathrm{Q}) \in [-\infty, \infty)$$

cannot take the value ∞, as mentioned after the definition (2.3) of relative entropy. Hence we can apply Proposition 6.4 in Section 6.3 to the local martingale (5.12) (in the reverse direction of time) and the deterministic stopping time $\tau = T$, to conclude that

$$\mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{\ell^{-1}(0, X(0))}{\ell^{-1}(T, X(T))}\right)\right] = -\mathbb{E}_{\mathbb{P}}\left[\int_0^T \frac{1}{2}\frac{\big|\nabla \ell(T-u, X(T-u))\big|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u\right],$$

where all terms are well-defined and finite. This shows that the local martingale M is bounded in $L^2(\mathbb{P})$, with

$$\tfrac{1}{2}\|M(0)\|^2_{L^2(\mathbb{P})} = H\big(P(0)\,|\,\mathrm{Q}\big) - H\big(P(T)\,|\,\mathrm{Q}\big) = \tfrac{1}{2}\int_0^T I\big(P(t)\,|\,\mathrm{Q}\big)\,\mathrm{d}t < \infty,$$

completing the proof of Theorem 4.1. □

5.2. The proof of Theorem 4.2

The first step in the proof of Theorem 4.2 is to compute the stochastic differentials of the time-reversed perturbed likelihood ratio process

$$\ell^\beta\big(T-s, X(T-s)\big) = \frac{p^\beta\big(T-s, X(T-s)\big)}{q\big(X(T-s)\big)}, \qquad 0 \leqslant s \leqslant T - t_0, \tag{5.14}$$

and its logarithm. By analogy with Proposition 5.1, the following result is well known (recall Theorems 4.1 and 5.1 of Part V) to hold under suitable regularity conditions, such as Assumptions 2.2. Recall that $(W^\beta(t))_{t\geqslant t_0}$ denotes the $\mathbb{P}^\beta$-Brownian motion (in the forward direction of time) defined in (3.9).

Proposition 5.3. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. The process*

$$\overline{W}^{\mathbb{P}^\beta}(T-s) := W^\beta(T-s) - W^\beta(T) - \int_0^s \nabla \log p^\beta\big(T-u, X(T-u)\big)\,\mathrm{d}u, \tag{5.15}$$

for $0 \leqslant s \leqslant T - t_0$, is a Brownian motion of the backwards filtration

$$\widetilde{\mathbb{G}} = \big(\mathcal{G}(T-s)\big)_{0\leqslant s\leqslant T-t_0}$$

under the probability measure $\mathbb{P}^\beta$. Furthermore, the semimartingale decomposition of the time-reversed canonical coordinate process $(X(T-s))_{0\leqslant s\leqslant T-t_0}$ is given by

$$\mathrm{d}X(T-s) = \Big(\nabla \log p^\beta\big(T-s, X(T-s)\big) + \big(\nabla\Psi + \beta\big)\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s) \tag{5.16}$$

$$= \Big(\nabla \log \ell^\beta\big(T-s, X(T-s)\big) - \big(\nabla\Psi - \beta\big)\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s), \tag{5.17}$$

for $0 \leqslant s \leqslant T - t_0$, with respect to the backwards filtration $\widetilde{\mathbb{G}}$.

Comparing the equation (5.3) with (5.16), we see that the reverse-time Brownian motions $\overline{W}^{\mathbb{P}^\beta}$ and $\overline{W}^{\mathbb{P}}$ are related as follows.

Lemma 5.4. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. For $0 \leqslant s \leqslant T - t_0$, we have*

$$\mathrm{d}(\overline{W}^{\mathbb{P}} - \overline{W}^{\mathbb{P}^\beta})(T-s) = \left(\beta\big(X(T-s)\big) + \nabla \log\left(\frac{p^\beta\big(T-s, X(T-s)\big)}{p\big(T-s, X(T-s)\big)}\right)\right)\mathrm{d}s. \tag{5.18}$$

By analogy with the proof of Proposition 5.2, we write now the perturbed Fokker–Planck equation (3.7) as

$$\partial_t p^\beta(t,x) = \tfrac{1}{2}\Delta p^\beta(t,x) + \langle \nabla p^\beta(t,x)\,,\nabla\Psi(x) + \beta(x)\rangle + p^\beta(t,x)\,(\Delta\Psi(x) + \operatorname{div}\beta(x)),$$

for $t > t_0$. Using the relation

$$p^\beta(t,x) = \ell^\beta(t,x)\,q(x) = \ell^\beta(t,x)\,\mathrm{e}^{-2\Psi(x)}, \qquad t \geqslant t_0,$$

determined computation shows that the perturbed likelihood ratio function $\ell^\beta(t,x)$ satisfies the partial differential equation

$$\begin{aligned}\partial_t \ell^\beta(t,x) = \tfrac{1}{2}\Delta\ell^\beta(t,x) + \langle \nabla\ell^\beta(t,x)\,,\,\beta(x) - \nabla\Psi(x)\rangle \\ + \ell^\beta(t,x)\,\Big(\operatorname{div}\beta(x) - \langle\beta(x)\,,\,2\,\nabla\Psi(x)\rangle\Big),\end{aligned} \tag{5.19}$$

for $t > t_0$; this is the analogue of the backwards Kolmogorov equation (5.10) in this "perturbed" context, and reduces to (5.10) when $\beta \equiv 0$.

With these preparations, we obtain the following stochastic differentials for our objects of interest.

Lemma 5.5. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. The time-reversed perturbed likelihood ratio process (5.14) and its logarithm satisfy the stochastic differential equations*

$$\begin{aligned}\frac{\mathrm{d}\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))} &= \Big(\langle\beta\,,\,2\,\nabla\Psi\rangle - \operatorname{div}\beta\Big)(X(T-s))\,\mathrm{d}s \\ &+ \frac{\big|\nabla\ell^\beta(T-s,X(T-s))\big|^2}{\ell^\beta(T-s,X(T-s))^2}\,\mathrm{d}s + \left\langle \frac{\nabla\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))}\,,\,\mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s)\right\rangle\end{aligned} \tag{5.20}$$

and

$$\begin{aligned}\mathrm{d}\log\ell^\beta(T-s,X(T-s)) &= \Big(\langle\beta\,,\,2\,\nabla\Psi\rangle - \operatorname{div}\beta\Big)(X(T-s))\,\mathrm{d}s \\ &+ \frac{1}{2}\frac{\big|\nabla\ell^\beta(T-s,X(T-s))\big|^2}{\ell^\beta(T-s,X(T-s))^2}\,\mathrm{d}s + \left\langle \frac{\nabla\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))}\,,\,\mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s)\right\rangle,\end{aligned} \tag{5.21}$$

respectively, for $0 \leqslant s \leqslant T - t_0$, with respect to the backwards filtration $\widetilde{\mathbb{G}}$.

Proof. The stochastic differential equations (5.20), (5.21) follow from Itô's formula together with (5.17), (5.19). □

We have assembled now all the ingredients needed for the proof of Theorem 4.2.

Proof of Theorem 4.2. Formally, the stochastic differential in (5.21) amounts to the conclusions (4.9) – (4.11) of Theorem 4.2. But as in the proof of Theorem 4.1, we still have to substantiate the claim, that the stochastic process M^β defined in (4.10) with representation (4.11) is indeed a $\mathbb{P}^\beta$-martingale of the backwards filtration $\widetilde{\mathbb{G}}$, and is bounded in $L^2(\mathbb{P}^\beta)$.

By (5.21) and the same stopping argument as in the proof of Theorem 4.1, the process M^β is a local $\mathbb{P}^\beta$-martingale. We have to show that $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0)\big] < \infty$.

We recall that $\beta = \nabla B$ and define the density

$$q^\beta(x) := \mathrm{e}^{-2(\Psi+B)(x)}, \qquad x \in \mathbb{R}^n.$$

This density function solves the stationary version of the perturbed Fokker–Planck equation (3.7). Equivalently, it induces an invariant measure for the stochastic differential equation (3.9). We now consider the "doubly perturbed" likelihood ratio function

$$\ell^\beta_\beta(t,x) := \frac{p^\beta(t,x)}{q^\beta(x)}, \qquad (t,x) \in [t_0,\infty) \times \mathbb{R}^n.$$

Since the Assumptions 2.2 are invariant under the passage from the potential Ψ to $\Psi + B$, we can apply Theorem 4.1 to the potential $\Psi + B$ and obtain that the process (cf. (4.3))

$$F^\beta_\beta(T-s) := \int_0^s \frac{1}{2} \frac{\big|\nabla \ell^\beta_\beta(T-u, X(T-u))\big|^2}{\ell^\beta_\beta(T-u, X(T-u))^2}\,\mathrm{d}u, \qquad 0 \leqslant s \leqslant T - t_0 \tag{5.22}$$

satisfies $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta_\beta(t_0)\big] < \infty$. This latter condition implies also $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0)\big] < \infty$, where the process F^β is defined in (4.9). Indeed, the function $\langle \beta\,,\, 2\,\nabla\Psi\rangle - \operatorname{div}\beta$ in (4.9) is bounded, so that

$$\mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{T-t_0} \Big|\langle \beta\,,\, 2\,\nabla\Psi\rangle - \operatorname{div}\beta\Big|\big(X(T-u)\big)\,\mathrm{d}u\right] < \infty.$$

As regards the remaining difference between (5.22) and (4.9), note that

$$\frac{\ell^\beta_\beta(t,x)}{\ell^\beta(t,x)} = \mathrm{e}^{2B(x)}$$

and consequently $\nabla \log \ell^\beta_\beta(t,x) - \nabla \log \ell^\beta(t,x) = 2\,\nabla B(x)$, which again is a bounded function.

In conclusion, we obtain that $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0)\big] < \infty$, implying that the process M^β is a martingale bounded in $L^2(\mathbb{P}^\beta)$. This finishes the proof of Theorem 4.2. □

5.3. Some useful lemmas

In this section we collect some useful results needed in order to justify the claims (4.18), (4.19) made in the course of the proof of Corollary 4.4, and to complete the proof of Proposition 4.7 in Section 5.4.

First, let us introduce the "perturbed-to-unperturbed" ratio

$$Y^\beta(t,x) := \frac{\ell^\beta(t,x)}{\ell(t,x)} = \frac{p^\beta(t,x)}{p(t,x)}, \qquad (t,x) \in [t_0,\infty)\times\mathbb{R}^n \tag{5.23}$$

and recall the backwards Kolmogorov-type equations (5.10), (5.19). These lead to the partial differential equation

$$\begin{aligned}\partial_t Y^\beta(t,x) = \tfrac{1}{2}\Delta Y^\beta(t,x) + \big\langle \nabla Y^\beta(t,x)\,,\, \beta(x) + \nabla\log p(t,x) + \nabla\Psi(x)\big\rangle \\ + Y^\beta(t,x)\Big(\operatorname{div}\beta(x) + \big\langle\beta(x)\,,\,\nabla\log p(t,x)\big\rangle\Big),\end{aligned}$$

for $t > t_0$, and with initial condition $Y^\beta(t_0,\,\cdot\,) = 1$, for the ratio in (5.23). In conjunction with (5.3), this equation yields by direct calculation the following backward dynamics.

Lemma 5.6. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. The time-reversed ratio process*

$$R^\beta(T-s) := Y^\beta\big(T-s, X(T-s)\big), \qquad 0 \leqslant s \leqslant T-t_0 \tag{5.24}$$

and its logarithm satisfy the stochastic differential equations

$$\begin{aligned}\frac{\mathrm{d}R^\beta(T-s)}{R^\beta(T-s)} = \left\langle \frac{\nabla R^\beta(T-s)}{R^\beta(T-s)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta\big(X(T-s)\big)\,\mathrm{d}s\right\rangle \\ - \Big(\operatorname{div}\beta\big(X(T-s)\big) + \Big\langle\beta\big(X(T-s)\big)\,,\nabla\log p\big(T-s,X(T-s)\big)\Big\rangle\Big)\,\mathrm{d}s\end{aligned} \tag{5.25}$$

and

$$\begin{aligned}\mathrm{d}\log R^\beta(T-s) = \left\langle \frac{\nabla R^\beta(T-s)}{R^\beta(T-s)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta\big(X(T-s)\big)\,\mathrm{d}s\right\rangle \\ - \Big(\operatorname{div}\beta\big(X(T-s)\big) + \Big\langle\beta\big(X(T-s)\big)\,,\nabla\log p\big(T-s,X(T-s)\big)\Big\rangle\Big)\,\mathrm{d}s \\ - \frac{1}{2}\frac{\big|\nabla R^\beta(T-s)\big|^2}{R^\beta(T-s)^2}\,\mathrm{d}s,\end{aligned} \tag{5.26}$$

respectively, for $0 \leqslant s \leqslant T-t_0$, relative to the backwards filtration $\widetilde{\mathbb{G}}$.

We first establish a preliminary control on $Y^\beta(\,\cdot\,,\,\cdot\,)$, which will be refined in Lemma 5.8 below.

Lemma 5.7. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. There is a real constant $C > 1$ such that*

$$\frac{1}{C} \leqslant Y^\beta(t,x) \leqslant C, \qquad (t,x) \in [t_0,T] \times \mathbb{R}^n. \tag{5.27}$$

Proof. In the forward direction of time, the canonical coordinate process $(X(t))_{t_0 \leqslant t \leqslant T}$ on path space $\Omega = \mathcal{C}([t_0,T];\mathbb{R}^n)$ satisfies the equations (1.3) and (3.9) with initial distribution $P(t_0)$ under the probability measures $\mathbb{P}$ and $\mathbb{P}^\beta$, respectively. Hence, the $\mathbb{P}$-Brownian motion $(W(t))_{t_0 \leqslant t \leqslant T}$ from (1.3) can be represented as

$$W(t) - W(t_0) = W^\beta(t) - W^\beta(t_0) - \int_{t_0}^t \beta(X(u))\,\mathrm{d}u, \qquad t_0 \leqslant t \leqslant T,$$

where $(W^\beta(t))_{t_0 \leqslant t \leqslant T}$ is the $\mathbb{P}^\beta$-Brownian motion appearing in (3.9). By the Girsanov theorem, this amounts, for $t_0 \leqslant t \leqslant T$, to the likelihood ratio computation

$$Z(t) := \frac{\mathrm{d}\mathbb{P}^\beta}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}(t)} = \exp\bigg(- \int_{t_0}^t \Big\langle \beta(X(u))\,,\,\mathrm{d}W(u) \Big\rangle - \tfrac{1}{2} \int_{t_0}^t |\beta(X(u))|^2\,\mathrm{d}u \bigg). \tag{5.28}$$

Now, for each $(t,x) \in [t_0,T] \times \mathbb{R}^n$, the ratio $Y^\beta(t,x) = p^\beta(t,x)/p(t,x)$ equals the conditional expectation of the random variable (5.28) with respect to the probability measure $\mathbb{P}$, where we condition on $X(t) = x$; to wit,

$$Y^\beta(t,x) = \mathbb{E}_{\mathbb{P}}\big[Z(t) \,|\, X(t) = x\big], \qquad (t,x) \in [t_0,T] \times \mathbb{R}^n.$$

Therefore, in order to obtain the estimate (5.27), it suffices to show that the log-density process $(\log Z(t))_{t_0 \leqslant t \leqslant T}$ is uniformly bounded. Since the perturbation β is smooth and has compact support, the Lebesgue integral inside the exponential of (5.28) is uniformly bounded, as required.

In order to handle the stochastic integral with respect to the $\mathbb{P}$-Brownian motion $(W(u))_{t_0 \leqslant u \leqslant t}$ inside the exponential (5.28), we invoke the assumption that the vector field β equals the gradient of a potential $B\colon \mathbb{R}^n \to \mathbb{R}$, which is of class $\mathcal{C}^\infty(\mathbb{R}^n;\mathbb{R})$ and has compact support. According to Itô's formula and (1.3), we can express the stochastic integral appearing in (5.28) as

$$\int_{t_0}^t \Big\langle \beta(X(u))\,,\,\mathrm{d}W(u) \Big\rangle = B(X(t)) - B(X(t_0)) + \int_{t_0}^t \Big(\langle \beta\,,\,\nabla\Psi \rangle - \tfrac{1}{2}\,\mathrm{div}\,\beta \Big)(X(u))\,\mathrm{d}u \tag{5.29}$$

for $t_0 \leqslant t \leqslant T$. At this stage it becomes obvious that the expression of (5.29) is uniformly bounded. This completes the proof of Lemma 5.7. □

The following Lemma 5.8 provides the crucial estimates (4.18) and (4.19), needed in the proofs of Corollary 4.4 and Proposition 4.7.

Lemma 5.8. *Under the Assumptions 2.2, we let $t_0 \geqslant 0$ and $T > t_0$. There is a constant $C > 0$ such that*

$$\left|Y^\beta(T-s,x) - 1\right| \leqslant C\,(T - t_0 - s), \tag{5.30}$$

as well as

$$\mathbb{E}_{\mathbb{P}}\left[\int_s^{T-t_0} \left|\nabla \log Y^\beta(T-u, X(T-u))\right|^2 \mathrm{d}u \;\middle|\; X(T-s) = x\right] \leqslant C\,(T-t_0-s)^2, \tag{5.31}$$

hold for all $0 \leqslant s \leqslant T - t_0$ and $x \in \mathbb{R}^n$. Furthermore, for every $t_0 > 0$ and $x \in \mathbb{R}^n$ we have the pointwise limiting assertion

$$\lim_{s \uparrow T-t_0} \frac{\log Y^\beta(T-s,x)}{T-t_0-s} = \operatorname{div} \beta(x) + \Big\langle \beta(x)\,,\, \nabla \log p(t_0,x)\Big\rangle, \tag{5.32}$$

where the fraction on the left-hand side of (5.32) is uniformly bounded on $[0, T-t_0] \times \mathbb{R}^n$.

Remark 5.9. The pointwise limiting assertion (5.32) is the deterministic analogue of the trajectorial relation (4.35) from Proposition 4.7. In Section 5.4 below we will prove that the limiting assertion (4.35) holds in L^1 under both $\mathbb{P}$ and $\mathbb{P}^\beta$, and is in fact valid for all $t_0 > 0$. ⋄

Proof of Lemma 5.8. As $\log Y^\beta = \log \ell^\beta - \log \ell$, we obtain from Theorems 4.1, 4.2 and (5.27) that the martingale parts of the processes in (5.25) and (5.26) are bounded in $L^2(\mathbb{P})$, i.e.,

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{T-t_0} \frac{\left|\nabla Y^\beta(T-u, X(T-u))\right|^2}{Y^\beta(T-u, X(T-u))^2}\,\mathrm{d}u\right] < \infty. \tag{5.33}$$

Once again using (5.27), we compare $\nabla Y^\beta / Y^\beta$ with ∇Y^β to see that (5.33) also implies

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{T-t_0} \left|\nabla Y^\beta(T-u, X(T-u))\right|^2 \mathrm{d}u\right] < \infty. \tag{5.34}$$

According to (5.25), the time-reversed ratio process (5.24) satisfies with respect to the backwards filtration $\widetilde{\mathbb{G}}$ the stochastic differential equation

$$\begin{aligned}\mathrm{d}R^\beta(T-s) &= \Big\langle \nabla R^\beta(T-s)\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta(X(T-s))\,\mathrm{d}s\Big\rangle \\ &\quad - R^\beta(T-s)\Big(\operatorname{div}\beta(X(t-s)) + \Big\langle \beta(X(T-s))\,,\nabla \log p(T-s, X(T-s))\Big\rangle\Big)\,\mathrm{d}s.\end{aligned} \tag{5.35}$$

In view of (5.34), the martingale part in (5.35) is bounded in $L^2(\mathbb{P})$. As regards the drift term of this equation, we observe that it vanishes when $X(T-s)$ takes values outside the compact support of the smooth vector field β. Consequently, the drift term is bounded, i.e., the constant

$$C_1 := \sup_{\substack{t_0 \leqslant t \leqslant T \\ y \in \mathbb{R}^n}} \left| - Y^\beta(t,y)\left(\operatorname{div}\beta(y) + \left\langle \beta(y)\,,\nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)}\right\rangle\right)\right|$$

is finite, and the processes

$$Y^\beta(T-s, X(T-s)) + C_1\, s \qquad \text{and} \qquad Y^\beta(T-s, X(T-s)) - C_1\, s$$

for $0 \leqslant s \leqslant T - t_0$, are a sub- and a supermartingale, respectively. We conclude that

$$\Big| Y^\beta(T-s,x) - \mathbb{E}_{\mathbb{P}}\Big[Y^\beta(t_0, X(t_0)) \;\Big|\; X(T-s) = x \Big] \Big| \leqslant C_1\,(T - t_0 - s)$$

holds for all $0 \leqslant s \leqslant T - t_0$ and $x \in \mathbb{R}^n$. Since $Y^\beta(t_0, \cdot\,) = 1$, this establishes the first estimate

$$\big| Y^\beta(T-s,x) - 1 \big| \leqslant C_1\,(T - t_0 - s). \tag{5.36}$$

Now we turn our attention to the second estimate (5.31). We fix $0 \leqslant s \leqslant T - t_0$ and $x \in \mathbb{R}^n$. By means of the stochastic differentials in (5.26) and (5.35), we find that the expression

$$\tfrac{1}{2}\, \mathbb{E}_{\mathbb{P}}\left[\int_s^{T-t_0} \Big| \nabla \log Y^\beta(T-u, X(T-u)) \Big|^2 \,\mathrm{d}u \;\Big|\; X(T-s) = x \right] \tag{5.37}$$

is equal to

$$\log Y^\beta(T-s,x) - Y^\beta(T-s,x) + 1 + \mathbb{E}_{\mathbb{P}}\left[\int_s^{T-t_0} G(T-u, X(T-u))\,\mathrm{d}u \;\Big|\; X(T-s) = x \right], \tag{5.38}$$

where we have set

$$G(t,y) := (Y^\beta(t,y) - 1) \left(\operatorname{div} \beta(y) + \left\langle \beta(y)\,, \nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)} \right\rangle \right),$$

for $t_0 \leqslant t \leqslant T$ and $y \in \mathbb{R}^n$. Introducing the finite constant

$$C_2 := \sup_{\substack{t_0 \leqslant t \leqslant T \\ y \in \mathbb{R}^n}} \left| \operatorname{div} \beta(y) + \left\langle \beta(y)\,, \nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)} \right\rangle \right|$$

and using the just proved estimate (5.36), we see that the absolute value of the conditional expectation appearing in (5.38) can be bounded by $C_1\, C_2\, (T - t_0 - s)^2$. In order to handle the remaining terms of (5.38), we apply the elementary inequality $\log p \leqslant p - 1$, which is valid for all $p > 0$, and obtain

$$\log Y^\beta(T-s,x) - Y^\beta(T-s,x) + 1 \leqslant 0. \tag{5.39}$$

This implies that the expression of (5.37) is bounded by $C_1\, C_2\, (T - t_0 - s)^2$, which establishes the second estimate (5.31).

We also note that the elementary inequality (5.39) in conjunction with the estimate (5.36) shows that

$$\log Y^\beta(T-s,x) \leqslant C_1\,(T - t_0 - s)$$

for all $0 \leqslant s \leqslant T - t_0$ and $x \in \mathbb{R}^n$; this implies that the fraction on the left-hand side of (5.32) is uniformly bounded on $[0, T - t_0] \times \mathbb{R}^n$.

Regarding the limiting assertion (5.32), we fix $t_0 > 0$, $x \in \mathbb{R}^n$ and $0 \leqslant s \leqslant T - t_0$, and take conditional expectations with respect to $X(T-s) = x$ in the integral version of the stochastic differential (5.26). On account of (5.33), the stochastic integral with respect to the $\mathbb{P}$-Brownian motion $(\overline{W}^{\mathbb{P}}(T-s))_{0 \leqslant s \leqslant T}$ in (5.26) vanishes. Dividing by $T - t_0 - s$ and passing to the limit as $s \uparrow T - t_0$, we can use the estimate (5.31) to deduce that the expression in the third line of (5.26) vanishes in the limit. After applying the Cauchy–Schwarz inequality, we see that the normalized integral involving the perturbation β appearing in the first line of (5.26) can be bounded by

$$\frac{1}{T - t_0 - s} \int_s^{T-t_0} \left| \nabla \log Y^\beta(T-u, X(T-u)) \right| \cdot \left| \beta(X(T-u)) \right| \mathrm{d}u. \tag{5.40}$$

By conditions (iv), (v) of Assumptions 2.2, the function $(t, x) \mapsto \nabla \log Y^\beta(t, x)$ is continuous on $(0, \infty) \times \mathbb{R}^n$, thus the expression in (5.40) is uniformly bounded on the rectangle $[0, T - t_0] \times \operatorname{supp} \beta$. As $\log Y^\beta(t_0, \cdot) = 0$, it converges $\mathbb{P}$-a.s. to zero, hence also the expectation

$$\mathbb{E}_{\mathbb{P}} \left[\frac{1}{T - t_0 - s} \int_s^{T-t_0} \left| \nabla \log Y^\beta(T-u, X(T-u)) \right| \cdot \left| \beta(X(T-u)) \right| \mathrm{d}u \;\middle|\; X(T-s) = x \right]$$

converges to zero, as $s \uparrow T - t_0$. Finally, continuity and uniform boundedness imply that the conditional expectations of the normalized integrals over the second line of (5.26) converge to the right-hand side of (5.32), as claimed. □

5.4. Completing the proof of Proposition 4.7

With the preparations of Section 5.3, we are now able to complete the proof of Proposition 4.7 by establishing the remaining limiting assertions (4.35) and (4.34) therein.

Proof of (4.35) in Proposition 4.7: Let $t_0 > 0$ and select $T > t_0$. Using the notation of (5.23) above, we have to calculate the limit

$$\lim_{s \uparrow T - t_0} \frac{\log Y^\beta(T-s, X(T-s))}{T - t_0 - s}. \tag{5.41}$$

Fix $0 \leqslant s \leqslant T - t_0$. According to the integral version of the stochastic differential (5.26), the fraction in (5.41) is equal to the sum of the following four normalized integral terms (5.42) – (5.44) and (5.46), whose behavior as $s \uparrow T - t_0$ we will study separately below. By conditions (iv), (v) of Assumptions 2.2, the function $(t, x) \mapsto \nabla \log Y^\beta(t, x)$ is continuous on $(0, \infty) \times \mathbb{R}^n$, thus the first expression

$$\frac{1}{T - t_0 - s} \int_s^{T-t_0} \Big(\operatorname{div} \beta(X(T-u)) + \big\langle \beta(X(T-u)) \,, \nabla \log p(T-u, X(T-u)) \big\rangle \Big) \, \mathrm{d}u \tag{5.42}$$

is uniformly bounded on $[0, T-t_0] \times \operatorname{supp}\beta$. Using continuity and uniform boundedness, we conclude that (5.42) converges $\mathbb{P}$-a.s. as well as in $L^1(\mathbb{P})$ to the right-hand side of (4.35), as required. Thus it remains to show that the three remaining terms converge to zero. Using continuity and uniform boundedness once again, we deduce from $\log Y^\beta(t_0, \cdot) = 0$ that the second integral term

$$\frac{1}{T-t_0-s}\int_s^{T-t_0}\left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}\,,\, \beta(X(T-u))\right\rangle \mathrm{d}u \tag{5.43}$$

converges to zero $\mathbb{P}$-a.s. and in $L^1(\mathbb{P})$. Since $\log Y^\beta(t_0, \cdot) = 0$ and because the integrand is continuous, we see that the third expression

$$\frac{1}{T-t_0-s}\int_s^{T-t_0}\frac{1}{2}\frac{\left|\nabla Y^\beta(T-u, X(T-u))\right|^2}{Y^\beta(T-u, X(T-u))^2}\,\mathrm{d}u \tag{5.44}$$

converges $\mathbb{P}$-a.s. to zero. Furthermore, owing to Lemma 5.8, there is a constant $C > 0$ such that

$$\mathbb{E}_{\mathbb{P}}\left[\frac{1}{T-t_0-s}\int_s^{T-t_0}\frac{\left|\nabla Y^\beta(T-u, X(T-u))\right|^2}{Y^\beta(T-u, X(T-u))^2}\,\mathrm{d}u\right] \leqslant C\,(T-t_0-s) \tag{5.45}$$

holds for all $0 \leqslant s \leqslant T-t_0$, which implies that (5.44) converges to zero also in $L^1(\mathbb{P})$. The fourth and last term is the stochastic integral

$$-\frac{1}{T-t_0-s}\int_s^{T-t_0}\left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u)\right\rangle. \tag{5.46}$$

The expression (5.44) converges to zero $\mathbb{P}$-a.s. and according to (5.45) we have

$$\mathbb{E}_{\mathbb{P}}\left[\frac{1}{(T-t_0-s)^2}\int_s^{T-t_0}\frac{\left|\nabla Y^\beta(T-u, X(T-u))\right|^2}{Y^\beta(T-u, X(T-u))^2}\,\mathrm{d}u\right] \leqslant C.$$

By means of the Itô isometry, we deduce that

$$\lim_{s\uparrow T-t_0}\mathbb{E}_{\mathbb{P}}\left[\left(\frac{1}{T-t_0-s}\int_s^{T-t_0}\left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u)\right\rangle\right)^2\right] = 0.$$

In other words, the normalized stochastic integral of (5.46) converges to zero in $L^2(\mathbb{P})$.

Summing up, we have shown that the limiting assertion (4.35) holds in $L^1(\mathbb{P})$ for every $t_0 > 0$. As we have seen in Lemma 5.7, the probability measures $\mathbb{P}$ and $\mathbb{P}^\beta$ are equivalent, the Radon–Nikodým derivatives $\frac{\mathrm{d}\mathbb{P}^\beta}{\mathrm{d}\mathbb{P}}$ and $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}^\beta}$ are bounded on the σ-algebra $\mathcal{F}(T) = \mathcal{G}(0)$, and therefore convergence in $L^1(\mathbb{P})$ is equivalent to convergence in $L^1(\mathbb{P}^\beta)$. This completes the proof of (4.35). □

Proof of (4.34) in Proposition 4.7: This is proved in very much the same way, as (4.33), (4.35). The only novelty here is the use of (5.18) to pass to the $\mathbb{P}$-Brownian motion $\overline{W}^{\mathbb{P}}$ from the $\mathbb{P}^\beta$-Brownian motion $\overline{W}^{\mathbb{P}^\beta}$, and the reliance on $\mathbb{E}_{\mathbb{P}^\beta}[F^\beta(t_0)] < \infty$ to ensure that the resulting stochastic integral is a (square-integrable) $\mathbb{P}$-martingale. We leave the details to the diligent reader. □

5.5. The rate of growth for the Wasserstein distance

Let us recapitulate the message of Corollaries 4.3 and 4.4: in these results we compare the rate of decay for the relative entropy with the rate of growth for the quadratic Wasserstein distance W_2 along the curves $(P(t))_{t\geqslant 0}$ and $(P^\beta(t))_{t\geqslant t_0}$ in $\mathscr{P}_2(\mathbb{R}^n)$. This is the essence of the gradient flow property formalized in Theorem 3.3.

In order to complete the proofs of Corollaries 4.3 and 4.4, we have to establish the limits (4.14) and (4.17). The limit (4.14) is well known (see [AGS08]) to exist, under suitable regularity assumptions, for *Lebesgue-a.e.* $t_0 \geqslant 0$. A similar remark pertains to the "perturbed" limit (4.17): if we replace t_0 by s_0 in (4.17), it is well known that this limit exists for *Lebesgue-a.e.* $s_0 \geqslant t_0$. But this is not what we need. We have to prove the validity of (4.17) for *the point* t_0 *itself*, in order to calculate the slope of the function

$$[t_0, \infty) \ni t \longmapsto H(P^\beta(t) \,|\, \mathrm{Q})$$

with respect to the Wasserstein distance *at time* t_0. After all, the deviation of $P^\beta(t)$ from $P(t)$ takes place at time t_0.

This technical aspect turns out to be quite delicate. We already needed a careful analysis (recall the estimates (4.18), (4.19)) to show that the exceptional set N of (3.6), defined in terms of the decay of entropy of the unperturbed curve $(P(t))_{t\geqslant 0}$, does not change when passing to the perturbed curve $(P^\beta(t))_{t\geqslant t_0}$. In addition, we have to show that this set N also cannot increase when passing from the unperturbed Wasserstein limit (4.14) to its perturbed counterpart (4.17). In order to do this, we have to rely here (and only here) on condition (vi) of Assumptions 2.4.

For a detailed discussion of metric measure spaces and in particular Wasserstein spaces, we refer to [AG13, AGS08] and [Stu06a, Stu06b].

For fixed $T \in (0, \infty)$, we define now the time-dependent velocity field

$$[0,T] \times \mathbb{R}^n \ni (t,x) \longmapsto v(t,x) := -\Big(\frac{1}{2}\frac{\nabla p(t,x)}{p(t,x)} + \nabla\Psi(x)\Big) = -\frac{1}{2}\frac{\nabla \ell(t,x)}{\ell(t,x)} \in \mathbb{R}^n. \quad (5.47)$$

According to condition (vi) in Assumptions 2.4, this gradient vector field $v(t,\,\cdot\,)$ is an element of the tangent space (see Definition 8.4.1 in [AGS08]) of $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$, i.e.,

$$v(t,\,\cdot\,) \in \mathrm{Tan}_{P(t)}\mathscr{P}_2(\mathbb{R}^n) := \overline{\big\{\nabla\varphi\colon\ \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n;\mathbb{R})\big\}}^{L^2(P(t))}. \quad (5.48)$$

We can now formulate the "unperturbed" version of our desired result.

Theorem 5.10 (Limiting behavior of the quadratic Wasserstein distance)**.**
Under the Assumptions 2.4, let $t_0 \geqslant 0$ *be such that the generalized de Bruijn identity (3.2), (4.13) is valid. Then we have the two-sided limit*

$$\lim_{t\to t_0} \frac{W_2(P(t),P(t_0))}{|t-t_0|} = \Big(\mathbb{E}_{\mathbb{P}}\big[\,|v(t_0,X(t_0))|^2\,\big]\Big)^{1/2} = \tfrac{1}{2}\sqrt{I(P(t_0)\,|\,\mathrm{Q})}. \quad (5.49)$$

Before dealing with Theorem 5.10, we will prove the more general Theorem 5.11 below, which amounts to the perturbed version of Theorem 5.10. For right-derivatives, the latter then simply follows by setting $\beta \equiv 0$ in the statement of Theorem 5.11.

We consider the "perturbed" curve $(P^\beta(t))_{t\geqslant t_0}$ in $\mathscr{P}_2(\mathbb{R}^n)$, as defined in (3.7) – (3.9), and define the time-dependent perturbed velocity field

$$[t_0,T]\times\mathbb{R}^n \ni (t,x) \longmapsto v^\beta(t,x) := -\Big(\frac{1}{2}\frac{\nabla p^\beta(t,x)}{p^\beta(t,x)} + \nabla\Psi(x) + \beta(x)\Big) \in \mathbb{R}^n. \tag{5.50}$$

At this point, we recall that the perturbation $\beta\colon \mathbb{R}^n \to \mathbb{R}^n$ is a *gradient vector field*, i.e., of the form $\beta = \nabla B$ for some smooth potential $B\colon \mathbb{R}^n \to \mathbb{R}$ with compact support. Since $p(t_0,\,\cdot\,) = p^\beta(t_0,\,\cdot\,)$, at time t_0 the vector fields of (5.47) and (5.50) are related via

$$v^\beta(t_0,x) = v(t_0,x) - \nabla B(x) = -\nabla\Big(\tfrac{1}{2}\log\ell(t_0,x) + B(x)\Big), \qquad x\in\mathbb{R}^n. \tag{5.51}$$

Using the regularity assumption that the potential B is of class $\mathcal{C}^\infty_c(\mathbb{R}^n;\mathbb{R})$, we conclude from (5.48) and (5.51) that the perturbed vector field $v^\beta(t_0,\,\cdot\,)$ is also an element of the tangent space of $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P^\beta(t_0) = P(t_0) \in \mathscr{P}_2(\mathbb{R}^n)$, i.e.,

$$v^\beta(t_0,\,\cdot\,) \in \mathrm{Tan}_{P^\beta(t_0)}\mathscr{P}_2(\mathbb{R}^n) = \overline{\{\nabla\varphi^\beta\colon\ \varphi^\beta\in\mathcal{C}^\infty_c(\mathbb{R}^n;\mathbb{R})\}}^{L^2(P^\beta(t_0))}. \tag{5.52}$$

Theorem 5.11 (Limiting behavior of the Wasserstein distance under perturbations)**.**
Under the Assumptions 2.4, for every point $t_0 \in \mathbb{R}_+ \setminus N$ (at which the right-sided limiting identity (3.6) is valid), we have the one-sided limit

$$\lim_{t\downarrow t_0}\frac{W_2(P^\beta(t),P^\beta(t_0))}{t-t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,|v^\beta(t_0,X(t_0))|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a+2b\|_{L^2(\mathbb{P})}. \tag{5.53}$$

Here $a = \nabla\log\ell(t_0,X(t_0))$ and $b = \beta(X(t_0))$ as in (3.10).

Proof. The second equality in (5.53) is apparent from the definition (5.50) of the time-dependent perturbed velocity field $(v^\beta(t,\,\cdot\,))_{t\geqslant t_0}$. The delicate point is to show that the limiting assertion (5.53) is valid for every $t_0 \in \mathbb{R}_+ \setminus N$.

In order to see this, let us fix some $t_0 \in \mathbb{R}_+ \setminus N$ so that the limiting identity (3.6) is valid. In the following steps we prove that then the limiting assertion (5.53) also holds.

Step 1. The gradient vector field $v^\beta(t_0,\,\cdot\,)$ induces a family of *linearized transport maps*

$$\mathcal{X}^\beta_t(x) := x + (t-t_0)\cdot v^\beta(t_0,x), \qquad x\in\mathbb{R}^n \tag{5.54}$$

for $t \geqslant t_0$ in the manner of (4.38), and we denote by $P^\beta_{\mathcal{X}}(t)$ the image measure of $P^\beta(t_0) = P(t_0)$ under the transport map $\mathcal{X}^\beta_t\colon \mathbb{R}^n \to \mathbb{R}^n$; i.e.,

$$P^\beta_{\mathcal{X}}(t) := (\mathcal{X}^\beta_t)_{\#}P^\beta(t_0), \qquad t\geqslant t_0.$$

To motivate the arguments that follow, let us first pretend that, for all $t > t_0$ sufficiently close to t_0, the map $\mathcal{X}_t^\beta$ is the *optimal quadratic Wasserstein transport* from $P^\beta(t_0)$ to $P_{\mathcal{X}}^\beta(t)$; i.e.,

$$W_2^2(P_{\mathcal{X}}^\beta(t), P^\beta(t_0)) = \mathbb{E}_{\mathbb{P}^\beta}\Big[\,\big|\mathcal{X}_t^\beta(X(t_0)) - X(t_0)\big|^2\,\Big] = \mathbb{E}_{\mathbb{P}}\Big[\,\big|\mathcal{X}_t^\beta(X(t_0)) - X(t_0)\big|^2\,\Big],$$

where we have used in the last equality the fact that $X(t_0)$ has the same distribution under $\mathbb{P}^\beta$ as it does under $\mathbb{P}$. Then, on account of (5.54), we could conclude that

$$\lim_{t\downarrow t_0} \frac{W_2(P_{\mathcal{X}}^\beta(t), P^\beta(t_0))}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}. \tag{5.55}$$

Furthermore, let us suppose that we can show the limiting identity

$$\lim_{t\downarrow t_0} \frac{W_2(P^\beta(t), P_{\mathcal{X}}^\beta(t))}{t - t_0} = 0, \tag{5.56}$$

which has the interpretation that "the straight line $(P_{\mathcal{X}}^\beta(t))_{t\geqslant t_0}$ is *tangential* to the curve $(P^\beta(t))_{t\geqslant t_0}$". Using (5.55) and (5.56), we could now derive the desired equality (5.53). Indeed, invoking the triangle inequality for the quadratic Wasserstein distance we obtain

$$\lim_{t\downarrow t_0} \frac{W_2(P_{\mathcal{X}}^\beta(t), P^\beta(t_0))}{t - t_0} \leqslant \lim_{t\downarrow t_0} \frac{W_2(P_{\mathcal{X}}^\beta(t), P^\beta(t))}{t - t_0} + \liminf_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0}, \tag{5.57}$$

and one more application of the triangle inequality yields

$$\limsup_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} \leqslant \lim_{t\downarrow t_0} \frac{W_2(P^\beta(t), P_{\mathcal{X}}^\beta(t))}{t - t_0} + \lim_{t\downarrow t_0} \frac{W_2(P_{\mathcal{X}}^\beta(t), P^\beta(t_0))}{t - t_0}. \tag{5.58}$$

Step 2. The bad news at this point is that there is little reason why, for $t > t_0$ sufficiently close to t_0, the map $\mathcal{X}_t^\beta$ defined in (5.54) of Step 1 should be an optimal transport for the quadratic Wasserstein distance; i.e., by Brenier's theorem [Bre91], be equal to the gradient of a *convex* function. The good news is that we can reduce the general case to the situation of optimal transports $\mathcal{X}_t^\beta$ as in Step 1 by localizing the vector field $v^\beta(t_0, \cdot\,)$ as well as the transport maps $(\mathcal{X}_t^\beta)_{t\geqslant t_0}$ to compact subsets of $\mathbb{R}^n$ (Steps 2 – 4); and that, after these localizations have been carried out, an analogue of the equality (5.56) also holds, allowing us to complete the argument (Steps 5 – 6).

To this end, we recall that the vector field $v^\beta(t_0, \cdot\,)$ from (5.51) is an element of the tangent space $\mathrm{Tan}_{P^\beta(t_0)}\mathscr{P}_2(\mathbb{R}^n)$ of the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P^\beta(t_0) \in \mathscr{P}_2(\mathbb{R}^n)$. Thus, we can choose a sequence of smooth and compactly supported functions $(\varphi_m^\beta(t_0, \cdot\,))_{m\geqslant 1} \subseteq \mathcal{C}_c^\infty(\mathbb{R}^n; \mathbb{R})$ such that

$$\lim_{m\to\infty} \mathbb{E}_{\mathbb{P}}\Big[\,\Big|v^\beta(t_0, X(t_0)) - \nabla\varphi_m^\beta(t_0, X(t_0))\Big|^2\,\Big] = 0. \tag{5.59}$$

Next, for each $m \in \mathbb{N}$, we define the *localized gradient vector fields*

$$v_m^\beta(t_0, x) := \nabla \varphi_m^\beta(t_0, x), \qquad x \in \mathbb{R}^n. \tag{5.60}$$

These have compact support, approximate the gradient vector field $v^\beta(t_0, \cdot\,)$ in $L^2(P(t_0))$ as in (5.59), and induce a family of *localized linear transports* $(\mathcal{X}_t^{\beta,m})_{t \geqslant t_0}$ defined by analogy with (5.54) via

$$\mathcal{X}_t^{\beta,m}(x) := x + (t - t_0) \cdot v_m^\beta(t_0, x), \qquad x \in \mathbb{R}^n. \tag{5.61}$$

We denote by $P_{\mathcal{X}}^{\beta,m}(t)$ the image measure of $P^\beta(t_0) = P(t_0)$ under this localized linear transport map $\mathcal{X}_t^{\beta,m} \colon \mathbb{R}^n \to \mathbb{R}^n$; i.e.,

$$P_{\mathcal{X}}^{\beta,m}(t) := (\mathcal{X}_t^{\beta,m})_\# P^\beta(t_0), \qquad t \geqslant t_0. \tag{5.62}$$

Step 3. We claim that, for every $m \in \mathbb{N}$, there exists some $\varepsilon_m > 0$ such that for all $t \in (t_0, t_0 + \varepsilon_m)$, the localized linear transport map $\mathcal{X}_t^{\beta,m} \colon \mathbb{R}^n \to \mathbb{R}^n$ of (5.61), constructed in Step 2, defines an optimal Wasserstein transport from $P^\beta(t_0)$ to $P_{\mathcal{X}}^{\beta,m}(t)$. Hence, by Brenier's theorem ([Bre91], [Vil03, Theorem 2.12]), we have to show that $\mathcal{X}_t^{\beta,m}$ is the *gradient of a convex function*, for all $t > t_0$ sufficiently close to t_0.

Indeed, from the definitions in (5.60), (5.61) we see that the functions $\mathcal{X}_t^{\beta,m}$ are gradients, for all $m \in \mathbb{N}$ and $t \geqslant t_0$. More precisely, we have

$$\mathcal{X}_t^{\beta,m}(x) = \nabla\Big(\tfrac{1}{2}|x|^2 + (t - t_0) \cdot \varphi_m^\beta(t_0, x)\Big), \qquad x \in \mathbb{R}^n. \tag{5.63}$$

As the Hessian matrix of $\varphi_m^\beta(t_0, \cdot\,)$ is uniformly bounded, the function in the bracket of (5.63) is a convex function of x for every $m \in \mathbb{N}$ and $t \in (t_0, t_0 + \varepsilon_m)$, for $\varepsilon_m > 0$ small enough. We also note for later use that $\mathcal{X}_t^{\beta,m}$ defines a Lipschitz bijection on $\mathbb{R}^n$, again for every $m \in \mathbb{N}$ and $t \in (t_0, t_0 + \varepsilon_m)$.

Step 4. From Step 3 we know that, for every $m \in \mathbb{N}$, there exists some $\varepsilon_m > 0$ such that for all $t \in (t_0, t_0 + \varepsilon_m)$, the localized map $\mathcal{X}_t^{\beta,m}$ is the optimal transport from $P^\beta(t_0)$ to $P_{\mathcal{X}}^{\beta,m}(t)$ with respect to the quadratic Wasserstein distance. Therefore, we can apply the considerations of Step 1 to the optimal map $\mathcal{X}_t^{\beta,m}$ in (5.61), and deduce that

$$\lim_{t \downarrow t_0} \frac{W_2(P_{\mathcal{X}}^{\beta,m}(t), P^\beta(t_0))}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v_m^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2}$$

holds for every $m \in \mathbb{N}$. Invoking (5.59) and (5.60), we obtain from this

$$\lim_{m \to \infty} \lim_{t \downarrow t_0} \frac{W_2(P_{\mathcal{X}}^{\beta,m}(t), P^\beta(t_0))}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}.$$

From the inequalities (5.57) and (5.58) of Step 1 (with $P_{\mathcal{X}}^{\beta,m}(t)$ instead of $P_{\mathcal{X}}^{\beta}(t)$) it follows that, in order to conclude (5.53), it remains to establish the analogue of the identity (5.56):

$$\lim_{m\to\infty} \lim_{t\downarrow t_0} \frac{W_2(P^\beta(t), P_{\mathcal{X}}^{\beta,m}(t))}{t-t_0} = 0. \tag{5.64}$$

Step 5. The time-dependent velocity field $(v^\beta(t,\,\cdot\,))_{t\geqslant t_0}$ induces a *curved flow* $(\mathcal{Y}_t^\beta)_{t\geqslant t_0}$, which is characterized by

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{Y}_t^\beta = v^\beta(t,\mathcal{Y}_t^\beta) \qquad \text{for all } t\geqslant t_0, \qquad \mathcal{Y}_{t_0}^\beta = \mathrm{Id}_{\mathbb{R}^n}. \tag{5.65}$$

Then, for every $t\geqslant t_0$, the map $\mathcal{Y}_t^\beta\colon \mathbb{R}^n\to\mathbb{R}^n$ transports the measure $P^\beta(t_0)=P(t_0)$ to $P^\beta(t)$, i.e., $(\mathcal{Y}_t^\beta)_\# P^\beta(t_0) = P^\beta(t)$.

On the other hand, the localized linear mappings $\mathcal{X}_t^{\beta,m}\colon \mathbb{R}^n\to\mathbb{R}^n$ of (5.61) transport $P^\beta(t_0)$ to $P_{\mathcal{X}}^{\beta,m}(t)$, as in (5.62). As mentioned at the end of Step 3, the inverse mappings $(\mathcal{X}_t^{\beta,m})^{-1}\colon \mathbb{R}^n\to\mathbb{R}^n$ are well-defined for all $m\in\mathbb{N}$ and $t\in(t_0,t_0+\varepsilon_m)$; they satisfy

$$\big((\mathcal{X}_t^{\beta,m})^{-1}\big)_\# P_{\mathcal{X}}^{\beta,m}(t) = P^\beta(t_0), \qquad t\in(t_0,t_0+\varepsilon_m).$$

From Step 4, our remaining task is to prove (5.64). To this end, we have to construct maps $\mathcal{Z}_t^{\beta,m}\colon\mathbb{R}^n\to\mathbb{R}^n$ that transport $P_{\mathcal{X}}^{\beta,m}(t)$ to $P^\beta(t)$, i.e., $(\mathcal{Z}_t^{\beta,m})_\# P_{\mathcal{X}}^{\beta,m}(t) = P^\beta(t)$, and satisfy

$$\lim_{m\to\infty}\lim_{t\downarrow t_0} \frac{1}{t-t_0}\left(\mathbb{E}_{\mathbb{P}_{\mathcal{X}}^{\beta,m}}\Big[\,\big|\mathcal{Z}_t^{\beta,m}(X(t))-X(t)\big|^2\,\Big]\right)^{1/2} = 0, \tag{5.66}$$

where $\mathbb{P}_{\mathcal{X}}^{\beta,m}$ denotes a probability measure on path space under which the random variable $X(t)$ has distribution $P_{\mathcal{X}}^{\beta,m}(t)$ as in (5.62). We define for this job the candidate maps

$$\mathcal{Z}_t^{\beta,m} := \mathcal{Y}_t^\beta \circ (\mathcal{X}_t^{\beta,m})^{-1}, \qquad t\in(t_0,t_0+\varepsilon_m); \tag{5.67}$$

recall that $(\mathcal{X}_t^{\beta,m})^{-1}$ transports $P_{\mathcal{X}}^{\beta,m}(t)$ to $P^\beta(t_0)$ while $\mathcal{Y}_t^\beta$ transports $P^\beta(t_0)$ to $P^\beta(t)$; and conclude that $\mathcal{Z}_t^{\beta,m}$ of (5.67) transports $P_{\mathcal{X}}^{\beta,m}(t)$ to $P^\beta(t)$. Thus, we obtain

$$\mathbb{E}_{\mathbb{P}_{\mathcal{X}}^{\beta,m}}\Big[\,\big|\mathcal{Z}_t^{\beta,m}(X(t))-X(t)\big|^2\,\Big] = \mathbb{E}_{\mathbb{P}}\Big[\,\big|\mathcal{Y}_t^\beta(X(t_0))-\mathcal{X}_t^{\beta,m}(X(t_0))\big|^2\,\Big]. \tag{5.68}$$

Combining (5.66) with (5.68), we see that we have to establish

$$\lim_{m\to\infty}\lim_{t\downarrow t_0}\frac{1}{(t-t_0)^2}\,\mathbb{E}_{\mathbb{P}}\Big[\,\big|\mathcal{Y}_t^\beta(X(t_0))-\mathcal{X}_t^{\beta,m}(X(t_0))\big|^2\,\Big] = 0. \tag{5.69}$$

Using (5.61) and the elementary inequality $|x+y|^2\leqslant 2(|x|^2+|y|^2)$, for $x,y\in\mathbb{R}^n$, we derive the estimate

$$\tfrac{1}{2}\,|\mathcal{Y}_t^\beta(x)-\mathcal{X}_t^{\beta,m}(x)|^2 \leqslant (t-t_0)^2\cdot|v^\beta(t_0,x)-v_m^\beta(t_0,x)|^2 \tag{5.70}$$

$$+\,\big|(\mathcal{Y}_t^\beta(x)-x)-(t-t_0)\cdot v^\beta(t_0,x)\big|^2. \tag{5.71}$$

Therefore, in order to establish (5.69), it suffices to show the limiting assertions (5.72) and (5.73) below; these correspond to (5.70) and (5.71), respectively. The first limiting identity we already have from (5.59), (5.60) of Step 2, namely,

$$\lim_{m\to\infty} \mathbb{E}_{\mathbb{P}}\Big[\, \big|v^{\beta}(t_0, X(t_0)) - v_m^{\beta}(t_0, X(t_0))\big|^2 \,\Big] = 0. \tag{5.72}$$

Step 6. Our final task is to justify that

$$\lim_{t\downarrow t_0} \mathbb{E}_{\mathbb{P}}\Big[\, \big|\tfrac{1}{t-t_0}\big(\mathcal{Y}_t^{\beta}(X(t_0)) - X(t_0)\big) - v^{\beta}(t_0, X(t_0))\big|^2 \,\Big] = 0. \tag{5.73}$$

To this end, we first note that by (5.65) we have for all $t \geqslant t_0$ the identity

$$\mathcal{Y}_t^{\beta}(x) = x + \int_{t_0}^{t} v^{\beta}(u, \mathcal{Y}_u^{\beta}(x))\, \mathrm{d}u, \qquad x \in \mathbb{R}^n,$$

on whose account the expectation in (5.73) is equal to

$$\mathbb{E}_{\mathbb{P}}\Big[\, \Big|\frac{1}{t-t_0}\int_{t_0}^{t} v^{\beta}\big(u, \mathcal{Y}_u^{\beta}(X(t_0))\big)\, \mathrm{d}u - v^{\beta}(t_0, X(t_0))\Big|^2 \,\Big].$$

As $\mathcal{Y}_t^{\beta}$ transports $P^{\beta}(t_0) = P(t_0)$ to $P^{\beta}(t)$, and because the random variable $X(t_0)$ has the same distribution under $\mathbb{P}^{\beta}$ as it does under $\mathbb{P}$, this expectation can also be expressed with respect to the probability measure $\mathbb{P}^{\beta}$, and it thus suffices to show the limiting assertion

$$\lim_{t\downarrow t_0} \mathbb{E}_{\mathbb{P}^{\beta}}\Big[\, \Big|\frac{1}{t-t_0}\int_{t_0}^{t} v^{\beta}(u, X(u))\, \mathrm{d}u - v^{\beta}(t_0, X(t_0))\Big|^2 \,\Big] = 0. \tag{5.74}$$

For this purpose, we first observe that by the continuity of the paths of the canonical coordinate process $(X(t))_{t\geqslant 0}$, the family of random variables

$$\Big(\Big|\frac{1}{t-t_0}\int_{t_0}^{t} v^{\beta}(u, X(u))\, \mathrm{d}u - v^{\beta}(t_0, X(t_0))\Big|^2 \Big)_{t\geqslant t_0} \tag{5.75}$$

converges $\mathbb{P}^{\beta}$-a.s. to zero, as $t \downarrow t_0$. In order to show that their expectations also converge to zero, i.e., that (5.74) does hold, we have to verify that the family of (5.75) is uniformly integrable with respect to $\mathbb{P}^{\beta}$. As the random variable $|v^{\beta}(t_0, X(t_0))|^2$ belongs to $L^1(\mathbb{P}^{\beta})$, and we have

$$\Big|\frac{1}{t-t_0}\int_{t_0}^{t} v^{\beta}(u, X(u))\, \mathrm{d}u\Big|^2 \leqslant \frac{1}{t-t_0}\int_{t_0}^{t} |v^{\beta}(u, X(u))|^2\, \mathrm{d}u, \qquad t \geqslant t_0$$

by Jensen's inequality, it is sufficient to prove the uniform integrability of the family

$$\Big(\frac{1}{t-t_0}\int_{t_0}^{t} |v^{\beta}(u, X(u))|^2\, \mathrm{d}u \Big)_{t\geqslant t_0}. \tag{5.76}$$

Invoking the definition of the time-dependent velocity field $(v^\beta(t,\,\cdot\,))_{t\geqslant t_0}$ in (5.50) and the fact that the perturbation β is smooth and compactly supported, the uniform integrability of the family in (5.76) above, is equivalent to the uniform integrability of the family

$$\left(\frac{1}{t-t_0} \int_{t_0}^{t} \frac{\left|\nabla \ell^\beta(u, X(u))\right|^2}{\ell^\beta(u, X(u))^2}\, \mathrm{d}u \right)_{t\geqslant t_0}. \tag{5.77}$$

Now by continuity, the family of (5.77) converges $\mathbb{P}^\beta$-a.s. to $|\nabla \log \ell(t_0, X(t_0))|^2$. Thus, to establish this uniform integrability, it suffices to show that the family of random variables in (5.77) converges in $L^1(\mathbb{P}^\beta)$. Hence, in view of *Scheffé's lemma* (Lemma 6.2), it remains to check that the corresponding expectations also converge. But at this point we use for the first time our choice of $t_0 \in \mathbb{R}_+ \setminus N$ and recall (4.21), (4.23) from the proof of Corollary 4.4, which gives us

$$\lim_{t\downarrow t_0} \mathbb{E}_{\mathbb{P}^\beta}\left[\frac{1}{t-t_0} \int_{t_0}^{t} \frac{\left|\nabla \ell^\beta(u, X(u))\right|^2}{\ell^\beta(u, X(u))^2}\, \mathrm{d}u \right] = \mathbb{E}_{\mathbb{P}}\left[\frac{\left|\nabla \ell(t_0, X(t_0))\right|^2}{\ell(t_0, X(t_0))^2} \right],$$

as required. This completes the proof of the claim made in the beginning of Step 6.

Summing up, in light of (5.70), (5.71) from Step 5, the limiting assertions (5.72) and (5.73) imply the limiting behavior (5.69). According to the results of Steps 4 and 5, the latter also entails the validity of the limiting identity (5.64), which completes the proof of Theorem 5.11. □

Equipped with Theorem 5.11, we can now easily deduce Theorem 5.10.

Proof of Theorem 5.10. The second equality in (5.49) follows from the representation of the relative Fisher information in (3.5) and the definition of the time-dependent velocity field $(v(t,\,\cdot\,))_{t\geqslant t_0}$ in (5.47). The first equality in (5.49) follows from Theorem 5.11 if we set $\beta \equiv 0$. However, the limit in (5.53) is only from the right, while the limit in (5.49) is two-sided. But the only reason for considering right-sided limits in Theorem 5.11, was the presence of the perturbation β at time $t \geqslant t_0$. If there is no such perturbation, one can replace all limits from the right by two-sided ones. This completes the proof of Theorem 5.10. □

6. Proofs of auxiliary results

In this chapter we collect useful auxiliary results and provide the proofs of some technical lemmas. In Section 6.1 we discuss the definition of relative entropy with respect to the σ-finite reference measure Q, introduced in Chapter 2. The proofs of Lemmas 2.3 and 3.2 are given in Section 6.2. In Section 6.3 we state two helpful results in probability and martingale theory we applied earlier. Finally, in Section 6.4 we prove Lemma 4.10, the crucial ingredient for our trajectorial proof of the HWI inequality.

6.1. Relative entropy with respect to the σ-finite reference measure Q

For two probability measures $\mathcal{P}$ and $\mathcal{Q}$ on the Borel sets $\mathscr{B}(\mathbb{R}^n)$ of $\mathbb{R}^n$, the *relative entropy* of $\mathcal{P}$ with respect to $\mathcal{Q}$ is defined as

$$H(\mathcal{P} \,|\, \mathcal{Q}) := \int_{\mathbb{R}^n} \log\left(\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{Q}}\right) \mathrm{d}\mathcal{P} \in [0, \infty], \tag{6.1}$$

if $\mathcal{P}$ is absolutely continuous with respect to $\mathcal{Q}$, and as $H(\mathcal{P} \,|\, \mathcal{Q}) := \infty$, if this is not the case.

Let us now consider the σ-finite measure Q on $\mathscr{B}(\mathbb{R}^n)$ with density $q(x) = \mathrm{e}^{-2\Psi(x)}$, as introduced in (2.2) of Chapter 2. Following the approach of [Léo14b, Section 3], we shall demonstrate that the same definition of relative entropy $H(P \,|\, \mathrm{Q})$ applies to the reference measure Q, provided that the probability measure P is an element of $\mathscr{P}_2(\mathbb{R}^n)$, with the only difference that the quantity (6.1) now takes values in $(-\infty, \infty]$.

To this end, we let $P \in \mathscr{P}_2(\mathbb{R}^n)$. The nonnegativity of the potential Ψ implies that for the function $\mathbb{R}^n \ni x \mapsto f(x) := \mathrm{e}^{-|x|^2}$ we have $\mathbb{E}_{\mathrm{Q}}[f] \in (0, \infty)$. Following [Léo14b, Section 3], we let $\mathcal{Q}$ be the probability measure on $\mathscr{B}(\mathbb{R}^n)$ having probability density function $f/\mathbb{E}_{\mathrm{Q}}[f]$ with respect to the measure Q, so that

$$\frac{\mathrm{d}P}{\mathrm{dQ}} = \frac{f}{\mathbb{E}_{\mathrm{Q}}[f]} \frac{\mathrm{d}P}{\mathrm{d}\mathcal{Q}}.$$

Taking first logarithms and then expectations with respect to P on both sides of this equation yields the formula

$$H(P \,|\, \mathrm{Q}) = H(P \,|\, \mathcal{Q}) - \int_{\mathbb{R}^n} |x|^2 \,\mathrm{d}P(x) - \log\left(\int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2 - 2\Psi(x)} \,\mathrm{d}x\right), \tag{6.2}$$

which is justified by (6.1) and the fact that $P \in \mathscr{P}_2(\mathbb{R}^n)$ as well as $\mathbb{E}_{\mathrm{Q}}[f] \in (0, \infty)$. In particular, we see that the right-hand side of (6.2) takes values in the interval $(-\infty, \infty]$.

Summing up, we can define well the relative entropy $H(P\,|\,\mathrm{Q})$ as in (6.1) provided that $P \in \mathscr{P}_2(\mathbb{R}^n)$, even when the σ-finite measure Q has infinite total mass.

Remark 6.1. Wherever we consider the relative entropy $H(P\,|\,\mathrm{Q})$ with respect to the σ-finite measure Q on $\mathscr{B}(\mathbb{R}^n)$ with density $\mathbb{R}^n \ni x \mapsto q(x) = \mathrm{e}^{-2\Psi(x)}$, the probability measure P will always be assumed to belong to $\mathscr{P}_2(\mathbb{R}^n)$. This is in accordance with Lemmas 2.3 and 3.2, as well as Lemma 4.10. In the latter, the constant-speed geodesic $(P_t)_{0\leqslant t\leqslant 1}$ joining two probability measures P_0 and P_1 in $\mathscr{P}_2(\mathbb{R}^n)$ is considered. Therefore, in all situations relevant to us, the relative entropy $H(P\,|\,\mathrm{Q})$ is well-defined and takes values in the interval $(-\infty, \infty]$. ⋄

6.2. The proofs of Lemmas 2.3 and 3.2

Proof of Lemma 2.3. Let the real constants $c, R \geqslant 0$ be as in condition (iii) of Assumptions 2.2, and denote

$$m_R := \max_{|x|\leqslant R} \bigl|n - \langle x\,,\, 2\,\nabla\Psi(x)\rangle\bigr| < \infty\,, \qquad \tau_k := \inf\,\{t \geqslant 0\colon\ |X(t)| > k\} \tag{6.3}$$

for integers $k > R$. Itô's formula gives

$$\mathrm{d}|X(t)|^2 = \Bigl(n - \Bigl\langle X(t)\,,\, 2\,\nabla\Psi\bigl(X(t)\bigr)\Bigr\rangle\Bigr)\,\mathrm{d}t + \langle 2\,X(t)\,,\,\mathrm{d}W(t)\rangle \tag{6.4}$$

for $t \geqslant 0$. We define $\varphi_k(t) := \mathbb{E}_{\mathbb{P}}\bigl[\,|X(t\wedge\tau_k)|^2\,\bigr]$ and $\varphi(t) := \mathbb{E}_{\mathbb{P}}\bigl[\,|X(t)|^2\,\bigr]$. Taking expectations in (6.4) yields

$$\varphi_k(t) = \varphi(0) + \mathbb{E}_{\mathbb{P}}\biggl[\int_0^{t\wedge\tau_k} \Bigl(n - \Bigl\langle X(u)\,,\, 2\,\nabla\Psi\bigl(X(u)\bigr)\Bigr\rangle\Bigr)\,\mathbb{1}_{\{|X(u)|\leqslant R\}}\,\mathrm{d}u\biggr] \tag{6.5}$$

$$+\,\mathbb{E}_{\mathbb{P}}\biggl[\int_0^{t\wedge\tau_k} \Bigl(n - \Bigl\langle X(u)\,,\, 2\,\nabla\Psi\bigl(X(u)\bigr)\Bigr\rangle\Bigr)\,\mathbb{1}_{\{|X(u)|>R\}}\,\mathrm{d}u\biggr] \tag{6.6}$$

$$\leqslant \varphi(0) + m_R\,\mathbb{E}_{\mathbb{P}}[t\wedge\tau_k] + \mathbb{E}_{\mathbb{P}}\biggl[\int_0^{t\wedge\tau_k} \bigl(n + 2\,c\,|X(u)|^2\bigr)\,\mathrm{d}u\biggr]$$

$$\leqslant \varphi(0) + (m_R + n)\,t + 2\,c\int_0^t \varphi_k(u)\,\mathrm{d}u.$$

The Gronwall inequality gives now

$$\varphi_k(t) \leqslant g(t) := \varphi(0) + (m_R + n)\,t + 2\,c\int_0^t \bigl(\varphi(0) + (m_R + n)\,u\bigr)\,\mathrm{e}^{2c(t-u)}\,\mathrm{d}u. \tag{6.7}$$

According to the second-moment condition in (2.6), the quantity $g(t)$ is finite for all $t \geqslant 0$, and independent of k; letting $k \uparrow \infty$ in (6.7), we get

$$\varphi(t) = \mathbb{E}_{\mathbb{P}}\bigl[\,|X(t)|^2\,\bigr] \leqslant g(t) < \infty, \qquad t \geqslant 0.$$

In other words, we have that $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$ for all $t \geqslant 0$. □

Proof of Lemma 3.2. The proof of Lemma 3.2 follows by analogy with the proof of Lemma 2.3 above. Indeed, one just has to add the perturbation β to the gradient $\nabla\Psi$ in the expressions (6.3) – (6.6), write $W^\beta(t)$ instead of $W(t)$ in (6.4), and replace all the $\mathbb{P}$-expectations by expectations with respect to the probability measure $\mathbb{P}^\beta$. Then the constant m_R in (6.3) is still finite and, because of its compact support, the perturbation β in the expression (6.6) vanishes, provided R is chosen large enough. Hence, by the same token as above, we conclude that $P^\beta(t) \in \mathscr{P}_2(\mathbb{R}^n)$ for all $t \geqslant t_0$. □

6.3. Some results in probability and martingale theory

In the proofs of Propositions 4.5 and 4.7 we have used a result about conditional expectations, which we formulate below as Proposition 6.3. We place ourselves on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a left-continuous filtration $(\mathcal{F}(t))_{t\geqslant 0}$. We first recall the following result, which is well known under the name of *Scheffé's lemma* [Wil91, 5.10].

Lemma 6.2 (Scheffé's lemma)**.** *For a sequence of integrable random variables $(X_n)_{n\in\mathbb{N}}$ which converges $\mathbb{P}$-a.s. to another integrable random variable X, the convergence of the $L^1(\mathbb{P})$-norms, i.e.,*

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n|\big] = \mathbb{E}\big[|X|\big],$$

is equivalent to the convergence in $L^1(\mathbb{P})$, i.e.,

$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|\big] = 0.$$

Proposition 6.3. *Let $(B(t))_{0\leqslant t\leqslant T}$ and $(C(t))_{0\leqslant t\leqslant T}$ be adapted continuous processes, which are nonnegative and uniformly bounded, respectively. Define the process*

$$A(t) := \int_0^t \big(B(u) + C(u)\big)\,\mathrm{d}u, \qquad 0 \leqslant t \leqslant T$$

and assume that $\mathbb{E}\big[\int_0^T B(u)\,\mathrm{d}u\big]$ is finite. By the Lebesgue differentiation theorem, for Lebesgue-a.e. $t_0 \in [0,T]$ we have

$$\lim_{t\uparrow t_0} \frac{\mathbb{E}[A(t_0) - A(t)]}{t_0 - t} = \lim_{t\uparrow t_0} \frac{1}{t_0 - t}\,\mathbb{E}\left[\int_t^{t_0} \big(B(u) + C(u)\big)\,\mathrm{d}u\right] = \mathbb{E}[B(t_0) + C(t_0)]. \quad (6.8)$$

Now fix a "Lebesgue point" $t_0 \in [0,T]$ for which (6.8) does hold. Then we have the analogous limiting assertion for the conditional expectations, i.e.,

$$\lim_{t\uparrow t_0} \frac{\mathbb{E}[A(t_0) - A(t) \mid \mathcal{F}(t)]}{t_0 - t} = B(t_0) + C(t_0), \quad (6.9)$$

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

Proof. We fix a Lebesgue point $t_0 \in [0,T]$ for which the limiting assertion (6.8) does hold. As the processes $(B(t))_{0\leqslant t\leqslant T}$ and $(C(t))_{0\leqslant t\leqslant T}$ are continuous, the fundamental theorem of calculus ensures that the limit

$$\begin{aligned}\lim_{t\uparrow t_0} \frac{A(t_0)-A(t)}{t_0-t} &= \lim_{t\uparrow t_0} \frac{1}{t_0-t}\int_t^{t_0} \big(B(u)+C(u)\big)\,\mathrm{d}u \\ &= B(t_0)+C(t_0)\end{aligned} \tag{6.10}$$

exists almost surely. Since the random variables appearing in (6.10) are integrable, and we already have the convergence of the $L^1(\mathbb{P})$-norms from (6.8), Lemma 6.2 allows us to conclude that the convergence of (6.10) holds also in $L^1(\mathbb{P})$, i.e.,

$$\lim_{t\uparrow t_0} \left\| \frac{A(t_0)-A(t)}{t_0-t} - \big(B(t_0)+C(t_0)\big)\right\|_{L^1(\mathbb{P})} = 0. \tag{6.11}$$

From (6.11) we deduce now the $L^1(\mathbb{P})$-convergence of (6.9) as follows. For $t<t_0$ we have the estimate

$$\begin{aligned}&\left\| \frac{\mathbb{E}\big[A(t_0)-A(t)\mid \mathcal{F}(t)\big]}{t_0-t} - \big(B(t_0)+C(t_0)\big)\right\|_{L^1(\mathbb{P})} \\ &\quad\leqslant \left\| \frac{A(t_0)-A(t)}{t_0-t} - \big(B(t_0)+C(t_0)\big)\right\|_{L^1(\mathbb{P})} \qquad (6.12)\\ &\qquad + \Big\| \mathbb{E}\big[B(t_0)+C(t_0)\,\big|\,\mathcal{F}(t)\big] - \big(B(t_0)+C(t_0)\big)\Big\|_{L^1(\mathbb{P})}. \qquad (6.13)\end{aligned}$$

According to (6.11), the expression of (6.12) converges to zero as $t\uparrow t_0$. The same is true also for the term in (6.13), on account of [Chu01, Theorem 9.4.8] and the left-continuity of the filtration $(\mathcal{F}(t))_{t\geqslant 0}$. This completes the proof of Proposition 6.3. □

In the proof of Theorem 4.1 we invoked the following result in martingale theory.

Proposition 6.4. *Suppose $(N(t))_{t\geqslant 0}$ is a strictly positive local martingale with continuous paths. Let τ be a $[0,\infty)$-valued stopping time such that $\log N(\tau)$ is integrable and $\mathbb{E}[(\log N(0))^+]<\infty$. Then $\log N(0)$ is integrable, and*

$$\mathbb{E}\big[\log N(\tau)\big] - \mathbb{E}\big[\log N(0)\big] = -\tfrac{1}{2}\,\mathbb{E}\big[\langle \log N\rangle(\tau)\big],$$

where $\langle \log N\rangle$ denotes the quadratic variation of $(\log N(t))_{t\geqslant 0}$.

Proof. We apply Lemma 2.48 in [KK21] to the continuous, strictly positive local martingale $\tilde{N}(t) := N(t)/N(0)$, $t\geqslant 0$. □

6.4. The proof of Lemma 4.10

Proof of Lemma 4.10. In order to show (4.40), we recall the notation of (4.38) and consider the time-dependent velocity field

$$[0,1]\times\mathbb{R}^n\ni(t,\xi)\longmapsto v_t(\xi):=\gamma((T_t^\gamma)^{-1}(\xi))\in\mathbb{R}^n,$$

which is well-defined P_t-a.s. for every $t\in[0,1]$. Then $(v_t)_{0\leqslant t\leqslant 1}$ is the velocity field associated with $(T_t^\gamma)_{0\leqslant t\leqslant 1}$, i.e.,

$$T_t^\gamma(x)=x+\int_0^t v_\theta(T_\theta^\gamma(x))\,\mathrm{d}\theta,$$

on account of (4.38). Let $p_t(\,\cdot\,)$ be the probability density function of the probability measure P_t in (4.38). Then, according to [Vil03, Theorem 5.34], the function $p_t(\,\cdot\,)$ satisfies the continuity equation

$$\partial_t p_t(x)+\operatorname{div}\big(v_t(x)\,p_t(x)\big)=0,\qquad (t,x)\in(0,1)\times\mathbb{R}^n,$$

which can be written equivalently as

$$-\partial_t p_t(x)=\operatorname{div}\big(v_t(x)\big)\,p_t(x)+\big\langle v_t(x)\,,\nabla p_t(x)\big\rangle,\qquad (t,x)\in(0,1)\times\mathbb{R}^n. \tag{6.14}$$

Recall that X_0 is a random variable with probability distribution P_0 on the probability space $(S,\mathcal{S},\nu)$. Then the integral equation

$$X_t=X_0+\int_0^t v_\theta(X_\theta)\,\mathrm{d}\theta,\qquad 0\leqslant t\leqslant 1 \tag{6.15}$$

defines random variables X_t with probability distributions $P_t=(T_t^\gamma)_\#(P_0)$ for $t\in[0,1]$, as in (4.38). We have

$$\mathrm{d}p_t(X_t)=\partial_t p_t(X_t)\,\mathrm{d}t+\big\langle\nabla p_t(X_t)\,,\,\mathrm{d}X_t\big\rangle=-p_t(X_t)\operatorname{div}\big(v_t(X_t)\big)\,\mathrm{d}t$$

on account of (6.14), (6.15), thus also

$$\mathrm{d}\log p_t(X_t)=-\operatorname{div}\big(v_t(X_t)\big)\,\mathrm{d}t,\qquad 0\leqslant t\leqslant 1. \tag{6.16}$$

Recall the function $q(x)=\mathrm{e}^{-2\Psi(x)}$, for which

$$\mathrm{d}\log q(X_t)=-\big\langle 2\,\nabla\Psi(X_t)\,,\,\mathrm{d}X_t\big\rangle=-\big\langle 2\,\nabla\Psi(X_t)\,,\,v_t(X_t)\big\rangle\,\mathrm{d}t. \tag{6.17}$$

For the likelihood ratio function $\ell_t(\,\cdot\,)$ of (4.37) we get from (6.16) and (6.17) that

$$\mathrm{d}\log\ell_t(X_t)=\big\langle 2\,\nabla\Psi(X_t)\,,\,v_t(X_t)\big\rangle\,\mathrm{d}t\;-\;\operatorname{div}\big(v_t(X_t)\big)\,\mathrm{d}t,\qquad 0\leqslant t\leqslant 1. \tag{6.18}$$

Taking expectations in the integral version of (6.18), we obtain that the difference

$$H(P_t\,|\,\mathrm{Q})-H(P_0\,|\,\mathrm{Q})=\mathbb{E}_\nu\big[\log\ell_t(X_t)\big]-\mathbb{E}_\nu\big[\log\ell_0(X_0)\big]$$

is equal to

$$\mathbb{E}_\nu\Big[\int_0^t \Big(\langle 2\,\nabla\Psi(X_\theta)\,,\,v_\theta(X_\theta)\rangle - \operatorname{div}\big(v_\theta(X_\theta)\big)\Big)\,\mathrm{d}\theta\Big]$$

for $t \in [0,1]$. Consequently,

$$\lim_{t\downarrow 0}\frac{H(P_t\,|\,\mathrm{Q}) - H(P_0\,|\,\mathrm{Q})}{t} = \mathbb{E}_\nu\Big[\langle 2\,\nabla\Psi(X_0)\,,\,v_0(X_0)\rangle - \operatorname{div}\big(v_0(X_0)\big)\Big]. \tag{6.19}$$

Integrating by parts, we see that

$$\mathbb{E}_\nu\big[\operatorname{div}\big(v_0(X_0)\big)\big] = \int_{\mathbb{R}^n} \operatorname{div}\big(v_0(x)\big)\,p_0(x)\,\mathrm{d}x \tag{6.20}$$

$$= -\int_{\mathbb{R}^n} \langle v_0(x)\,,\nabla p_0(x)\rangle\,\mathrm{d}x \tag{6.21}$$

$$= -\big\langle \nabla \log p_0(X_0)\,,\,v_0(X_0)\big\rangle_{L^2(\nu)}. \tag{6.22}$$

Recalling (6.19), and combining it with the relation $\nabla \log \ell_t(x) = \nabla \log p_t(x) + 2\,\nabla\Psi(x)$, as well as with (6.20) – (6.22), we get

$$\lim_{t\downarrow 0}\frac{H(P_t\,|\,\mathrm{Q}) - H(P_0\,|\,\mathrm{Q})}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\,v_0(X_0)\big\rangle_{L^2(\nu)}.$$

Since $v_0 = \gamma$, this leads to (4.40). □

Part II.

A variational characterization of Langevin–Smoluchowski diffusions

Joint work with Ioannis Karatzas

Abstract

We show that Langevin–Smoluchowski measure on path space is invariant under time-reversal, followed by stochastic control of the drift with a novel entropic-type criterion. Repeated application of these forward-backward steps leads to a sequence of stochastic control problems, whose initial/terminal distributions converge to the Gibbs probability measure of the diffusion, and whose values decrease to zero along the relative entropy of the Langevin–Smoluchowski flow with respect to this Gibbs measure.

1. Introduction

Diffusions of Langevin–Smoluchowski type have some important properties. They possess invariant (Gibbs) probability measures described very directly in terms of their potentials and towards which, under appropriate conditions, their time-marginals converge as time increases to infinity and in a manner that conforms to the second law of thermodynamics: the relative entropy of the current distribution, with respect to the invariant one, decreases to zero. The seminal paper [JKO98] revealed another remarkable, local aspect of this decrease towards equilibrium: the family of time-marginals is, at (almost) every point in time, a curve of steepest descent among all probability density functions with finite second moment, when distances are measured according to the Wasserstein metric in configuration space.

We establish in this work yet another variational property of such diffusions, this time a global one: their law is invariant under the combined effects of time-reversal, and of stochastic control of the drift under a novel, entropic-type criterion. Here, one minimizes over admissible controls the relative entropy of the "terminal" state with respect to the invariant measure, plus an additional term thought of as "entropic cost of time-reversal": the difference in relative entropy with respect to the Langevin–Smoluchowski measure on path space, computed based on the "terminal" state as opposed to on the entire path. Quite similar, but different, cost criteria have been considered in [DPP89, DPP90, FS85, Pav89, PW91].

The setting under consideration bears similarities to the celebrated Schrödinger bridge problem, but also considerable differences. Both problems are posed on a fixed time horizon of finite length, and both involve the relative entropy with respect to the invariant measure. But here this entropy is modified by the addition of the above-mentioned entropic cost of time-reversal, and there is no fixed, target distribution on the terminal state. Yet the trajectory that emerges as the solution of the stochastic control problem has time-marginals that replicate exactly those of the original Langevin–Smoluchowski flow, whence the "invariance" property mentioned in the abstract.

We refer to the literature [CGP16a, CGP16b, CGP21, Léo14a] for overviews on the classical Schrödinger bridge problem, to [TT13] for the related semimartingale transport problem, and to the recent paper [BCGL20] for a detailed study of the mean-field Schrödinger problem. A related controllability problem for a Fokker–Planck equation and its connection to Schrödinger systems and stochastic optimal control, is considered in [Bla92]. More information about the Schrödinger equation, diffusion theory, and time reversal, can be found in the book [Nag93].

1.1. Preview

In Chapter 2 we introduce the Langevin–Smoluchowski measure $\mathbb{P}$ on path space, under which the canonical process $(X(t))_{t\geqslant 0}$ has dynamics (2.1) with initial distribution $P(0)$. Then, in Section 2.3, this process is studied under time-reversal. That is, we fix a terminal time $T \in (0,\infty)$ and consider the time-reversed process $\overline{X}(s) = X(T-s)$, $0 \leqslant s \leqslant T$. Classical theory of time reversal shows that $\overline{X}$ is again a diffusion, and gives an explicit description of its dynamics.

Chapter 3 develops our main result, Theorem 3.1. An equivalent change of probability measure $\mathbb{P}^{\gamma} \sim \mathbb{P}$ adds to the drift of the time-reversed process $\overline{X}$ a measurable, adapted process $\boldsymbol{\gamma}(T-s)$, $0 \leqslant s \leqslant T$. In broad brushes, this allows us to define, in terms of relative entropies, the quantities

$$H^{\gamma} := H\big(\mathbb{P}^{\gamma} \,|\, \mathbb{Q}\big)\Big|_{\boldsymbol{\sigma}(\overline{X}(T))}, \qquad D^{\gamma} := H\big(\mathbb{P}^{\gamma} \,|\, \mathbb{P}\big)\Big|_{\boldsymbol{\sigma}(\overline{X})} - H\big(\mathbb{P}^{\gamma} \,|\, \mathbb{P}\big)\Big|_{\boldsymbol{\sigma}(\overline{X}(T))}. \tag{1.1}$$

Here, $\mathbb{Q}$ denotes the probability measure on path space, inherited from the Langevin–Smoluchowski dynamics (2.1) with initial distribution given by the invariant Gibbs probability measure Q. Theorem 3.1 establishes then the variational characterization

$$\inf_{\gamma} \big(H^{\gamma} + D^{\gamma}\big) = H\big(P(T) \,|\, \mathrm{Q}\big), \tag{1.2}$$

where $P(T)$ denotes the distribution of the random variable $X(T)$ under $\mathbb{P}$. The process $\boldsymbol{\gamma}_*$ that realizes the infimum in (1.2) gives rise to a probability measure $\mathbb{P}^{\gamma_*}$, under which the time-reversed diffusion $\overline{X}$ is of Langevin–Smoluchowski type in its own right, but now with initial distribution $P(T)$. In other words, with the constraint of minimizing the sum of the entropic quantities H^{γ} and D^{γ} of (1.1), Langevin–Smoluchowski measure on path space is invariant under time-reversal.

Chapters 4 – 6 develop ramifications of the main result, including the following consistency property: starting with the time-reversal $\overline{X}$ of the Langevin–Smoluchowski diffusion, the solution of a related optimization problem, whose value is now $H(P(2T) \,|\, \mathrm{Q})$, leads to the original forward Langevin–Smoluchowski dynamics as in (2.1), but now with initial distribution $P(2T)$. Iterating these procedures we obtain an alternating sequence of forward-backward Langevin–Smoluchowski dynamics with initial distributions $(P(kT))_{k\in\mathbb{N}_0}$ converging to Q in total variation, along which the values of the corresponding optimization problems as in (1.2) are given by $(H(P(kT) \,|\, \mathrm{Q}))_{k\in\mathbb{N}}$ and decrease to zero.

2. The setting

Let us consider a Langevin–Smoluchowski diffusion process $(X(t))_{t\geqslant 0}$ of the form

$$\mathrm{d}X(t) = -\nabla\Psi\big(X(t)\big)\,\mathrm{d}t + \mathrm{d}W(t), \tag{2.1}$$

with values in $\mathbb{R}^n$. Here $(W(t))_{t\geqslant 0}$ is standard n-dimensional Brownian motion, and the "potential" $\Psi\colon \mathbb{R}^n \to [0,\infty)$ is a C^∞-function growing, along with its derivatives of all orders, at most exponentially as $|x| \to \infty$; we stress that no convexity assumptions are imposed on this potential. We posit also an "initial condition" $X(0) = \Xi$, a random variable independent of the driving Brownian motion and with given distribution $P(0)$. For concreteness, we shall assume that this initial distribution has a continuous probability density function $p_0(\,\cdot\,)$.

Under these conditions, the Langevin–Smoluchowski equation (2.1) admits a pathwise unique, strong solution, up until an "explosion time" $\mathfrak{e}$; such explosion never happens, i.e., $\mathbb{P}(\mathfrak{e} = \infty) = 1$, if in addition the second moment condition (2.10) and the coercivity condition (2.9) below hold. The condition (2.9) propagates the finiteness of the second moment to the entire collection of time-marginal distributions $P(t) = \mathrm{Law}(X(t))$, $t \geqslant 0$, which are then determined uniquely. In fact, adapting the arguments in [Rog85] to the present situation, we check that each time-marginal distribution $P(t)$ has probability density $p(t,\,\cdot\,)$ such that the resulting function $(t,x) \mapsto p(t,x)$ is continuous and strictly positive on $(0,\infty)\times\mathbb{R}^n$; differentiable with respect to the temporal variable t for each $x \in \mathbb{R}^n$; smooth in the spatial variable x for each $t > 0$; and such that the logarithmic derivative $(t,x) \mapsto \nabla \log p(t,x)$ is continuous on $(0,\infty)\times\mathbb{R}^n$. These arguments also lead to the *Fokker–Planck* [Fri75, Gar09, Ris96, Sch80], or *forward Kolmogorov* [Kol31], equation

$$\partial_t p(t,x) = \tfrac{1}{2}\Delta p(t,x) + \mathrm{div}\,\big(\nabla\Psi(x)\,p(t,x)\big), \qquad (t,x) \in (0,\infty)\times\mathbb{R}^n \tag{2.2}$$

with initial condition

$$p(0,x) = p_0(x), \qquad x \in \mathbb{R}^n.$$

Here and throughout this work, ∂_t denotes differentiation with respect to the temporal argument; whereas ∇, Δ and div stand, respectively, for gradient, Laplacian and divergence with respect to the spatial argument.

2.1. Invariant measure, likelihood ratio, and relative entropy

We introduce now the function

$$q(x) := \mathrm{e}^{-2\Psi(x)}, \qquad x \in \mathbb{R}^n \tag{2.3}$$

and note that it satisfies the stationary version

$$\tfrac{1}{2}\Delta q(x) + \operatorname{div}\left(\nabla\Psi(x)\, q(x)\right) = 0, \qquad x \in \mathbb{R}^n \tag{2.4}$$

of the forward Kolmogorov equation (2.2). We introduce also the σ-finite measure Q on the Borel subsets $\mathscr{B}(\mathbb{R}^n)$ of $\mathbb{R}^n$, which has density q as in (2.3) with respect to n-dimensional Lebesgue measure. This measure Q is invariant for the diffusion of (2.1); see Exercise 5.6.18 in [KS98]. When finite, Q can be normalized to an invariant probability measure for (2.1), to which the time-marginals $P(t)$ "converge" as $t \to \infty$; more about this convergence can be found in Chapter 6. We shall always assume tacitly that such a normalization has taken place, when Q is finite, i.e., when

$$\mathrm{Q}(\mathbb{R}^n) = \int_{\mathbb{R}^n} q(x)\, \mathrm{d}x = \int_{\mathbb{R}^n} \mathrm{e}^{-2\Psi(x)}\, \mathrm{d}x < \infty. \tag{2.5}$$

One way to think of the above-mentioned convergence, is by considering the likelihood ratio

$$\ell(t,x) := \frac{p(t,x)}{q(x)}, \qquad (t,x) \in (0,\infty) \times \mathbb{R}^n. \tag{2.6}$$

It follows from (2.2), (2.4) that this function satisfies the *backward* Kolmogorov equation

$$\partial_t \ell(t,x) = \tfrac{1}{2}\Delta\ell(t,x) - \left\langle \nabla\ell(t,x)\,, \nabla\Psi(x)\right\rangle, \qquad (t,x) \in (0,\infty) \times \mathbb{R}^n. \tag{2.7}$$

In terms of the likelihood ratio function (2.6), let us consider now the relative entropy

$$H\big(P(t)\,|\,\mathrm{Q}\big) := \mathbb{E}_{\mathbb{P}}\big[\log \ell(t, X(t))\big] = \int_{\mathbb{R}^n} \log\left(\frac{p(t,x)}{q(x)}\right) p(t,x)\, \mathrm{d}x, \qquad t \geqslant 0 \tag{2.8}$$

of the probability distribution $P(t)$ with respect to the invariant measure Q. The expectation in (2.8) is well-defined in $[0,\infty]$, if Q is a probability measure. As we are not imposing this as a blanket assumption, we shall rely on Section 6.1 of Part I, where it is shown that the relative entropy $H(P(t)\,|\,\mathrm{Q})$ is well-defined and takes values in $(-\infty,\infty]$, whenever $P(t)$ belongs to the space $\mathscr{P}_2(\mathbb{R}^n)$ of probability measures with finite second moment (see also [Léo14b] for a more general discussion). This, in turn, is the case whenever $P(0)$ has finite second moment, and the coercivity condition

$$\forall\, x \in \mathbb{R}^n, |x| \geqslant R\colon \qquad \left\langle x\,, \nabla\Psi(x)\right\rangle \geqslant -c\,|x|^2 \tag{2.9}$$

is satisfied by the potential Ψ in (2.1), for some real constants $c \geqslant 0$ and $R \geqslant 0$; see the first problem on page 125 of [Fri75], or Lemma 2.3 and Section 6.2 of Part I. The prototypical such potential is $\Psi(x) = \frac{1}{2}|x|^2$, leading to Ornstein–Uhlenbeck dynamics in (2.1); but $\Psi \equiv 0$ and the "double well" $\Psi(x) = (x^2 - a^2)^2$ for $a > 0$, are also allowed. In particular, the coercivity condition (2.9) does not imply that the potential Ψ is convex.

We shall impose throughout Chapters 2 – 5 the coercivity condition (2.9), as well as the finite second moment condition

$$\int_{\mathbb{R}^n} |x|^2\, p_0(x)\, \mathrm{d}x < \infty. \tag{2.10}$$

This amounts to $P(0) \in \mathscr{P}_2(\mathbb{R}^n)$, as has been already alluded to. In Chapter 6 we will see that these two conditions (2.9) and (2.10) are not needed when Q is a probability measure.

However, *we shall impose throughout the entire work the crucial assumption that the initial relative entropy is finite*, i.e.,

$$H(P(0) \,|\, \mathrm{Q}) = \int_{\mathbb{R}^n} \log\left(\frac{p_0(x)}{q(x)}\right) p_0(x)\, \mathrm{d}x < \infty. \tag{2.11}$$

Under either the conditions "(2.9) + (2.10)", or the condition (2.5), the decrease of the relative entropy function

$$[0,\infty) \ni t \longmapsto H(P(t) \,|\, \mathrm{Q}) \in (-\infty, \infty] \tag{2.12}$$

implies then that the quantity $H(P(t) \,|\, \mathrm{Q})$ in (2.8) is finite for all $t \geqslant 0$ whenever (2.11) holds. The decrease of the function (2.12) is a classical aspect of thermodynamics. For a proof of this fact under the conditions "(2.9) + (2.10)" and without assuming finiteness of the invariant measure Q, see Theorem 3.1 of Part I; when Q is a probability measure, we refer to Section 6.1.

In fact, under the conditions (2.9) – (2.11), the rate of decrease for the relative entropy, measured with respect to distances traveled in $\mathscr{P}_2(\mathbb{R}^n)$ in terms of the quadratic Wasserstein metric

$$W_2(\mu,\nu) := \Big(\inf_{Y\sim\mu, Z\sim\nu} \mathbb{E}\big[|Y-Z|^2\big] \Big)^{1/2}, \qquad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^n)$$

(cf. [AG13, AGS08, Vil03]) is, at Lebesgue-almost all times $t_0 \in [0,\infty)$, the *steepest possible* along the Langevin–Smoluchowski curve $(P(t))_{t\geqslant 0}$ of probability measures. Here, we are comparing the curve $(P(t))_{t\geqslant 0}$ against all such curves $(P^\beta(t))_{t\geqslant t_0}$ of probability measures generated as in (2.1) — but with an additional drift ∇B for suitable (smooth and compactly supported) perturbations B of the potential Ψ in (2.1). This local optimality property of Langevin–Smoluchowski diffusions is due to [JKO98]; it was established in Theorem 3.3 of Part I in the form just described. We develop in this work yet another, global this time, optimality property for such diffusions.

2.2. The probabilistic setting

In (2.8) and throughout this work, we are denoting by $\mathbb{P}$ the unique probability measure on the space $\Omega = C([0,\infty);\mathbb{R}^n)$ of continuous, $\mathbb{R}^n$-valued functions, under which the canonical coordinate process $X(t,\omega) = \omega(t)$, $t \geqslant 0$ has the property that

$$W(t) := X(t) - X(0) + \int_0^t \nabla\Psi(X(\theta))\, \mathrm{d}\theta, \qquad t \geqslant 0 \tag{2.13}$$

is standard $\mathbb{R}^n$-valued Brownian motion, and independent of the random variable $X(0)$ with distribution

$$\mathbb{P}[X(0) \in A] = \int_A p_0(x)\, \mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n).$$

The $\mathbb{P}$-Brownian motion $(W(t))_{t\geqslant 0}$ of (2.13) is adapted to, in fact generates, the canonical filtration $\mathbb{F} = (\mathcal{F}(t))_{t\geqslant 0}$ with

$$\mathcal{F}(t) := \boldsymbol{\sigma}\big(X(s)\colon\ 0 \leqslant s \leqslant t\big). \tag{2.14}$$

By analogy to the terminology "Wiener measure", we call the probability measure $\mathbb{P}$ the "Langevin–Smoluchowski measure" associated with the potential Ψ.

2.3. Reversal of time

The densities $p(t,\,\cdot\,)$ and $q(\,\cdot\,)$ satisfy the forward Kolmogorov equations (2.2) and (2.4), respectively. Whereas, their likelihood ratio $\ell(t,\,\cdot\,)$ in (2.6) satisfies the *backward* Kolmogorov equation (2.7). This suggests that, in the study of relative entropy and of its temporal dissipation, it might make sense to look at the underlying Langevin–Smoluchowski diffusion under time-reversal. Such an approach proved very fruitful in [DPP89], [Pav89], [FJ16] and [KST22]; it will be important in our context here as well.

Thus, we fix an arbitrary terminal time $T \in (0,\infty)$ and consider the time-reversed process

$$\overline{X}(s) := X(T-s), \qquad \overline{\mathcal{G}}(s) := \boldsymbol{\sigma}\big(\overline{X}(u)\colon\ 0 \leqslant u \leqslant s\big); \qquad 0 \leqslant s \leqslant T \tag{2.15}$$

along with the filtration $\overline{\mathbb{G}} = (\overline{\mathcal{G}}(s))_{0\leqslant s\leqslant T}$ this process generates. Then standard theory on time-reversal shows that the process

$$\overline{W}(s) := \overline{X}(s) - \overline{X}(0) + \int_0^s \Big(\nabla\Psi\big(\overline{X}(u)\big) - \nabla L\big(T-u, \overline{X}(u)\big)\Big)\,\mathrm{d}u, \tag{2.16}$$

for $0 \leqslant s \leqslant T$, is a $\overline{\mathbb{G}}$-standard Brownian motion under $\mathbb{P}$, with values in $\mathbb{R}^n$ and independent of $\overline{X}(0) = X(T)$ (see, for instance, [Föl85, Föl86], [HP86], [Mey94], [Nel01], and [Par86] for the classical results; an extensive presentation of the relevant facts regarding the time reversal of diffusion processes can be found in Part V). Here

$$L(t,x) := \log \ell(t,x), \qquad (t,x) \in (0,\infty)\times\mathbb{R}^n$$

is the logarithm of the likelihood ratio function in (2.6); and on the strength of (2.7), this function solves the semilinear Schrödinger-type equation

$$\partial_t L(t,x) = \tfrac{1}{2}\Delta L(t,x) - \big\langle \nabla L(t,x)\,,\nabla\Psi(x)\big\rangle + \tfrac{1}{2}|\nabla L(t,x)|^2. \tag{2.17}$$

Another way to express this, is by saying that the so-called *Hopf–Cole transform* $\ell = \mathrm{e}^L$ turns the semilinear equation (2.17), into the linear backward Kolmogorov equation (2.7). This observation is not new; it has been used in stochastic control to good effect by Fleming [Fle77, Fle82], Holland [Hol77], and in a context closer in spirit to this work by Dai Pra and Pavon [DPP90], Dai Pra [DP91].

3. A stochastic control problem

Yet another way to cast the equation (2.17), is in the *Hamilton–Jacobi–Bellman* form

$$\partial_t L(t,x) = \tfrac{1}{2}\Delta L(t,x) - \big\langle \nabla L(t,x)\,,\nabla\Psi(x)\big\rangle - \min_{g\in\mathbb{R}^n}\Big(\big\langle \nabla L(t,x)\,,\,g\big\rangle + \tfrac{1}{2}|g|^2\Big), \tag{3.1}$$

where the minimization is attained by the gradient $g_* = -\nabla L(t,x)$. This, in turn, suggests a *stochastic control problem* related to the backwards diffusive dynamics

$$\mathrm{d}\overline{X}(s) = \Big(\nabla L\big(T-s,\overline{X}(s)\big) - \nabla\Psi\big(\overline{X}(s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}(s) \tag{3.2}$$

of (2.16), which we formulate now as follows.

For any measurable process $[0,T]\times\Omega \ni (t,\omega)\mapsto \boldsymbol{\gamma}(t,\omega)\in\mathbb{R}^n$ such that the time-reversed process $\boldsymbol{\gamma}(T-s)$, $0\leqslant s\leqslant T$ is adapted to the backward filtration $\overline{\mathbb{G}}$ of (2.15), and which satisfies the condition

$$\mathbb{P}\bigg[\int_0^T |\boldsymbol{\gamma}(T-s)|^2\,\mathrm{d}s < \infty\bigg] = 1, \tag{3.3}$$

we consider the exponential $(\overline{\mathbb{G}},\mathbb{P})$-local martingale

$$Z^{\boldsymbol{\gamma}}(s) := \exp\bigg(\int_0^s \big\langle \boldsymbol{\gamma}(T-u)\,,\,\mathrm{d}\overline{W}(u)\big\rangle - \tfrac{1}{2}\int_0^s |\boldsymbol{\gamma}(T-u)|^2\,\mathrm{d}u\bigg), \tag{3.4}$$

for $0\leqslant s\leqslant T$. We denote by $\boldsymbol{\Gamma}$ the collection of all processes $\boldsymbol{\gamma}$ as above, for which $Z^{\boldsymbol{\gamma}}$ is a true $(\overline{\mathbb{G}},\mathbb{P})$-martingale. This collection is not empty: it contains all such uniformly bounded processes $\boldsymbol{\gamma}$, and quite a few more (e.g., conditions of Novikov [KS98, Corollary 3.5.13] and Kazamaki [RY99, Proposition VIII.1.14]).

Now, for every $\boldsymbol{\gamma}\in\boldsymbol{\Gamma}$, we introduce an equivalent probability measure $\mathbb{P}^{\boldsymbol{\gamma}}\sim\mathbb{P}$ on path space, via

$$\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\bigg|_{\overline{\mathcal{G}}(s)} = Z^{\boldsymbol{\gamma}}(s), \qquad 0\leqslant s\leqslant T. \tag{3.5}$$

Then, by the Girsanov theorem [KS98, Theorem 3.5.1], the process

$$\overline{W}^{\boldsymbol{\gamma}}(s) := \overline{W}(s) - \int_0^s \boldsymbol{\gamma}(T-u)\,\mathrm{d}u, \qquad 0\leqslant s\leqslant T \tag{3.6}$$

is standard $\mathbb{R}^n$-valued $\mathbb{P}^{\boldsymbol{\gamma}}$-Brownian motion of the filtration $\overline{\mathbb{G}}$, thus independent of the random variable $\overline{X}(0) = X(T)$. Under the probability measure $\mathbb{P}^{\boldsymbol{\gamma}}$, the backwards dynamics of (3.2) take the form

$$\mathrm{d}\overline{X}(s) = \Big(\nabla L\big(T-s,\overline{X}(s)\big) + \boldsymbol{\gamma}(T-s) - \nabla\Psi\big(\overline{X}(s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\boldsymbol{\gamma}}(s); \tag{3.7}$$

and it follows readily from this decomposition and the semilinear parabolic equation (2.17), that the process

$$M^{\boldsymbol{\gamma}}(s) := L(T-s, \overline{X}(s)) + \tfrac{1}{2}\int_0^s |\boldsymbol{\gamma}(T-u)|^2 \,\mathrm{d}u, \qquad 0 \leqslant s \leqslant T \tag{3.8}$$

is a local $\overline{\mathbb{G}}$-submartingale under $\mathbb{P}^{\boldsymbol{\gamma}}$, with decomposition

$$\mathrm{d}M^{\boldsymbol{\gamma}}(s) = \tfrac{1}{2}\big|\nabla L(T-s, \overline{X}(s)) + \boldsymbol{\gamma}(T-s)\big|^2 \,\mathrm{d}s + \Big\langle \nabla L(T-s, \overline{X}(s)) \,,\, \mathrm{d}\overline{W}^{\boldsymbol{\gamma}}(s) \Big\rangle. \tag{3.9}$$

In fact, introducing the $\overline{\mathbb{G}}$-stopping times

$$\sigma_n := \inf\left\{ s \geqslant 0 \colon \int_0^s \Big(\big|\nabla L(T-u, \overline{X}(u))\big|^2 + |\boldsymbol{\gamma}(T-u)|^2 \Big) \,\mathrm{d}u \geqslant n \right\} \wedge T \tag{3.10}$$

for $n \in \mathbb{N}_0$, with $\sigma_n \uparrow T$, we see that the stopped process $M^{\boldsymbol{\gamma}}(\,\cdot \wedge \sigma_n)$ is a $\overline{\mathbb{G}}$-submartingale under $\mathbb{P}^{\boldsymbol{\gamma}}$, for every $n \in \mathbb{N}_0$. In particular, we observe

$$\begin{aligned} H\big(P(T) \,|\, \mathrm{Q}\big) &= \mathbb{E}_{\mathbb{P}}\big[L(T, X(T))\big] = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\big[L(T, \overline{X}(0))\big] \\ &\leqslant \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\bigg[L(T-\sigma_n, \overline{X}(\sigma_n)) + \tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}(T-u)|^2 \,\mathrm{d}u \bigg], \end{aligned} \tag{3.11}$$

since $\mathbb{P}^{\boldsymbol{\gamma}} = \mathbb{P}$ on $\overline{\mathcal{G}}(0) = \boldsymbol{\sigma}(\overline{X}(0)) = \boldsymbol{\sigma}(X(T))$. Now (3.11) holds for every $n \in \mathbb{N}_0$, thus

$$H\big(P(T) \,|\, \mathrm{Q}\big) \leqslant \liminf_{n\to\infty} \; \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\bigg[L(T-\sigma_n, \overline{X}(\sigma_n)) + \tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}(T-u)|^2 \,\mathrm{d}u \bigg]. \tag{3.12}$$

But as we remarked already, the minimum in (3.1) is attained by $g_* = -\nabla L(t,x)$; likewise, the drift term in (3.9) vanishes, if we select the process $\boldsymbol{\gamma}_* \in \boldsymbol{\Gamma}$ via

$$\boldsymbol{\gamma}_*(t,\omega) := -\nabla L(t, \omega(t)), \qquad \text{thus} \qquad \boldsymbol{\gamma}_*(T-s) = -\nabla L(T-s, \overline{X}(s)), \tag{3.13}$$

for $0 \leqslant s \leqslant T$. With this choice, the backwards dynamics of (3.7) take the form

$$\mathrm{d}\overline{X}(s) = -\nabla \Psi\big(\overline{X}(s)\big) \,\mathrm{d}s + \mathrm{d}\overline{W}^{\boldsymbol{\gamma}_*}(s); \tag{3.14}$$

that is, *precisely of the Langevin–Smoluchowski type* (2.1), but now with the "initial condition" $\overline{X}(0) = X(T)$ and independent driving $\overline{\mathbb{G}}$-Brownian motion $\overline{W}^{\boldsymbol{\gamma}_*}$, under $\mathbb{P}^{\boldsymbol{\gamma}_*}$. Since $\mathbb{P}^{\boldsymbol{\gamma}_*} = \mathbb{P}$ holds on the $\boldsymbol{\sigma}$-algebra $\overline{\mathcal{G}}(0) = \boldsymbol{\sigma}(\overline{X}(0)) = \boldsymbol{\sigma}(X(T))$, the initial distribution of $\overline{X}(0)$ under $\mathbb{P}^{\boldsymbol{\gamma}_*}$ is equal to $P(T)$. Furthermore, with $\boldsymbol{\gamma} = \boldsymbol{\gamma}_*$, the process of (3.8), (3.9) becomes a $\mathbb{P}^{\boldsymbol{\gamma}_*}$-local martingale, namely

$$M^{\boldsymbol{\gamma}_*}(s) = L(T, X(T)) + \int_0^s \Big\langle \nabla L(T-u, \overline{X}(u)) \,,\, \mathrm{d}\overline{W}^{\boldsymbol{\gamma}_*}(u) \Big\rangle; \tag{3.15}$$

and we have equality in (3.11), thus also

$$H\big(P(T) \,|\, \mathrm{Q}\big) = \lim_{n\to\infty} \; \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}_*}}\bigg[L(T-\sigma_n, \overline{X}(\sigma_n)) + \tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}_*(T-u)|^2 \,\mathrm{d}u \bigg]. \tag{3.16}$$

We conclude that the infimum over $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ of the right-hand side in (3.12) is attained by the process $\boldsymbol{\gamma}_*$ of (3.13), which gives rise to the Langevin–Smoluchowski dynamics (3.14) for the time-reversed process $\overline{X}(s) = X(T-s)$, $0 \leqslant s \leqslant T$, under $\mathbb{P}^{\boldsymbol{\gamma}_*}$. We formalize this discussion as follows.

Theorem 3.1. *Consider the stochastic control problem of minimizing over the class* $\boldsymbol{\Gamma}$ *of measurable, adapted processes* $\boldsymbol{\gamma}$ *satisfying* (3.3) *and inducing an exponential martingale* $Z^{\boldsymbol{\gamma}}$ *in* (3.4)*, with the notation of* (3.10) *and with the backwards dynamics of* (3.7)*, the expected cost*

$$\mathcal{I}(\boldsymbol{\gamma}) := \liminf_{n\to\infty} \; \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\Big[L(T-\sigma_n, \overline{X}(\sigma_n)) + \tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}(T-u)|^2 \,\mathrm{d}u\Big]. \tag{3.17}$$

Under the assumptions of Chapter 2, the infimum $\inf_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \mathcal{I}(\boldsymbol{\gamma})$ *is equal to the relative entropy* $H(P(T)\,|\,\mathrm{Q})$ *and is attained by the "score process"* $\boldsymbol{\gamma}_*$ *of* (3.13)*. This choice leads to the backwards Langevin–Smoluchowski dynamics* (3.14)*, and with* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_*$ *the limit in* (3.17) *exists as in* (3.16)*.*

Proof. It only remains to check that the minimizing process of (3.13) belongs indeed to the collection $\boldsymbol{\Gamma}$ of admissible processes. By its definition, this process $\boldsymbol{\gamma}_*$ is measurable, and its time-reversal is adapted to the backward filtration $\overline{\mathbb{G}}$ of (2.15). Theorem 4.1 of Part I gives

$$\mathbb{E}_{\mathbb{P}}\Big[\int_0^T \big|\nabla L(T-u, \overline{X}(u))\big|^2 \,\mathrm{d}u\Big] = \mathbb{E}_{\mathbb{P}}\Big[\int_0^T \big|\nabla L(\theta, X(\theta))\big|^2 \,\mathrm{d}\theta\Big] < \infty, \tag{3.18}$$

which implies *a fortiori* that the condition in (3.3) is satisfied for $\boldsymbol{\gamma} = \boldsymbol{\gamma}_*$.

We must also show that the process $Z^{\boldsymbol{\gamma}_*}$ defined in the manner of (3.4), is a true martingale. A very mild dose of stochastic calculus leads to

$$\mathrm{d}L(T-s, \overline{X}(s)) = \Big\langle \nabla L(T-s, \overline{X}(s)) \,,\, \mathrm{d}\overline{W}(s)\Big\rangle + \tfrac{1}{2}\big|\nabla L(T-s, \overline{X}(s))\big|^2 \,\mathrm{d}s$$

on account of (2.17), (3.2). Therefore, we have

$$\begin{aligned}
&\int_0^s \langle \boldsymbol{\gamma}_*(T-u) \,,\, \mathrm{d}\overline{W}(u)\rangle - \tfrac{1}{2}\int_0^s |\boldsymbol{\gamma}_*(T-u)|^2 \,\mathrm{d}u \\
&\qquad = -\int_0^s \Big\langle \nabla L(T-u, \overline{X}(u)) \,,\, \mathrm{d}\overline{W}(u)\Big\rangle - \tfrac{1}{2}\int_0^s \big|\nabla L(T-u, \overline{X}(u))\big|^2 \,\mathrm{d}u \\
&\qquad = L(T, X(T)) - L(T-s, \overline{X}(s)) = \log\Big(\frac{\ell(T, X(T))}{\ell(T-s, \overline{X}(s))}\Big),
\end{aligned}$$

which expresses the exponential process of (3.4) with $\boldsymbol{\gamma} = \boldsymbol{\gamma}_*$ as

$$Z^{\boldsymbol{\gamma}_*}(s) = \frac{\ell(T, X(T))}{\ell(T-s, \overline{X}(s))}, \qquad 0 \leqslant s \leqslant T.$$

Now, let us argue that the process Z^{γ_*} is a true $(\overline{\mathbb{G}}, \mathbb{P})$-martingale. It is a positive local martingale, thus a supermartingale. It will be a martingale, if it has constant expectation. But $Z^{\gamma_*}(0) \equiv 1$, so it is enough to show that $\mathbb{E}_{\mathbb{P}}[Z^{\gamma_*}(T)] = 1$. Let us denote by $P(s, y; t, \xi)$ the transition kernel of the Langevin–Smoluchowski dynamics, so that

$$\mathbb{P}[X(s) \in \mathrm{d}y, X(t) \in \mathrm{d}\xi] = p(s, y)\, P(s, y; t, \xi)\, \mathrm{d}y\, \mathrm{d}\xi$$

for $0 \leqslant s < t \leqslant T$ and $(y, \xi) \in \mathbb{R}^n \times \mathbb{R}^n$. Then the invariance of Q gives

$$\int_{\mathbb{R}^n} q(y)\, P(0, y; T, \xi)\, \mathrm{d}y = q(\xi), \qquad \xi \in \mathbb{R}^n; \tag{3.19}$$

consequently

$$\mathbb{E}_{\mathbb{P}}\big[Z^{\gamma_*}(T)\big] = \mathbb{E}_{\mathbb{P}}\left[\frac{p(T, X(T))}{q(X(T))} \frac{q(X(0))}{p(0, X(0))}\right] \tag{3.20}$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{p(T, \xi)}{q(\xi)} \frac{q(y)}{p(0, y)}\, p(0, y)\, P(0, y; T, \xi)\, \mathrm{d}y\, \mathrm{d}\xi \tag{3.21}$$

$$= \int_{\mathbb{R}^n} \frac{p(T, \xi)}{q(\xi)} \left(\int_{\mathbb{R}^n} q(y)\, P(0, y; T, \xi)\, \mathrm{d}y \right) \mathrm{d}\xi \tag{3.22}$$

$$= \int_{\mathbb{R}^n} p(T, \xi)\, \mathrm{d}\xi = 1, \tag{3.23}$$

implying that Z^{γ_*} is a true martingale and completing the proof of Theorem 3.1. □

Results related to Theorem 3.1 have been established in the publications [DPP89, DPP90, FS85, Pav89, PW91].

Remark 3.2 (Reincarnation of time-marginals). Let us denote by $\overline{P}_*(s)$ the distribution of the random variable $\overline{X}(s) = X(T - s)$ under the probability measure $\mathbb{P}^{\gamma_*}$, for $0 \leqslant s \leqslant T$. Since $(\overline{X}(s))_{0 \leqslant s \leqslant T}$ is under $\mathbb{P}^{\gamma_*}$ a Langevin–Smoluchowski diffusion in its own right, we deduce

$$\overline{P}_*(s) = P(T + s), \qquad 0 \leqslant s \leqslant T \tag{3.24}$$

on the strength of uniqueness in distribution for the Langevin–Smoluchowski flow, and of its time-homogeneity. In other words, the branch $(P(T + s))_{0 \leqslant s \leqslant T}$ of the original Langevin–Smoluchowski curve of time-marginals, gets "reincarnated" as $(\overline{P}_*(s))_{0 \leqslant s \leqslant T}$, the curve of time-marginals arising from the solution of the stochastic control problem in Theorem 3.1. But now, under the probability measure $\mathbb{P}^{\gamma_*}$, the states of the Langevin–Smoluchowski diffusion $(\overline{X}(s))_{0 \leqslant s \leqslant T}$ corresponding to the curve $(\overline{P}_*(s))_{0 \leqslant s \leqslant T}$ traverse the time interval $[0, T]$ in the opposite temporal direction. ◇

3.1. Entropic interpretation of the expected cost when $\mathrm{Q}(\mathbb{R}^n) < \infty$

Let us observe from (3.4) – (3.6) that

$$\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\Big|_{\overline{\mathcal{G}}(\sigma_n)}\right) = \int_0^{\sigma_n} \langle \boldsymbol{\gamma}(T-u)\,,\, \mathrm{d}\overline{W}^{\boldsymbol{\gamma}}(u)\rangle + \tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}(T-u)|^2\,\mathrm{d}u \tag{3.25}$$

holds for every $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $n \in \mathbb{N}_0$. Thus, as the $\mathbb{P}^{\boldsymbol{\gamma}}$-expectation of the stochastic integral in (3.25) vanishes, the expected quadratic cost, or "energy", term in (3.17) is itself a relative entropy:

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\tfrac{1}{2}\int_0^{\sigma_n} |\boldsymbol{\gamma}(T-u)|^2\,\mathrm{d}u\right] = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\Big|_{\overline{\mathcal{G}}(\sigma_n)}\right)\right].$$

By contrast, when Q is a probability measure on $\mathscr{B}(\mathbb{R}^n)$, and denoting by $\mathbb{Q}$ the probability measure induced on $\Omega = C([0,\infty);\mathbb{R}^n)$ by the canonical process driven by the dynamics (2.1) with Q as the distribution of $X(0)$, the first term in (3.17) can be cast as

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[L(T-\sigma_n, \overline{X}(\sigma_n))\right] &= \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(\overline{X}(\sigma_n))}\right)\right] \\ &= \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(\overline{X}(\sigma_n))}\right)\right] - \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\Big|_{\boldsymbol{\sigma}(\overline{X}(\sigma_n))}\right)\right].\end{aligned} \tag{3.26}$$

It follows that, in this case, the expected cost of (3.17) is equal to the sum $H_n^{\boldsymbol{\gamma}} + D_n^{\boldsymbol{\gamma}}$ of two nonnegative quantities:

$$H_n^{\boldsymbol{\gamma}} := \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(\overline{X}(\sigma_n))}\right)\right],$$

the relative entropy of the probability measure $\mathbb{P}^{\boldsymbol{\gamma}}$ with respect to the probability measure $\mathbb{Q}$ when both are restricted to the $\boldsymbol{\sigma}$-algebra generated by the random variable $\overline{X}(\sigma_n)$; and

$$D_n^{\boldsymbol{\gamma}} := \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\Big|_{\overline{\mathcal{G}}(\sigma_n)}\right)\right] - \mathbb{E}_{\mathbb{P}^{\boldsymbol{\gamma}}}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\boldsymbol{\gamma}}}{\mathrm{d}\mathbb{P}}\Big|_{\boldsymbol{\sigma}(\overline{X}(\sigma_n))}\right)\right], \tag{3.27}$$

the difference between the relative entropies of the probability measure $\mathbb{P}^{\boldsymbol{\gamma}}$ with respect to the probability measure $\mathbb{P}$, when restricted to the $\boldsymbol{\sigma}$-algebra generated by the collection of random variables $(\overline{X}(u\wedge\sigma_n))_{0\leqslant u\leqslant T}$ and by the random variable $\overline{X}(\sigma_n)$, respectively. The difference in (3.27) is nonnegative, because conditioning on a smaller $\boldsymbol{\sigma}$-algebra can only decrease the relative entropy; this difference can be thought of as an "entropic cost of time-reversal".

It develops from this discussion that the expected cost on the right-hand side of (3.17) is nonnegative, when Q is a probability measure.

4. From local to square-integrable martingales

Whenever the process M^{γ} of (3.8), (3.9) happens to be a true submartingale under $\mathbb{P}^{\gamma}$ (as, for instance, with $\boldsymbol{\gamma} \equiv 0$ on account of Theorem 4.1 of Part I), the expected cost (3.17) takes the form

$$\mathbb{E}_{\mathbb{P}^{\gamma}}\Big[L(0, X(0)) + \tfrac{1}{2}\int_0^T |\boldsymbol{\gamma}(T-u)|^2 \,\mathrm{d}u\Big].$$

Likewise, we derive from Theorem 3.1 the identity

$$H\big(P(T) \,|\, \mathrm{Q}\big) = \mathbb{E}_{\mathbb{P}^{\gamma_*}}\Big[L(0, X(0)) + \tfrac{1}{2}\int_0^T \big|\nabla L(T-u, \overline{X}(u))\big|^2 \,\mathrm{d}u\Big],$$

whenever the process M^{γ_*} of (3.15) is a true $\mathbb{P}^{\gamma_*}$-martingale. This is the case, for instance, whenever

$$\mathbb{E}_{\mathbb{P}^{\gamma_*}}\Big[\langle M^{\gamma_*}, M^{\gamma_*}\rangle(T)\Big] = \mathbb{E}_{\mathbb{P}^{\gamma_*}}\bigg[\int_0^T \big|\nabla L(T-u, \overline{X}(u))\big|^2 \,\mathrm{d}u\bigg] < \infty \tag{4.1}$$

holds; and then the stochastic integral in (3.15) is an $L^2(\mathbb{P}^{\gamma_*})$-bounded martingale (see, for instance, [KS98, Proposition 3.2.10] or [RY99, Corollary IV.1.25]). Using (3.24), we can express the expectation of (4.1) more explicitly as

$$\int_{\mathbb{R}^n}\int_0^T |\nabla \log \ell(t,x)|^2 \, p(2T-t, x) \,\mathrm{d}t \,\mathrm{d}x. \tag{4.2}$$

The shift in the temporal variable makes it difficult to check whether the quantity in (4.2) is finite and one cannot directly apply arguments similar to those in Theorem 4.1 of Part I, where the expectation (4.1) is taken with respect to the probability measure $\mathbb{P}$, in the manner of (3.18) (and thus, the argument $2T-t$ in (4.2) is replaced by t). This problem is consonant with the fact that the term in (3.26) is not quite a relative entropy, but a linear combination of relative entropies.

The goal of this chapter is to find a square-integrable $\mathbb{P}^{\gamma_*}$-martingale $\overline{M}^{\gamma_*}$, which is closely related to the local martingale M^{γ_*} of (3.15). The idea is to correct the shift in the temporal variable appearing in (4.2), by reversing time once again. First, we need to introduce some notation. We denote by $\overline{p}_*(s, \cdot\,)$ the probability density function of the random variable $\overline{X}(s) = X(T-s)$ under the probability measure $\mathbb{P}^{\gamma_*}$, for $0 \leqslant s \leqslant T$. From (3.24), we deduce the relation

$$\overline{p}_*(s,x) = p(T+s, x), \qquad (s,x) \in [0,T] \times \mathbb{R}^n.$$

The associated likelihood ratio function and its logarithm are defined respectively by

$$\lambda(s,x) := \frac{\overline{p}_*(s,x)}{q(x)}, \qquad \Lambda(s,x) := \log \lambda(s,x) = L(T+s,x), \tag{4.3}$$

for $(s,x) \in [0,T] \times \mathbb{R}^n$. From the definition (4.3), and the equations (2.17), (3.1), we see that the function $(s,x) \mapsto \Lambda(s,x)$ satisfies again the semilinear Schrödinger-type equation

$$-\partial_s \Lambda(s,x) + \tfrac{1}{2}\Delta\Lambda(s,x) - \langle \nabla\Lambda(s,x)\,,\nabla\Psi(x)\rangle = \min_{b\in\mathbb{R}^n}\Big(\langle \nabla\Lambda(s,x)\,,\,b\rangle + \tfrac{1}{2}|b|^2\Big). \tag{4.4}$$

In the setting introduced above, the relative entropy with respect to Q of the distribution $\overline{P}_*(s)$ of $\overline{X}(s)$ under $\mathbb{P}^{\gamma_*}$ is

$$H\big(\overline{P}_*(s)\,|\,\mathrm{Q}\big) = \mathbb{E}_{\mathbb{P}^{\gamma_*}}\big[\Lambda(s,\overline{X}(s))\big] = \int_{\mathbb{R}^n} \log\left(\frac{\overline{p}_*(s,x)}{q(x)}\right)\overline{p}_*(s,x)\,\mathrm{d}x. \tag{4.5}$$

Again, the assumption that $H(P(0)\,|\,\mathrm{Q})$ is finite, and the decrease of the relative entropy function

$$[0,\infty) \ni t \longmapsto H\big(P(t)\,|\,\mathrm{Q}\big) \in (-\infty,\infty],$$

imply that the relative entropy in (4.5) is finite, for all $0 \leqslant s \leqslant T$. Finally, the relative Fisher information of $\overline{P}_*(s)$ with respect to Q is defined as

$$I\big(\overline{P}_*(s)\,|\,\mathrm{Q}\big) := \mathbb{E}_{\mathbb{P}^{\gamma_*}}\Big[\,\big|\nabla\Lambda(s,\overline{X}(s))\big|^2\,\Big] = \int_{\mathbb{R}^n} \big|\nabla\Lambda(s,x)\big|^2\,\overline{p}_*(s,x)\,\mathrm{d}x. \tag{4.6}$$

4.1. Reversing time once again

Let us consider on the filtered probability space $(\Omega,\mathbb{F},\mathbb{P}^{\gamma_*})$ the canonical process $X(t)$, for $0 \leqslant t \leqslant T$, whose time-reversal (2.15) satisfies the backwards Langevin–Smoluchowski dynamics (3.14). Reversing time once again, we find that the process $X(t) = \overline{X}(T-t)$, for $0 \leqslant t \leqslant T$ satisfies the stochastic differential equation

$$\mathrm{d}X(t) = \Big(\nabla\Lambda(T-t,X(t)) - \nabla\Psi(X(t))\Big)\,\mathrm{d}t + \mathrm{d}W^{\gamma_*}(t), \tag{4.7}$$

where the process

$$W^{\gamma_*}(t) := \overline{W}^{\gamma_*}(T-t) - \overline{W}^{\gamma_*}(T) - \int_0^t \nabla \log \overline{p}_*(T-\theta,X(\theta))\,\mathrm{d}\theta,$$

for $0 \leqslant t \leqslant T$, is Brownian motion on $(\Omega,\mathbb{F},\mathbb{P}^{\gamma_*})$. We recall here Theorems 4.1 and 5.1 of Part V. Comparing the equation (4.7) with (2.1), we see that the $\mathbb{P}^{\gamma_*}$-Brownian motion $(W^{\gamma_*}(t))_{0\leqslant t\leqslant T}$ and the $\mathbb{P}$-Brownian motion $(W(t))_{0\leqslant t\leqslant T}$ are related via

$$W(t) = W^{\gamma_*}(t) + \int_0^t \nabla\Lambda(T-\theta,X(\theta))\,\mathrm{d}\theta, \qquad 0 \leqslant t \leqslant T.$$

4.2. The dynamics of the relative entropy process

We look now at the *relative entropy process*

$$\Lambda(T-t, X(t)) = \log\left(\frac{\overline{p}_*(T-t, X(t))}{q(X(t))}\right), \qquad 0 \leqslant t \leqslant T \tag{4.8}$$

on $(\Omega, \mathbb{F}, \mathbb{P}^{\gamma_*})$. Applying Itô's formula and using the equation (4.4), together with the forward dynamics (4.7), we obtain the following result.

Proposition 4.1. *On the filtered probability space $(\Omega, \mathbb{F}, \mathbb{P}^{\gamma_*})$, the relative entropy process (4.8) is a submartingale with stochastic differential*

$$\mathrm{d}\Lambda(T-t, X(t)) = \tfrac{1}{2}\left|\nabla\Lambda(T-t, X(t))\right|^2 \mathrm{d}t + \left\langle \nabla\Lambda(T-t, X(t))\,,\, \mathrm{d}W^{\gamma_*}(t)\right\rangle. \tag{4.9}$$

In particular, the process

$$\overline{M}^{\gamma_*}(t) := \Lambda(T-t, X(t)) - \Lambda(T, X(0)) - \tfrac{1}{2}\int_0^t \left|\nabla\Lambda(T-\theta, X(\theta))\right|^2 \mathrm{d}\theta,$$

for $0 \leqslant t \leqslant T$, is an $L^2(\mathbb{P}^{\gamma_*})$-bounded martingale, with stochastic integral representation

$$\overline{M}^{\gamma_*}(t) = \int_0^t \left\langle \nabla\Lambda(T-\theta, X(\theta))\,,\, \mathrm{d}W^{\gamma_*}(\theta)\right\rangle. \tag{4.10}$$

Proof. The last thing we need to verify for the proof of Proposition 4.1, is that

$$\mathbb{E}_{\mathbb{P}^{\gamma_*}}\left[\langle \overline{M}^{\gamma_*}, \overline{M}^{\gamma_*}\rangle(T)\right] = \mathbb{E}_{\mathbb{P}^{\gamma_*}}\left[\int_0^T \left|\nabla\Lambda(T-t, X(t))\right|^2 \mathrm{d}t\right] < \infty. \tag{4.11}$$

We observe that the expectation in (4.11) is equal to

$$\mathbb{E}_{\mathbb{P}^{\gamma_*}}\left[\int_0^T \left|\nabla\Lambda(s, \overline{X}(s))\right|^2 \mathrm{d}s\right] = \mathbb{E}_{\mathbb{P}}\left[\int_T^{2T} \left|\nabla L(t, X(t))\right|^2 \mathrm{d}t\right]. \tag{4.12}$$

This is because (2.15), (4.3) give

$$\nabla\Lambda(s, \overline{X}(s)) = \nabla L(t, X(2T-t))$$

with $t = T + s \in [T, 2T]$; and because the $\mathbb{P}^{\gamma_*}$-distribution of $X(2T-t) = \overline{X}(s)$ is the same as the $\mathbb{P}$-distribution of $X(T+s) = X(t)$, on account of (3.24). But, as (3.18) holds for any finite time horizon $T > 0$, the quantity in (4.12) is finite as well. □

4.3. Relative entropy dissipation

Exploiting the trajectorial evolution of the relative entropy process (4.8), provided by Proposition 4.1, allows us to derive some immediate consequences on the decrease of the relative entropy function $[0, T] \ni s \mapsto H(\overline{P}_*(s)\,|\,\mathrm{Q}) \in (-\infty, \infty)$ and its rate of dissipation. The submartingale-property of the relative entropy process (4.8) shows once more, that this function is nondecreasing. More precisely, we have the following rate of change for the relative entropy.

Corollary 4.2. *For all $s, s_0 \geqslant 0$, we have*

$$H\big(\overline{P}_*(s) \,|\, \mathrm{Q}\big) - H\big(\overline{P}_*(s_0) \,|\, \mathrm{Q}\big) = -\tfrac{1}{2} \int_{s_0}^{s} I\big(\overline{P}_*(u) \,|\, \mathrm{Q}\big)\, \mathrm{d}u. \tag{4.13}$$

Proof. Let $s, s_0 \geqslant 0$ and choose $T \geqslant \max\{s, s_0\}$. Taking expectations under $\mathbb{P}^{\gamma_*}$ in (4.9), and noting that the stochastic integral process in (4.10) is a martingale, leads to

$$\mathbb{E}_{\mathbb{P}^{\gamma_*}}\big[\Lambda\big(s, \overline{X}(s)\big)\big] - \mathbb{E}_{\mathbb{P}^{\gamma_*}}\big[\Lambda\big(s_0, \overline{X}(s_0)\big)\big] = -\tfrac{1}{2} \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}^{\gamma_*}}\Big[\big|\nabla\Lambda\big(u, \overline{X}(u)\big)\big|^2\Big]\, \mathrm{d}u.$$

Recalling the definitions of relative entropy (4.5) and relative Fisher information (4.6), we obtain (4.13). □

Corollary 4.3. *For Lebesgue-almost every $s \geqslant 0$, the rate of relative entropy dissipation is equal to*

$$\tfrac{\mathrm{d}}{\mathrm{d}s}\, H\big(\overline{P}_*(s) \,|\, \mathrm{Q}\big) = -\tfrac{1}{2}\, I\big(\overline{P}_*(s) \,|\, \mathrm{Q}\big). \tag{4.14}$$

Proof. By the Lebesgue differentiation theorem, the monotone function

$$s \longmapsto H\big(\overline{P}_*(s) \,|\, \mathrm{Q}\big) \in (-\infty, \infty)$$

is differentiable for Lebesgue-a.e. $s \geqslant 0$, in which case (4.13) yields the identity (4.14). □

5. From backwards dynamics "back" to forward dynamics

Starting with the forward Langevin–Smoluchowski dynamics (2.1), we have seen in Chapter 3 that the combined effects of time-reversal, and of stochastic control of the drift under an entropic-type criterion, lead to the backwards dynamics

$$\mathrm{d}\overline{X}(s) = -\nabla\Psi\big(\overline{X}(s)\big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\gamma_*}(s), \qquad 0 \leqslant s \leqslant T, \tag{5.1}$$

which are again of the Langevin–Smoluchowski type, but now viewed on the filtered probability space $(\Omega, \overline{\mathbb{G}}, \mathbb{P}^{\gamma_*})$. We will see now that this universal property of Langevin–Smoluchowski measure is consistent in the following sense: starting with the backwards Langevin–Smoluchowski dynamics of (5.1), after another reversal of time, the solution of a related stochastic control problem leads to the original forward Langevin–Smoluchowski dynamics (2.1) we started with. This consistency property should come as no surprise, but its formal proof requires the results of Chapter 4, which perhaps appeared artificial at first sight.

Let us recall from equation (4.7) of Section 4.1 that reversing time in (5.1) leads to the forward dynamics

$$\mathrm{d}X(t) = \Big(\nabla\Lambda\big(T-t, X(t)\big) - \nabla\Psi\big(X(t)\big)\Big)\,\mathrm{d}t + \mathrm{d}W^{\gamma_*}(t), \tag{5.2}$$

for $0 \leqslant t \leqslant T$, on the filtered probability space $(\Omega, \mathbb{F}, \mathbb{P}^{\gamma_*})$. By analogy with Chapter 3, we define an equivalent probability measure $\Pi^{\boldsymbol{\beta}} \sim \mathbb{P}^{\gamma_*}$ as follows.

For any measurable process $[0,T]\times\Omega \ni (t,\omega) \mapsto \boldsymbol{\beta}(t,\omega) \in \mathbb{R}^n$, adapted to the forward filtration $\mathbb{F}$ of (2.14) and satisfying the condition

$$\mathbb{P}^{\gamma_*}\bigg[\int_0^T |\boldsymbol{\beta}(t)|^2\,\mathrm{d}t < \infty\bigg] = 1, \tag{5.3}$$

we consider the exponential $(\mathbb{F}, \mathbb{P}^{\gamma_*})$-local martingale

$$Z^{\boldsymbol{\beta}}(t) := \exp\bigg(\int_0^t \big\langle \boldsymbol{\beta}(\theta)\,,\, \mathrm{d}W^{\gamma_*}(\theta)\big\rangle - \tfrac{1}{2}\int_0^t |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta\bigg), \tag{5.4}$$

for $0 \leqslant t \leqslant T$. We denote by $\mathcal{B}$ the collection of all processes $\boldsymbol{\beta}$ as above, for which $Z^{\boldsymbol{\beta}}$ is a true $(\mathbb{F}, \mathbb{P}^{\gamma_*})$-martingale.

Now, for every $\boldsymbol{\beta} \in \mathcal{B}$, we introduce an equivalent probability measure $\Pi^{\boldsymbol{\beta}} \sim \mathbb{P}^{\gamma_*}$ on path space, via

$$\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\bigg|_{\mathcal{F}(t)} = Z^{\boldsymbol{\beta}}(t), \qquad 0 \leqslant t \leqslant T. \tag{5.5}$$

Then we deduce from the Girsanov theorem that, under the probability measure $\Pi^{\boldsymbol{\beta}}$, the process

$$W^{\boldsymbol{\beta}}(t) := W^{\gamma_*}(t) - \int_0^t \boldsymbol{\beta}(\theta)\,\mathrm{d}\theta, \qquad 0 \leqslant t \leqslant T \tag{5.6}$$

is $\mathbb{R}^n$-valued $\mathbb{F}$-Brownian motion, and the dynamics (5.2) become

$$\mathrm{d}X(t) = \Big(\nabla\Lambda\big(T-t, X(t)\big) + \boldsymbol{\beta}(t) - \nabla\Psi\big(X(t)\big)\Big)\,\mathrm{d}t + \mathrm{d}W^{\boldsymbol{\beta}}(t), \tag{5.7}$$

for $0 \leqslant t \leqslant T$. We couple these dynamics with the stochastic differential (4.9) and deduce that the process

$$N^{\boldsymbol{\beta}}(t) := \Lambda\big(T-t, X(t)\big) + \tfrac{1}{2}\int_0^t |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta, \qquad 0 \leqslant t \leqslant T \tag{5.8}$$

is a local $\Pi^{\boldsymbol{\beta}}$-submartingale with decomposition

$$\mathrm{d}N^{\boldsymbol{\beta}}(t) = \tfrac{1}{2}\big|\nabla\Lambda\big(T-t, X(t)\big) + \boldsymbol{\beta}(t)\big|^2\,\mathrm{d}t + \Big\langle \nabla\Lambda\big(T-t, X(t)\big)\,,\, \mathrm{d}\overline{W}^{\boldsymbol{\beta}}(t)\Big\rangle. \tag{5.9}$$

In fact, introducing the $\mathbb{F}$-stopping times

$$\tau_n := \inf\left\{t \geqslant 0\colon \int_0^t \Big(\big|\nabla\Lambda\big(T-\theta, X(\theta)\big)\big|^2 + |\boldsymbol{\beta}(\theta)|^2\Big)\,\mathrm{d}\theta \geqslant n\right\} \wedge T \tag{5.10}$$

for $n \in \mathbb{N}_0$, with $\tau_n \uparrow T$, we see that the stopped process $N^{\boldsymbol{\beta}}(\,\cdot\,\wedge\tau_n)$ is an $\mathbb{F}$-submartingale under $\Pi^{\boldsymbol{\beta}}$, for every $n \in \mathbb{N}_0$. In particular, we observe

$$\begin{aligned} H\big(P(2T)\,\big|\,\mathrm{Q}\big) = H\big(\overline{P}_*(T)\,\big|\,\mathrm{Q}\big) &= \mathbb{E}_{\mathbb{P}^{\gamma_*}}\big[\Lambda\big(T, \overline{X}(T)\big)\big] = \mathbb{E}_{\Pi^{\boldsymbol{\beta}}}\big[\Lambda\big(T, X(0)\big)\big] \\ &\leqslant \mathbb{E}_{\Pi^{\boldsymbol{\beta}}}\bigg[\Lambda\big(T-\tau_n, X(\tau_n)\big) + \tfrac{1}{2}\int_0^{\tau_n} |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta\bigg], \end{aligned} \tag{5.11}$$

since $\Pi^{\boldsymbol{\beta}} = \mathbb{P}^{\gamma_*}$ on $\mathcal{F}(0) = \boldsymbol{\sigma}(X(0)) = \boldsymbol{\sigma}(\overline{X}(T))$. Now (5.11) holds for every $n \in \mathbb{N}_0$, thus

$$H\big(P(2T)\,\big|\,\mathrm{Q}\big) \leqslant \liminf_{n\to\infty}\ \mathbb{E}_{\Pi^{\boldsymbol{\beta}}}\bigg[\Lambda\big(T-\tau_n, X(\tau_n)\big) + \tfrac{1}{2}\int_0^{\tau_n} |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta\bigg]. \tag{5.12}$$

But the minimum in (4.4) is attained by $b_* = -\nabla\Lambda(s,x)$; likewise, the drift term in (5.9) vanishes, if we select the process $\boldsymbol{\beta}_* \in \mathcal{B}$ via

$$\boldsymbol{\beta}_*(t,\omega) := -\nabla\Lambda\big(T-t, \omega(t)\big), \qquad \text{thus} \qquad \boldsymbol{\beta}_*(t) = -\nabla\Lambda\big(T-t, X(t)\big), \tag{5.13}$$

for $0 \leqslant t \leqslant T$. With this choice, the forward dynamics of (5.7) take the form

$$\mathrm{d}X(t) = -\nabla\Psi\big(X(t)\big)\,\mathrm{d}t + \mathrm{d}W^{\boldsymbol{\beta}_*}(t); \tag{5.14}$$

that is, *precisely the forward Langevin–Smoluchowski dynamics* (2.1) we started with, but now with the "initial condition" $X(0) = \overline{X}(T)$ and independent driving $\mathbb{F}$-Brownian

motion W^{β_*}, under the probability measure Π^{β_*}. Since the equality $\Pi^{\beta_*} = \mathbb{P}^{\gamma_*}$ holds on the $\boldsymbol{\sigma}$-algebra $\mathcal{F}(0) = \boldsymbol{\sigma}(X(0)) = \boldsymbol{\sigma}(\overline{X}(T))$, the initial distribution of $X(0)$ under Π^{β_*} is equal to $P(2T)$. Furthermore, with $\boldsymbol{\beta} = \boldsymbol{\beta}_*$, the process of (5.8), (5.9) becomes a Π^{β_*}-local martingale, namely

$$N^{\beta_*}(t) = \Lambda(T, X(0)) + \int_0^t \Big\langle \nabla\Lambda(T-\theta, X(\theta)) \,,\, \mathrm{d}\overline{W}^{\beta_*}(\theta) \Big\rangle; \tag{5.15}$$

and we have equality in (5.11), thus also

$$H(P(2T) \,|\, \mathrm{Q}) = \lim_{n\to\infty} \mathbb{E}_{\Pi^{\beta_*}} \Big[\Lambda(T-\tau_n, X(\tau_n)) + \tfrac{1}{2} \int_0^{\tau_n} |\boldsymbol{\beta}_*(\theta)|^2 \,\mathrm{d}\theta \Big]. \tag{5.16}$$

We conclude that the infimum over $\boldsymbol{\beta} \in \mathcal{B}$ of the right-hand side in (5.12) is attained by the process $\boldsymbol{\beta}_*$ of (5.13), which gives rise to the Langevin–Smoluchowski dynamics (5.14) for the process $(X(t))_{0\leqslant t\leqslant T}$, under Π^{β_*}. We formalize this result as follows.

Theorem 5.1. *Consider the stochastic control problem of minimizing over the class $\mathcal{B}$ of measurable, adapted processes $\boldsymbol{\beta}$ satisfying (5.3) and inducing an exponential martingale Z^{β} in (5.4), with the notation of (5.10) and with the forward dynamics of (5.7), the expected cost*

$$\mathcal{J}(\boldsymbol{\beta}) := \liminf_{n\to\infty} \; \mathbb{E}_{\Pi^{\beta}} \Big[\Lambda(T-\tau_n, X(\tau_n)) + \tfrac{1}{2} \int_0^{\tau_n} |\boldsymbol{\beta}(\theta)|^2 \,\mathrm{d}\theta \Big]. \tag{5.17}$$

Under the assumptions of Chapter 2, the infimum $\inf_{\beta\in\mathcal{B}} \mathcal{J}(\boldsymbol{\beta})$ is equal to the relative entropy $H(P(2T) \,|\, \mathrm{Q})$ and is attained by the "score process" $\boldsymbol{\beta}_*$ of (5.13). This choice leads to the forward Langevin–Smoluchowski dynamics (5.14), and with $\boldsymbol{\beta} = \boldsymbol{\beta}_*$ the limit in (5.17) exists as in (5.16).

Proof. We have to show that the minimizing process $\boldsymbol{\beta}_*$ belongs to the collection $\mathcal{B}$ of admissible processes. By its definition in (5.13), the process $\boldsymbol{\beta}_*$ is measurable, and adapted to the forward filtration $\mathbb{F}$ of (2.14). Thanks to (4.11) in Proposition 4.1, we have

$$\mathbb{E}_{\mathbb{P}^{\gamma_*}} \Big[\int_0^T |\boldsymbol{\beta}_*(t)|^2 \,\mathrm{d}t \Big] = \mathbb{E}_{\mathbb{P}^{\gamma_*}} \Big[\int_0^T \big|\nabla\Lambda(T-t, X(t))\big|^2 \,\mathrm{d}t \Big] < \infty,$$

which implies a fortiori that the condition in (5.3) is satisfied for $\boldsymbol{\beta} = \boldsymbol{\beta}_*$.

It remains to check that the process Z^{β_*} defined in the manner of (5.4), is a true martingale. From Proposition 4.1 we have the stochastic differential

$$\mathrm{d}\Lambda(T-t, X(t)) = \tfrac{1}{2}\big|\nabla\Lambda(T-t, X(t))\big|^2 \,\mathrm{d}t + \Big\langle \nabla\Lambda(T-t, X(t)) \,,\, \mathrm{d}W^{\gamma_*}(t) \Big\rangle,$$

and therefore

$$\int_0^t \langle \boldsymbol{\beta}_*(\theta)\,,\, \mathrm{d}W^{\gamma_*}(\theta)\rangle - \tfrac{1}{2}\int_0^t |\boldsymbol{\beta}_*(\theta)|^2\,\mathrm{d}\theta$$

$$= -\int_0^t \Big\langle \nabla\Lambda(T-\theta, X(\theta))\,,\, \mathrm{d}W^{\gamma_*}(\theta)\Big\rangle - \tfrac{1}{2}\int_0^t \big|\nabla\Lambda(T-\theta, X(\theta))\big|^2\,\mathrm{d}\theta$$

$$= \Lambda(T, X(0)) - \Lambda(T-t, X(t)) = \log\left(\frac{\lambda(T, X(0))}{\lambda(T-t, X(t))}\right),$$

which expresses the exponential process of (5.4) with $\boldsymbol{\beta} = \boldsymbol{\beta}_*$ as

$$Z^{\boldsymbol{\beta}_*}(t) = \frac{\lambda(T, X(0))}{\lambda(T-t, X(t))}, \qquad 0 \leqslant t \leqslant T.$$

The process $Z^{\boldsymbol{\beta}_*}$ is a positive local martingale, thus a supermartingale. To see that it is a true $(\mathbb{F}, \mathbb{P}^{\gamma_*})$-martingale, it suffices to argue that it has constant expectation. But $Z^{\boldsymbol{\beta}_*}(0) \equiv 1$, so we have to show $\mathbb{E}_{\mathbb{P}^{\gamma_*}}[Z^{\boldsymbol{\beta}_*}(T)] = 1$. We denote again by $P(s, y; t, \xi)$ the transition kernel of the Langevin–Smoluchowski dynamics, note that

$$\mathbb{P}^{\gamma_*}\big[\overline{X}(s) \in \mathrm{d}y, \overline{X}(t) \in \mathrm{d}\xi\big] = \overline{p}_*(s, y)\, P(s, y; t, \xi)\,\mathrm{d}y\,\mathrm{d}\xi$$

for $0 \leqslant s < t \leqslant T$ and $(y, \xi) \in \mathbb{R}^n \times \mathbb{R}^n$, and recall the invariance property (3.19) of Q, to deduce $\mathbb{E}_{\mathbb{P}^{\gamma_*}}[Z^{\boldsymbol{\beta}_*}(T)] = 1$ in the manner of (3.20) – (3.23). This implies that $Z^{\boldsymbol{\beta}_*}$ is a true martingale and completes the proof of Theorem 5.1. □

5.1. Entropic interpretation of the expected cost when $\mathrm{Q}(\mathbb{R}^n) < \infty$

By analogy with Section 3.1, we interpret now the expected cost on the right-hand side of (5.17) in terms of relative entropies. From (5.4) – (5.6), we deduce that

$$\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\bigg|_{\mathcal{F}(\tau_n)}\right) = \int_0^{\tau_n} \langle \boldsymbol{\beta}(\theta)\,,\, \mathrm{d}W^{\boldsymbol{\beta}}(\theta)\rangle + \tfrac{1}{2}\int_0^{\tau_n} |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta \tag{5.18}$$

holds for every $\boldsymbol{\beta} \in \mathcal{B}$ and $n \in \mathbb{N}_0$. Therefore, as the $\Pi^{\boldsymbol{\beta}}$-expectation of the stochastic integral in (5.18) vanishes, the expected quadratic cost, or "energy", term in (5.17) is equal to the relative entropy

$$\mathbb{E}_{\Pi^{\boldsymbol{\beta}}}\left[\tfrac{1}{2}\int_0^{\tau_n} |\boldsymbol{\beta}(\theta)|^2\,\mathrm{d}\theta\right] = \mathbb{E}_{\Pi^{\boldsymbol{\beta}}}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\bigg|_{\mathcal{F}(\tau_n)}\right)\right].$$

In order to interpret the first term in (5.17), let us assume that $\mathbb{Q}$, and thus also the induced measure $\mathbb{Q}$ on path space, are probability measures. Then we have

$$\begin{aligned}\mathbb{E}_{\Pi^\beta}\big[\Lambda(T-\tau_n, X(\tau_n))\big] &= \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\mathbb{P}^{\gamma_*}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(X(\tau_n))}\right)\right] \\ &= \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(X(\tau_n))}\right)\right] - \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\Big|_{\boldsymbol{\sigma}(X(\tau_n))}\right)\right].\end{aligned}$$

We conclude that, in this case, the expected cost of (5.17) is equal to the sum $H_n^{\boldsymbol{\beta}} + D_n^{\boldsymbol{\beta}}$ of two nonnegative quantities:

$$H_n^{\boldsymbol{\beta}} := \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{Q}}\Big|_{\boldsymbol{\sigma}(X(\tau_n))}\right)\right],$$

the relative entropy of the probability measure $\Pi^{\boldsymbol{\beta}}$ with respect to the probability measure $\mathbb{Q}$ when both are restricted to the $\boldsymbol{\sigma}$-algebra generated by the random variable $X(\tau_n)$; and

$$D_n^{\boldsymbol{\beta}} := \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\Big|_{\mathcal{F}(\tau_n)}\right)\right] - \mathbb{E}_{\Pi^\beta}\left[\log\left(\frac{\mathrm{d}\Pi^{\boldsymbol{\beta}}}{\mathrm{d}\mathbb{P}^{\gamma_*}}\Big|_{\boldsymbol{\sigma}(X(\tau_n))}\right)\right],$$

the difference between the relative entropies of the probability measure $\Pi^{\boldsymbol{\beta}}$ with respect to the probability measure $\mathbb{P}^{\gamma_*}$, when restricted to the $\boldsymbol{\sigma}$-algebra generated by the collection of random variables $(X(\theta\wedge\tau_n))_{0\leqslant\theta\leqslant T}$ and by the random variable $X(\tau_n)$, respectively.

6. The case of finite invariant measure, and an iterative procedure

Let us suppose now that the diffusion process $(X(t))_{t\geqslant 0}$ as in (2.1) is well-defined, along with the curve $P(t) = \mathrm{Law}(X(t))$, $t \geqslant 0$ of its time-marginals; and that the invariant measure Q of Section 2.1 is finite, i.e., (2.5) holds, and is thus normalized to a probability measure.

Then, neither the coercivity condition (2.9), nor the finite second moment condition (2.10), are needed for the results of Chapters 3 – 5. The reason is that the relative entropy $H(P(t)\,|\,\mathrm{Q})$ is now well-defined and nonnegative, as both $P(t)$ and Q are probability measures. Since the function $t \mapsto H(P(t)\,|\,\mathrm{Q})$ is decreasing and the initial relative entropy $H(P(0)\,|\,\mathrm{Q})$ is finite on account of (2.11), it follows that this function takes values in $[0,\infty)$. It can also be shown in this case that

$$\lim_{t\to\infty} \downarrow H(P(t)\,|\,\mathrm{Q}) = 0, \tag{6.1}$$

i.e., the relative entropy decreases down to zero; see [FJ16, Proposition 1.9] for a quite general version of this result. This, in turn, implies that the time-marginals $(P(t))_{t\geqslant 0}$ converge to Q in total variation as $t \to \infty$, on account of the *Pinsker–Csiszár inequality*

$$2\,\|P(t) - \mathrm{Q}\|_{\mathrm{TV}}^2 \leqslant H(P(t)\,|\,\mathrm{Q}).$$

The entropic decrease to zero is actually exponentially fast, whenever the Hessian of the potential Ψ dominates a positive multiple of the identity matrix; see, e.g., [BÉ85], [MV00, Section 5], [OV00, Proposition 1'], [Vil03, Formal Corollary 9.3], or [KST20, Remark 3.23]. As another consequence of (6.1), the initial relative entropy $H(P(0)\,|\,\mathrm{Q})$ can be expressed as

$$H(P(0)\,|\,\mathrm{Q}) = \tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\Big[\int_0^\infty |\nabla L(t, X(t))|^2\,\mathrm{d}t\Big]. \tag{6.2}$$

We prove (6.1) and (6.2) in Section 6.1.

In this context, i.e., with (2.5) replacing (2.9) and (2.10), and always under the standing assumption (2.11), Theorem 4.1 of Part I continues to hold, as do the results in Chapters 3 – 5. By combining time-reversal with stochastic control of the drift, these results lead to an alternating sequence of forward and backward Langevin–Smoluchowski dynamics, with time-marginals starting at $P(0)$ and converging along $(P(kT))_{k\in\mathbb{N}_0}$ in total variation to the invariant probability measure Q. Along the way, the values of the corresponding stochastic control problems decrease along $(H(P(kT)\,|\,\mathrm{Q}))_{k\in\mathbb{N}}$ to zero.

6.1. The decrease of the relative entropy without convexity assumption

We present a probabilistic proof of (6.1) and (6.2), which complements the proof of the more general Proposition 1.9 in [FJ16]. We stress that no convexity assumptions are imposed on the potential Ψ.

Proof of (6.1) and (6.2): Since Q is assumed to be a probability measure in Chapter 6, the relative entropy $H(P(t)\,|\,\mathrm{Q})$ is nonnegative for every $t \geqslant 0$. Thus, Corollary 4.3 of Part I gives the inequality

$$\begin{aligned} H(P(0)\,|\,\mathrm{Q}) &= H(P(T)\,|\,\mathrm{Q}) + \tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla L(t,X(t))\big|^2\,\mathrm{d}t\bigg] \\ &\geqslant \tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla L(t,X(t))\big|^2\,\mathrm{d}t\bigg] \end{aligned} \tag{6.3}$$

for every $T \in (0,\infty)$. Letting $T \uparrow \infty$ in (6.3), we deduce from the monotone convergence theorem that

$$\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\int_0^\infty \big|\nabla L(t,X(t))\big|^2\,\mathrm{d}t\bigg] \leqslant H(P(0)\,|\,\mathrm{Q}). \tag{6.4}$$

By analogy with Section 2.2, we denote by $\mathbb{Q}$ the Langevin–Smoluchowski measure associated with the potential Ψ, but now with distribution

$$\mathbb{Q}\big[X(0) \in A\big] = \mathrm{Q}(A) = \int_A q(x)\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n)$$

for the random variable $X(0)$. Since by assumption Q is a probability measure, the Langevin–Smoluchowski measure $\mathbb{Q}$ is a well-defined probability measure on path space $\Omega = C([0,\infty);\mathbb{R}^n)$.

Let us recall now the likelihood ratio of (2.6). We denote the corresponding likelihood ratio process by $\vartheta(t) := \ell(t,X(t))$, $t \geqslant 0$. The following remarkable insight is due to Pavon [Pav89], and was rediscovered by Fontbona and Jourdain [FJ16]: For any given $T \in (0,\infty)$, the time-reversed likelihood ratio process

$$\overline{\vartheta}(s) := \ell\big(T-s,\overline{X}(s)\big) = \frac{p\big(T-s,X(T-s)\big)}{q\big(\overline{X}(s)\big)}, \qquad 0 \leqslant s \leqslant T \tag{6.5}$$

is a $\mathbb{Q}$-martingale with respect to the backwards filtration $\overline{\mathbb{G}} = (\overline{\mathcal{G}}(s))_{0\leqslant s\leqslant T}$ of (2.15). For a simple proof of this result in the setting of Chapter 6 we refer to Appendix E of [KST20].

Let us pick arbitrary times $0 \leqslant t_1 < t_2 < \infty$. For any given $T \in (t_2,\infty)$, the martingale property of the process (6.5) amounts, with $s_1 = T - t_1$ and $s_2 = T - t_2$, to

$$\mathbb{E}_{\mathbb{Q}}\Big[\overline{\vartheta}(s_1)\,\Big|\,\overline{\mathcal{G}}(s_2)\Big] = \overline{\vartheta}(s_2) \qquad \Longleftrightarrow \qquad \mathbb{E}_{\mathbb{Q}}\Big[\vartheta(t_1)\,\Big|\,\boldsymbol{\sigma}\big(X(\theta)\colon t_2 \leqslant \theta \leqslant T\big)\Big] = \vartheta(t_2). \tag{6.6}$$

Because $T \in (t_2, \infty)$ is arbitrary in (6.6), this gives

$$\mathbb{E}_{\mathbb{Q}}\big[\vartheta(t_1) \,|\, \mathcal{H}(t_2)\big] = \vartheta(t_2), \qquad \mathcal{H}(t) := \boldsymbol{\sigma}\big(X(\theta) \colon t \leqslant \theta < \infty\big). \tag{6.7}$$

In other words, the likelihood ratio process $(\vartheta(t))_{t\geqslant 0}$ is a backwards $\mathbb{Q}$-martingale of the backwards filtration $\mathbb{H} = (\mathcal{H}(t))_{t\geqslant 0}$. We denote by

$$\mathcal{H}(\infty) := \bigcap_{t\geqslant 0} \mathcal{H}(t)$$

the tail $\boldsymbol{\sigma}$-algebra of the Langevin–Smoluchowski diffusion $(X(t))_{t\geqslant 0}$. The ergodicity of this process under the probability measure $\mathbb{Q}$ implies that the tail $\boldsymbol{\sigma}$-algebra $\mathcal{H}(\infty)$ is $\mathbb{Q}$-trivial, i.e., $\mathcal{H}(\infty) = \{\varnothing, \Omega\}$ modulo $\mathbb{Q}$; see Section 6.2 for a proof of this claim.

We recall now the martingale version of Theorem 9.4.7 (backwards submartingale convergence) in [Chu01]. This says that $(\vartheta(t))_{t\geqslant 0}$ is a $\mathbb{Q}$-uniformly integrable family, that the limit

$$\vartheta(\infty) := \lim_{t\to\infty} \vartheta(t) \tag{6.8}$$

exists $\mathbb{Q}$-a.e., that the convergence in (6.8) holds also in $L^1(\mathbb{Q})$, and that for every $t \geqslant 0$ we have

$$\mathbb{E}_{\mathbb{Q}}\big[\vartheta(t) \,|\, \mathcal{H}(\infty)\big] = \vartheta(\infty), \qquad \mathbb{Q}\text{-a.e.} \tag{6.9}$$

But since the tail $\boldsymbol{\sigma}$-algebra $\mathcal{H}(\infty)$ is $\mathbb{Q}$-trivial, the random variable $\vartheta(\infty)$ is $\mathbb{Q}$-a.e. constant, and (6.9) identifies this constant as $\vartheta(\infty) \equiv 1$.

In terms of the function $f(x) := x \log x$ for $x > 0$ (and with $f(0) := 0$), we can express the relative entropy $H(P(t) \,|\, \mathbb{Q})$ as

$$H(P(t) \,|\, \mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\big[\log \vartheta(t)\big] = \mathbb{E}_{\mathbb{Q}}\big[f(\vartheta(t))\big] \geqslant 0, \qquad t \geqslant 0. \tag{6.10}$$

The convexity of f, in conjunction with (6.7), shows that the process $(f(\vartheta(t)))_{t\geqslant 0}$ is a backwards $\mathbb{Q}$-submartingale of the backwards filtration $\mathbb{H}$, with decreasing expectation as in (6.10). By appealing to the backwards submartingale convergence theorem [Chu01, Theorem 9.4.7] once again, we deduce that $(f(\vartheta(t)))_{t\geqslant 0}$ is a $\mathbb{Q}$-uniformly integrable family, which converges, a.e. and in L^1 under $\mathbb{Q}$, to

$$\lim_{t\to\infty} f(\vartheta(t)) = f(\vartheta(\infty)) = f(1) = 0.$$

In particular,

$$\lim_{t\to\infty} \downarrow H(P(t) \,|\, \mathbb{Q}) = \lim_{t\to\infty} \mathbb{E}_{\mathbb{Q}}\big[f(\vartheta(t))\big] = \mathbb{E}_{\mathbb{Q}}\big[\lim_{t\to\infty} f(\vartheta(t))\big] = 0, \tag{6.11}$$

proving (6.1). From (6.11) and (6.3) it follows now that (6.4) holds as equality, proving (6.2). □

6.2. The triviality of the tail $\boldsymbol{\sigma}$-algebra $\mathcal{H}(\infty)$

We recall the filtered probability space $(\Omega, \mathcal{F}(\infty), \mathbb{F}, \mathbb{Q})$. Here, $\Omega = C([0,\infty);\mathbb{R}^n)$ is the path space of continuous functions, $\mathcal{F}(\infty) = \boldsymbol{\sigma}(\bigcup_{t\geqslant 0}\mathcal{F}(t))$, the canonical filtration $\mathbb{F} = (\mathcal{F}(t))_{t\geqslant 0}$ is as in (2.14), and the Langevin–Smoluchowski measure $\mathbb{Q}$ is represented by

$$\mathbb{Q}(B) = \int_{\mathbb{R}^n} \mathbb{P}^x(B)\,\mathrm{d}\mathrm{Q}(x), \qquad B \in \mathscr{B}(\Omega), \tag{6.12}$$

where $\mathbb{P}^x$ denotes the Langevin–Smoluchowski measure with initial distribution δ_x, for every $x \in \mathbb{R}^n$, and $\mathscr{B}(\Omega)$ is the Borel $\boldsymbol{\sigma}$-field on Ω. We note that here are several equivalent ways to define the Borel $\boldsymbol{\sigma}$-field on Ω. Two possible constructions appear in Problems 2.4.1 and 2.4.2 in [KS98].

For every $s \geqslant 0$, we define a measurable map $\theta_s\colon \Omega \to \Omega$, called *shift transformation*, by requiring that $\theta_s(\omega)(t) = \omega(s+t)$ holds for all $\omega \in \Omega$ and $t \geqslant 0$. A Borel set $B \in \mathscr{B}(\Omega)$ is called *shift-invariant* if $\theta_s^{-1}(B) = B$ holds for any $s \geqslant 0$. Since the Gibbs probability measure Q is the unique invariant measure for the Langevin–Smoluchowski diffusion $(X(t))_{t\geqslant 0}$, [Bel06, Theorem 3.8] implies that the probability measure $\mathbb{Q}$ of (6.12) is *ergodic*, meaning that $\mathbb{Q}(B) \in \{0,1\}$ holds for every shift-invariant set B. As a consequence of the ergodicity of $\mathbb{Q}$, the *Birkhoff Ergodic Theorem* [Bel06, Theorem 3.4] implies that, for every $A \in \mathscr{B}(\mathbb{R}^n)$, the limit

$$\lim_{t\to\infty} \tfrac{1}{t}\int_0^t \mathbb{1}_A(X(s))\,\mathrm{d}s = \mathrm{Q}(A) \tag{6.13}$$

exists $\mathbb{Q}$-a.e.

Lemma 6.1. *The tail $\boldsymbol{\sigma}$-algebra $\mathcal{H}(\infty)$ is $\mathbb{Q}$-trivial, i.e., $\mathcal{H}(\infty) = \{\varnothing, \Omega\}$ modulo $\mathbb{Q}$.*

Proof. We follow a reasoning similar to that in [FJ16, Remark 1.10]. According to [Kun97, Theorem 1.3.9], it suffices to show that the Langevin–Smoluchowski diffusion $(X(t))_{t\geqslant 0}$ is *recurrent in the sense of Harris*, i.e.,

$$\mathbb{P}^x\Big[\int_0^\infty \mathbb{1}_A(X(s))\,\mathrm{d}s = \infty\Big] = 1 \tag{6.14}$$

is satisfied for every $x \in \mathbb{R}^n$ and all $A \in \mathscr{B}(\mathbb{R}^n)$ with $\mathrm{Q}(A) > 0$.

For the proof of (6.14), we fix $x \in \mathbb{R}^n$ and $A \in \mathscr{B}(\mathbb{R}^n)$ with $\mathrm{Q}(A) > 0$. By its definition, the event

$$B := \Big\{\int_0^\infty \mathbb{1}_A(X(s))\,\mathrm{d}s = \infty\Big\}$$

is shift-invariant. Thus, by the ergodicity of $\mathbb{Q}$, the probability $\mathbb{Q}(B)$ is equal to either zero or one. Outside the set B, we have

$$\lim_{t\to\infty} \tfrac{1}{t}\int_0^t \mathbb{1}_A(X(s))\,\mathrm{d}s = 0.$$

But since $\mathrm{Q}(A) > 0$, the $\mathbb{Q}$-a.e. limit (6.13) implies that $\mathbb{Q}(B^{\mathrm{c}}) = 0$ and hence $\mathbb{Q}(B) = 1$. From the definition (6.12) of the probability measure $\mathbb{Q}$ it follows that $\mathbb{P}^x(B) = 1$ for Q-a.e. $x \in \mathbb{R}^n$. Since Q is equivalent to Lebesgue measure, we also have that $\mathbb{P}^x(B) = 1$ for Lebesgue-a.e. $x \in \mathbb{R}^n$.

Furthermore, the shift-invariance of B and the Markov property of the Langevin–Smoluchowski diffusion give

$$\begin{aligned} \mathbb{P}^x(B) &= \mathbb{P}^x(\theta_t^{-1}(B)) = \mathbb{E}_{\mathbb{P}^x}\big[\mathbb{P}^{X(t)}(B)\big] \\ &= T_t(\mathbb{P}^x(B)) = \int_{\mathbb{R}^n} P(0,x;t,\mathrm{d}y)\,\mathbb{P}^y(B) \end{aligned} \tag{6.15}$$

for every $t \geqslant 0$. Here, $P(0,x;t,y)$ denotes the transition kernel of the Langevin–Smoluchowski dynamics, so that $\mathbb{P}^x[X(t) \in \mathrm{d}y] = P(0,x;t,y)\,\mathrm{d}y$; and T_t denotes the operator

$$T_t f(x) := \int_{\mathbb{R}^n} P(0,x;t,\mathrm{d}y)\, f(y)\,, \qquad (t,x) \in [0,\infty) \times \mathbb{R}^n$$

acting on bounded measurable functions $f\colon \mathbb{R}^n \to \mathbb{R}$. Since $(T_t)_{t\geqslant 0}$ is a strong Feller semigroup under the assumptions of Chapter 6, the function $\mathbb{R}^n \ni x \mapsto T_t f(x)$ is continuous. Now (6.15) implies the continuity of the function $\mathbb{R}^n \ni x \mapsto \mathbb{P}^x(B)$. On the other hand, we have already seen that the function $\mathbb{R}^n \ni x \mapsto \mathbb{P}^x(B) \in [0,1]$ is Lebesgue-a.e. equal to one. But such a function is constant everywhere, i.e., $\mathbb{P}^x(B) = 1$ for every $x \in \mathbb{R}^n$, proving (6.14). □

Part III.

A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: gradient flows and HWBI inequalities

Joint work with Lane Chun Yeung

Abstract

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).

1. Introduction

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

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla(W * \mathrm{P}_t)(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T, \tag{1.1}$$

where X_0 has some given initial probability distribution P_0 on $\mathbb{R}^n$. Here, the functions $V, W\colon \mathbb{R}^n \to [0,\infty)$ play the roles of confinement and interaction potentials and are assumed to be suitably regular, $\mathrm{P}_t \coloneqq \mathrm{Law}(X_t)$ denotes the distribution of the random vector X_t, the symbol $*$ stands for the standard convolution operator, and $(B_t)_{0\leqslant t\leqslant T}$ is a standard n-dimensional Brownian motion. In particular, this stochastic differential equation is nonlocal (or nonlinear) in the sense that the drift term depends on the distribution of the state variable. Nonlocal equations of this form arise in the modeling of weakly interacting diffusion equations, after the seminal work of McKean [McK66].

Since the work of Carrillo–McCann–Villani [CMV03, CMV06], 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 [CJM$^+$01, Mal03, BGG13, Tug13, CG14, CGPS20]. In a broader context, [MMN18, HRŠS21] recently applied entropy methods to the mean-field theory of neural networks.

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

$$\mathscr{P}_{\mathrm{ac}}(\mathbb{R}^n) \ni p \longmapsto \mathscr{F}(p) \coloneqq \mathcal{U}(p) + \mathcal{V}(p) + \mathcal{W}(p) \tag{1.2}$$

is defined as the sum of the energy functionals

$$\mathcal{U}(p) \coloneqq \int_{\mathbb{R}^n} p(x)\log p(x)\,\mathrm{d}x, \tag{1.3}$$

$$\mathcal{V}(p) \coloneqq \int_{\mathbb{R}^n} V(x)\,p(x)\,\mathrm{d}x, \tag{1.4}$$

$$\mathcal{W}(p) \coloneqq \tfrac{1}{2}\int_{\mathbb{R}^n} (W * p)(x)\,p(x)\,\mathrm{d}x \tag{1.5}$$

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

$$\mathscr{P}_{\mathrm{ac}}(\mathbb{R}^n) \ni p \longmapsto \mathscr{D}(p) \coloneqq \int_{\mathbb{R}^n} |\nabla \log p(x) + \nabla V(x) + \nabla(W * p)(x)|^2\,p(x)\,\mathrm{d}x, \tag{1.6}$$

the well-known relative entropy dissipation identity takes the form

$$\mathscr{F}(p_t) - \mathscr{F}(p_{t_0}) = -\int_{t_0}^{t} \mathscr{D}(p_u)\,\mathrm{d}u. \tag{1.7}$$

This identity is of a deterministic nature: it only depends on the curve of probability density functions $(p_t)_{0\leqslant t\leqslant T}$, but not on the trajectories of the underlying process $(X_t)_{0\leqslant t\leqslant T}$ itself. It is then natural to ask whether there is a process-level analogue of the relative entropy dissipation identity (1.7), depending directly on the trajectories of the McKean–Vlasov process $(X_t)_{0\leqslant t\leqslant T}$. The main contribution of this work 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 [FJ16, KST22] based on a simpler (linear) setting without interaction, i.e., $W \equiv 0$. In this case, the McKean–Vlasov stochastic differential equation (1.1) reduces to a Langevin–Smoluchowski diffusion equation of the form

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T. \tag{1.8}$$

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 [Gar09, JK96, Ris96]) with density proportional to $\mathbb{R}^n \ni x \mapsto q(x) \coloneqq \mathrm{e}^{-V(x)}$. Defining the likelihood ratio function (or Radon–Nikodym derivative) $\ell_t(x) \coloneqq p_t(x)/q(x)$, 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

$$\log \ell_t(X_t) = \log p_t(X_t) + V(X_t), \qquad 0 \leqslant t \leqslant T \tag{1.9}$$

is called *free energy* or *relative entropy process*. As shown in [FJ16, KST22], the time-reversal

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

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

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

Here, $(M_{T-s})_{0\leqslant s\leqslant T}$ is a martingale and $(F_{T-s})_{0\leqslant s\leqslant T}$ 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 (1.7).

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

$$\log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t), \qquad 0 \leqslant t \leqslant T. \tag{1.10}$$

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 [BRTV98, Ver06, HT10, Tug13, Tug14a, Tug14b, DT16, DT18]. 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 (1.10) 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 [Lio07]. We refer to the monograph [CD18, 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 (1.10), in terms of the semimartingale decomposition

$$\log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}) - \log \ell_T(X_T, \mathrm{P}_T) = M_{T-s} + F_{T-s}, \qquad 0 \leqslant s \leqslant T, \tag{1.11}$$

where $(M_{T-s})_{0\leqslant s\leqslant T}$ is a martingale and $(F_{T-s})_{0\leqslant s\leqslant T}$ 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 (1.7) can then be recovered by taking expectations.

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*

$$\partial_t p_t(x) = \operatorname{div}\Big(\nabla p_t(x) + p_t(x)\nabla V(x) + p_t(x)\nabla(W * p_t)(x)\Big), \qquad (t,x) \in (0,T)\times\mathbb{R}^n, \tag{1.12}$$

which describes the evolution of the curve of probability density functions $(p_t)_{0\leqslant t\leqslant T}$ corresponding to the McKean–Vlasov diffusion $(X_t)_{0\leqslant t\leqslant T}$ of (1.1). When $n = 1$, this partial differential equation appears in the modeling of the time evolution of granular media [BCCP98, Vil06, CGM08]; in that context, the granular medium is modeled as a 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 (1.12) reduces to a linear *Fokker–Planck equation*. As is well known from [CMV03, CMV06], this curve of probability densities can be characterized as a *gradient flow* in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, the space of absolutely continuous probability measures with finite second moments. Roughly speaking, this is an optimality property stating that the curve $(p_t)_{0\leqslant t\leqslant T}$ evolves in the direction of

steepest possible descent for the free energy functional (1.2) with respect to the quadratic Wasserstein distance

$$W_2(\mu,\nu) \coloneqq \Big(\inf_{Y\sim\mu, Z\sim\nu} \mathbb{E}[|Y-Z|^2] \Big)^{1/2}, \qquad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^n).$$

The Wasserstein gradient flow structure of the linear Fokker–Planck equation was first discovered by Jordan, Kinderlehrer and Otto in the seminal work [JKO98]. In the paper [OV00], 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 [Ott01], 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 [AGS08]. Recently, a trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions [KST22] and Markov chains [KMS21] 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 [EFLS16].

Returning to the setting of this work, 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

$$\mathrm{d}X_t = -\big(\nabla V(X_t) + \nabla\boldsymbol{\beta}(X_t) + \nabla(W * \mathrm{P}_t^{\boldsymbol{\beta}})(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\boldsymbol{\beta}}, \qquad t_0 \leqslant t \leqslant T \quad (1.13)$$

which is constructed by adding a perturbation $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ to the confinement potential of the original McKean–Vlasov stochastic differential equation (1.1). 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 $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0\leqslant t\leqslant T}$ also evolves differently from the unperturbed curve $(\mathrm{P}_t)_{t_0\leqslant t\leqslant T}$. In parallel with the unperturbed case, we may compute the dynamics of the perturbed relative entropy process associated with (1.13). 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 $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0\leqslant t\leqslant T}$ can be computed based on the general theory of metric derivative of absolutely continuous curves, see [AGS08]. Finally, comparing these two rates in both the perturbed and unperturbed settings, allows us to establish the gradient flow property.

1.2. The HWBI inequality

The second application of our trajectorial approach deals with the HWBI inequality [AGK04, Theorem 4.2], which is an extension of the HWI inequality [OV00]. 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 $\nu,\mu \in \mathscr{P}(\mathbb{R}^n)$, the *relative entropy* of ν with respect to μ is defined by

$$H(\nu\,|\,\mu) \coloneqq \begin{cases} \int_{\mathbb{R}^n} \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu})\,\mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \quad (1.14)$$

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

$$I(\nu \,|\, \mu) := \begin{cases} \int_{\mathbb{R}^n} |\nabla \log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu})|^2 \,\mathrm{d}\mu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases} \tag{1.15}$$

and the *barycenter* of a probability measure $\nu \in \mathscr{P}_2(\mathbb{R}^n)$ is defined as

$$b(\nu) := \int_{\mathbb{R}^n} x \,\mathrm{d}\nu(x) \in \mathbb{R}^n,$$

where the integral is understood as a Bochner integral. Informally, the HWBI inequality then states that any two probability measures $\nu_0, \nu_1 \in \mathscr{P}_2(\mathbb{R}^n)$ satisfy

$$H(\nu_0 \,|\, \mu_0) - H(\nu_1 \,|\, \mu_1) \leqslant \sqrt{I(\nu_0 \,|\, \mu_0^{\uparrow})}\, W_2(\nu_0, \nu_1) - \tfrac{\kappa_V + \kappa_W}{2}\, W_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2}\, |b(\nu_0) - b(\nu_1)|^2, \tag{1.16}$$

where $\mu_0, \mu_1, \mu_0^{\uparrow}$ are some appropriate σ-finite reference measures depending on the potentials V, W (see Section 3.3 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* $(\nu_t)_{0 \leqslant t \leqslant 1}$ between ν_0 and ν_1. Compared with the HWI inequality, there are two additional terms on the right-hand side of (1.16) contributed by the interaction energy functional $\mathcal{W}$ of (1.5). Intuitively, the κ_W-uniform convexity of W leads to the first additional term $-\frac{\kappa_W}{2} W_2^2(\nu_0, \nu_1)$, which alone would correspond to the κ_W-uniform displacement convexity of $\mathcal{W}$ along $(\nu_t)_{0 \leqslant t \leqslant 1}$. But since $\mathcal{W}(p)$ is invariant under any translation of p, the functional $\mathcal{W}$ might fail to be uniformly displacement convex when the barycenter shifts. This suggests that the barycentric shift along $(\nu_t)_{0 \leqslant t \leqslant 1}$ should be factored out of the consideration of the displacement convexity of $\mathcal{W}$, which is intuitively why the second additional term $\frac{\kappa_W}{2} |b(\nu_0) - b(\nu_1)|^2$ in (1.16) appears.

Coming back to our second application, we illustrate how our approach yields a trajectorial proof of the inequality (1.16), in the slightly strengthened form of [CEGH04, Theorem 4.1] and [FK06, Theorem D.50]. Much of this consists of arguments similar in spirit to our main result (1.11), but with one key difference: instead of the time-marginals of the McKean–Vlasov diffusion, we apply the trajectorial approach to the displacement interpolation $(\nu_t)_{0 \leqslant t \leqslant 1}$. In this regard, our derivation can be seen as a generalization of the trajectorial proof of the HWI inequality in [KST22, Section 4.2]; see also [KMS21, 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 3.15) on the potentials V, W. We do not require these assumptions in the rest of this work.

In the literature, similar trajectorial approaches have also been applied in the context of martingale inequalities [ABP+13, BS15], functional inequalities [Cat04, Leh13, BCGL20, GLRT20], and their stability estimates [ELS20, EM20]. In particular, we refer to [BCGL20, Corollary 1.4] for a related HWI inequality derived from the entropic interpolation of the mean-field Schrödinger problem.

1.3. Preview

We set up the probabilistic framework and discuss some frequently used regularity assumptions in Chapter 2. In Chapter 3 we state our main trajectorial results, Theorem 3.1 and Theorem 3.7, and develop two explicit examples for illustration. As immediate consequences, we derive the classical relative entropy dissipation identities in Corollary 3.4 and Corollary 3.8. Building on these results, we formulate the gradient flow property of the granular media equation in Theorem 3.12. The HWBI inequality is then stated in Theorem 3.16. The proofs of the trajectorial results and of the HWBI inequality are developed in Chapter 4. Some proofs of auxiliary results postponed in previous chapters are contained in Chapter 5.

2. The probabilistic framework

2.1. The setting

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

As will be shown in Lemma 2.2, under the Assumptions 2.1 below, the stochastic differential equation (1.1) 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 $(B_t)_{0 \leqslant t \leqslant T}$ such that the stochastic differential equation (1.1) holds. We write $\mathbb{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant T}$ 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 $\mathbb{R}^n$. The density functions $(p_t)_{0 \leqslant t \leqslant T}$ then solve the granular media equation (1.12).

2.2. Regularity assumptions

Assumptions 2.1. The following regularity assumptions will be used frequently.

(i) The functions $V, W \colon \mathbb{R}^n \to [0, \infty)$ 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

$$\forall\, x \in \mathbb{R}^n \colon \qquad |\nabla V(x)| \leqslant C(1 + |x|), \qquad |\nabla W(x)| \leqslant C(1 + |x|). \tag{2.1}$$

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

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

These assumptions ensure that the equation (1.1) belongs to a broad class of strongly solvable McKean–Vlasov stochastic differential equations. We relegate the proof of the following result to Section 5.1.

Lemma 2.2. *Suppose Assumptions 2.1 hold. Then on an arbitrary filtered probability space, the McKean–Vlasov stochastic differential equation (1.1) has a pathwise unique, strong solution $(X_t)_{0\leqslant t\leqslant T}$ satisfying*

$$\mathbb{E}\Big[\sup_{0\leqslant t\leqslant T}|X_t|^2\Big] < \infty. \tag{2.2}$$

Moreover, its marginal distributions $(\mathrm{P}_t)_{0\leqslant t\leqslant T}$ belong to $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, and the corresponding curve of probability density functions $(p_t)_{0\leqslant t\leqslant T}$ is a classical solution of the granular media equation (1.12).

2.3. 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 $\Psi\colon \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n) \to [0,\infty)$ and its close relative $\Psi^{\uparrow}$ given by

$$\Psi(x,\mu) \coloneqq V(x) + \tfrac{1}{2}(W*\mu)(x), \qquad \Psi^{\uparrow}(x,\mu) \coloneqq V(x) + (W*\mu)(x) \tag{2.3}$$

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

$$q(x,\mu) \coloneqq \mathrm{e}^{-\Psi(x,\mu)}, \qquad q^{\uparrow}(x,\mu) \coloneqq \mathrm{e}^{-\Psi^{\uparrow}(x,\mu)}, \qquad q^{\downarrow}(x) \coloneqq \mathrm{e}^{-V(x)} \tag{2.4}$$

and the corresponding generalized likelihood ratio functions

$$\ell_t(x,\mu) \coloneqq \frac{p_t(x)}{q(x,\mu)}, \qquad \ell_t^{\uparrow}(x,\mu) \coloneqq \frac{p_t(x)}{q^{\uparrow}(x,\mu)}, \qquad \ell_t^{\downarrow}(x) \coloneqq \frac{p_t(x)}{q^{\downarrow}(x)} \tag{2.5}$$

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 $\mathscr{B}(\mathbb{R}^n)$ of $\mathbb{R}^n$, given by

$$\mathrm{Q}_t(A) \coloneqq \int_A q(x,\mathrm{P}_t)\,\mathrm{d}x, \qquad \mathrm{Q}_t^{\uparrow}(A) \coloneqq \int_A q^{\uparrow}(x,\mathrm{P}_t)\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n). \tag{2.6}$$

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 (1.8), 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 (1.2) and (1.6), in probabilistic terms: the relative entropy (defined in (1.14)) of P_t with respect to Q_t and the relative Fisher information (defined in (1.15)) of P_t with respect to $\mathrm{Q}_t^{\uparrow}$ can be expressed respectively as

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = \mathbb{E}_{\mathbb{P}}\big[\log \ell_t(X_t,\mathrm{P}_t)\big], \qquad I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}) = \mathbb{E}_{\mathbb{P}}\big[|\nabla \log \ell_t^{\uparrow}(X_t,\mathrm{P}_t)|^2\big]; \tag{2.7}$$

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*

$$\log \ell_t(X_t, \mathrm{P}_t) = \log p_t(X_t) + V(X_t) + \tfrac{1}{2}(W * \mathrm{P}_t)(X_t), \qquad 0 \leqslant t \leqslant T. \tag{2.8}$$

The dynamics of this stochastic process, together with its perturbed counterpart to be introduced in Section 3.2 below, will be our main objects of interest.

Remark 2.3. If the reference measure Q_t in (2.6) is a probability measure, then the expression (2.7) matches the classical definition of relative entropy given in (1.14). In the general case when Q_t is a σ-finite measure, the definition (1.14) 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 Section 6.1 of Part I or [Léo14b, Section 3] for the details. ⋄

3. Main results

3.1. Trajectorial dissipation of relative entropy for McKean–Vlasov diffusions

Our first main result is the semimartingale decomposition of the relative entropy process (2.8). It describes the dissipation of relative entropy along every trajectory of a particle undergoing the McKean–Vlasov dynamics (1.1). In the same spirit as the trajectorial approaches of [FJ16] and [KST22], 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

$$\overline{X}_s := X_{T-s}, \qquad 0 \leqslant s \leqslant T \tag{3.1}$$

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

In order to formulate Theorem 3.1 below, we introduce the time-reversed *Fisher information process*

$$\begin{aligned} \overline{I}_s := \Big(\big| \nabla \log \overline{\ell}^{\downarrow}_s \big|^2 + \tfrac{1}{2} \big| \nabla (W * \overline{\mathrm{P}}_s) \big|^2 + \Big\langle \tfrac{1}{2} \nabla (W * \overline{\mathrm{P}}_s) \,,\, 2 \nabla \log \overline{\ell}^{\downarrow}_s + \nabla V \Big\rangle \Big) (\overline{X}_s) \\ - \mathbb{E}_{\tilde{\mathbb{P}}} \Big[\Big\langle \tfrac{1}{2} \nabla W (\overline{X}_s - \overline{Y}_s) \,,\, \Big(2 \nabla \log \overline{\ell}^{\downarrow}_s - \nabla V + \nabla (W * \overline{\mathrm{P}}_s) \Big) (\overline{Y}_s) \Big\rangle \Big] \end{aligned} \tag{3.2}$$

for $0 \leqslant s \leqslant T$. Here, the stochastic process $(\overline{Y}_s)_{0 \leqslant s \leqslant T}$ is defined on another probability space $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}})$ such that the tuple $(\tilde{\Omega}, \tilde{\mathbb{G}}, \tilde{\mathbb{P}}, (\overline{Y}_s)_{0 \leqslant s \leqslant T})$ is an exact copy of $(\Omega, \mathbb{G}, \mathbb{P}, (\overline{X}_s)_{0 \leqslant s \leqslant T})$. A bar over a letter means that time is reversed as in (3.1).

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

$$\overline{F}_s := \int_0^s \overline{I}_u \, \mathrm{d}u, \qquad 0 \leqslant s \leqslant T. \tag{3.3}$$

This process will act as the compensator in the semimartingale decomposition of the time-reversed relative entropy process

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) = \log \overline{p}_s(\overline{X}_s) + V(\overline{X}_s) + \tfrac{1}{2} (W * \overline{\mathrm{P}}_s)(\overline{X}_s),$$

for $0 \leqslant s \leqslant T$. Its relation with the relative Fisher information (2.7) will be given in (3.7) below.

Theorem 3.1. *Suppose Assumptions 2.1 hold. On* $(\Omega, \mathbb{G}, \mathbb{P})$*, the time-reversed relative entropy process*

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) = \log \ell_{T-s}(X_{T-s}, \mathrm{P}_{T-s}), \qquad 0 \leqslant s \leqslant T \tag{3.4}$$

admits the semimartingale decomposition

$$\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) - \log \overline{\ell}_0(\overline{X}_0, \overline{\mathrm{P}}_0) = \overline{M}_s + \overline{F}_s. \tag{3.5}$$

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

$$\overline{M}_s \coloneqq \int_0^s \Big\langle \nabla \log \overline{\ell}_u(\overline{X}_u, \overline{\mathrm{P}}_u)\,,\, \sqrt{2}\,\mathrm{d}\overline{B}_u \Big\rangle, \tag{3.6}$$

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

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s I(\overline{\mathrm{P}}_u \,|\, \overline{\mathrm{Q}}_u^{\uparrow})\,\mathrm{d}u = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^s \big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2\,\mathrm{d}u\bigg] < \infty. \tag{3.7}$$

3.1.1. Examples

We give two concrete examples to illustrate Theorem 3.1.

Example 3.2. We set $n = 1$ and specialize Theorem 3.1 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 $(B_t)_{0\leqslant t\leqslant T}$ and to be normally distributed with mean 0 and variance $\sigma_0^2 > 0$. In this case, the stochastic differential equation (1.1) becomes

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T \tag{3.8}$$

and its solution is given by the Ornstein–Uhlenbeck process

$$X_t = \mathrm{e}^{-t}X_0 + \sqrt{2}\int_0^t \mathrm{e}^{u-t}\,\mathrm{d}B_u, \qquad 0 \leqslant t \leqslant T \tag{3.9}$$

with probability density function

$$p_t(x) = \tfrac{1}{\sqrt{2\pi\sigma_t^2}}\exp\big(-\tfrac{x^2}{2\sigma_t^2}\big), \qquad \sigma_t^2 \coloneqq 1 + \mathrm{e}^{-2t}(\sigma_0^2 - 1). \tag{3.10}$$

Recalling (2.3) – (2.5) and using (3.10), we see that in this setting the cumulative Fisher information process of (3.3) is explicitly given by

$$\overline{F}_s^{\mathrm{OU}} = \int_0^s \Big(\nabla \log \overline{\ell}_u^{\downarrow}(\overline{X}_u)\Big)^2\,\mathrm{d}u = \int_0^s \Big(\nabla \log \overline{p}_u(\overline{X}_u) + \overline{X}_u\Big)^2\,\mathrm{d}u = \int_0^s \big(1 - \tfrac{1}{\overline{\sigma}_u^2}\big)^2\,\overline{X}_u^2\,\mathrm{d}u \tag{3.11}$$

for $0 \leqslant s \leqslant T$. In particular, the nonnegativity of the integrand in (3.11) 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 3.1 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. ◊

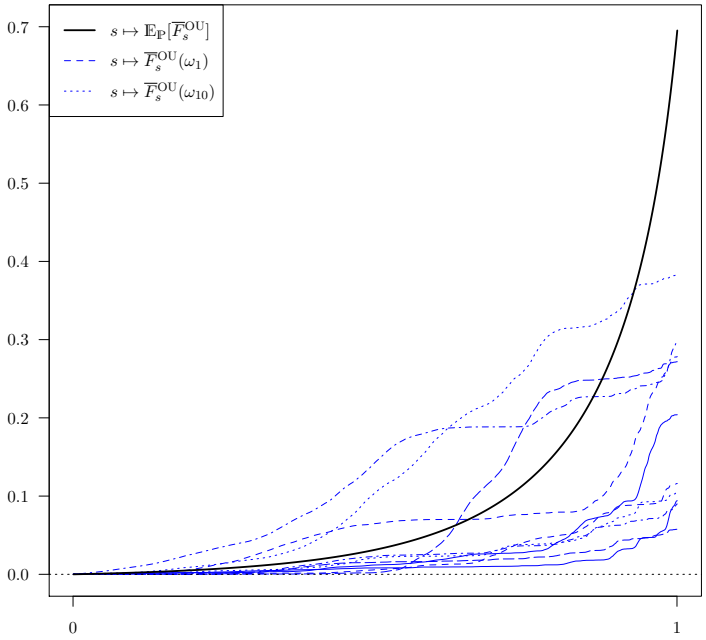


Figure 3.1.: Simulations of the cumulative Fisher information process (3.11) for the Ornstein–Uhlenbeck diffusion (3.8).

Example 3.3. 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 one-dimensional Brownian motion $(B_t)_{t \geqslant 0}$. In this case, for any $t \geqslant 0$, the drift term of the stochastic differential equation (1.1) is equal to

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

In particular, the drift term depends on the distribution P_t only through its mean. Substituting it into (1.1), this stochastic differential equation reduces to

$$\mathrm{d}X_t = -\big(X_t - \mathbb{E}[X_t]\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T. \tag{3.12}$$

This type of nonlinear, self-interacting stochastic differential equation has been studied since [BRTV98], where it was shown that its solution is also given by the Ornstein–Uhlenbeck process of (3.9). Therefore, similar computations as in Example 3.2 show that in this setting the cumulative Fisher information process is given by

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\overline{X}_u^2 + \mathbb{E}_{\tilde{\mathbb{P}}}\Big[\tfrac{1}{2}\big(\overline{X}_u - \overline{Y}_u\big)\big(\tfrac{2}{\overline{\sigma}_u^2}\overline{Y}_u - \overline{Y}_u\big)\Big]\Big)\,\mathrm{d}u \tag{3.13}$$

for $0 \leqslant s \leqslant T$. 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 (3.10), we can compute the expectation appearing in (3.13) and obtain

$$\overline{F}_s^{\mathrm{NL}} = \int_0^s \Big(\big(\tfrac{1}{\overline{\sigma}_u^4} + \tfrac{1}{2} - \tfrac{1}{\overline{\sigma}_u^2}\big)\overline{X}_u^2 + \tfrac{\overline{\sigma}_u^2}{2} - 1 \Big)\, \mathrm{d}u. \tag{3.14}$$

Clearly, $\overline{F}_s^{\mathrm{OU}} \neq \overline{F}_s^{\mathrm{NL}}$. In particular, the integrand in (3.14) is nonnegative if and only if $\overline{X}_u^2 \geqslant \big(\frac{1}{\overline{\sigma}_u^4} + \frac{1}{2} - \frac{1}{\overline{\sigma}_u^2}\big)^{-1}\big(1 - \frac{\overline{\sigma}_u^2}{2}\big)$. In other words, as opposed to Example 3.2, relative entropy only decreases along a trajectory if $\overline{X}_u$ is far from its mean. However, after taking expectations in (3.11) and (3.14), we see that the expected rate of relative entropy dissipation in both cases is equal to

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{OU}}\big] = \mathbb{E}_{\mathbb{P}}\big[\overline{F}_s^{\mathrm{NL}}\big] = \int_0^s \big(\overline{\sigma}_u - \tfrac{1}{\overline{\sigma}_u}\big)^2\, \mathrm{d}u, \qquad 0 \leqslant s \leqslant T. \tag{3.15}$$

Now we set again $T = 1$ and $\sigma_0^2 = 0.1$. In the same vein as in Figure 3.1, we plot in Figure 3.2 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 (3.14), and hence its integrand, can both be negative. Finally, the thick black line in Figure 3.2 follows the expected path $s \mapsto \mathbb{E}_{\mathbb{P}}[\overline{F}_s^{\mathrm{NL}}]$ of all possible trajectories. According to (3.15), this is the same black line as in Figure 3.1. ◊

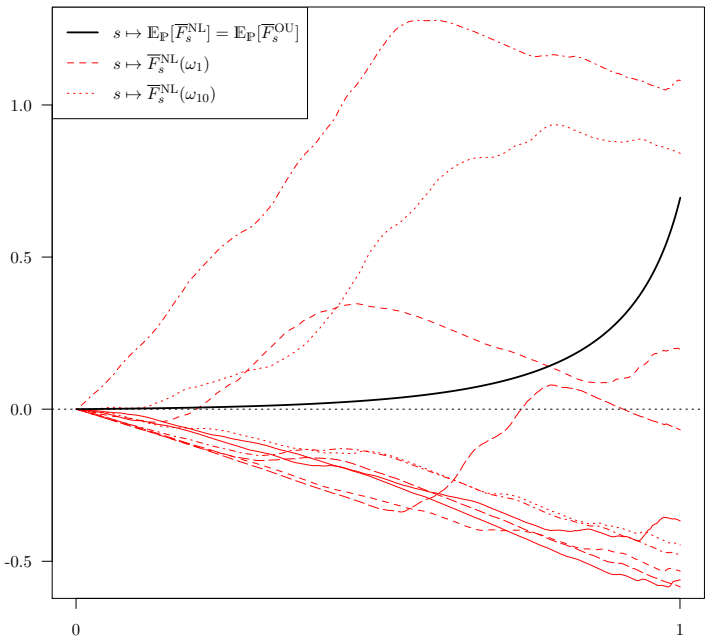


Figure 3.2.: Simulations of the cumulative Fisher information process (3.14) for the nonlinear, self-interacting diffusion (3.12).

3.1.2. Consequences of Theorem 3.1

We now return to the general statement of Theorem 3.1 and deduce some direct consequences. By averaging the trajectorial result of Theorem 3.1 according to the path measure $\mathbb{P}$, we derive the well-known *relative entropy identity* (3.16) and the *dissipation of relative entropy* (3.17) below. A sketch of proof for the latter result was first given in [CMV03, Proposition 2.1].

Corollary 3.4. *Suppose Assumptions 2.1 hold. For all $t, t_0 \in [0,T]$, we have the relative entropy identity*

$$H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0}) = -\int_{t_0}^{t} I(\mathrm{P}_u \,|\, \mathrm{Q}_u^{\uparrow})\,\mathrm{d}u. \tag{3.16}$$

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

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, H(\mathrm{P}_t \,|\, \mathrm{Q}_t) = -I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow}). \tag{3.17}$$

The relation (3.17) 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 $I(\mathrm{P}_t \,|\, \mathrm{Q}_t^{\uparrow})$.

Proof of Corollary 3.4. The identity (3.16) follows by taking expectations with respect to the probability measure $\mathbb{P}$ in (3.5), recalling the definitions of (2.7), using (3.7), and invoking the fact that the $\mathbb{P}$-expectation of the martingale (3.6) vanishes. Finally, applying the Lebesgue differentiation theorem to the monotone function $t \mapsto H(\mathrm{P}_t \,|\, \mathrm{Q}_t)$ gives (3.17). □

Finally, let us place ourselves again on the filtered probability space $(\Omega, \mathbb{G}, \mathbb{P})$ as in Theorem 3.1 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 (3.17) and deduce the following trajectorial rate of relative entropy dissipation.

Corollary 3.5. *Suppose Assumptions 2.1 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*

$$\lim_{s \downarrow s_0} \frac{\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big)}{s - s_0} = \overline{I}_{s_0}(\omega). \tag{3.18}$$

Recalling (3.7), we observe that $\mathbb{E}_{\mathbb{P}}[\overline{I}_{s_0}] = I(\overline{\mathrm{P}}_{s_0} \,|\, \overline{\mathrm{Q}}_{s_0}^{\uparrow})$. Therefore the limiting assertion (3.18) can indeed be viewed as a trajectorial version of the deterministic relative entropy dissipation identity (3.17).

Proof of Corollary 3.5. We let $0 \leqslant s_0 \leqslant s \leqslant T$. By (3.5), (3.3) and Fubini's theorem, we have for $\mathbb{P}$-a.e. $\omega \in \Omega$

$$\mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) \,|\, \mathcal{G}_{s_0}\big](\omega) - \log \overline{\ell}_{s_0}\big(\overline{X}_{s_0}(\omega), \overline{\mathrm{P}}_{s_0}\big) = \int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega)\,\mathrm{d}u. \tag{3.19}$$

Furthermore, Jensen's inequality gives

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

which implies

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

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

$$\lim_{s \downarrow s_0} \frac{\int_{s_0}^{s} \mathbb{E}_{\mathbb{P}}\big[\overline{I}_u \,|\, \mathcal{G}_{s_0}\big](\omega) \,\mathrm{d}u}{s - s_0} = \mathbb{E}_{\mathbb{P}}\big[\overline{I}_{s_0} \,|\, \mathcal{G}_{s_0}\big](\omega) = \overline{I}_{s_0}(\omega) \tag{3.20}$$

holds for every $s_0 \in [0,T] \setminus N_\omega$. Finally, combining (3.19) and (3.20) proves (3.18). □

3.2. Gradient flow structure of the granular media equation

In this section we apply the trajectorial approach of Section 3.1 in order to formulate the gradient flow property of the granular media equation (1.12). To this end, we consider a function $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$, which will be treated as a perturbation potential. We denote by $V^{\boldsymbol{\beta}} \coloneqq V + \boldsymbol{\beta}$ the perturbed confinement potential and invoke the following regularity assumptions.

Assumptions 3.6. The function $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ is smooth and compactly supported, and we require that Assumptions 2.1 are still satisfied if we replace V by $V^{\boldsymbol{\beta}}$.

Note that Assumptions 3.6 allow us to apply Lemma 2.2 to the "perturbed" McKean–Vlasov stochastic differential equation

$$\mathrm{d}X_t = -\big(\nabla V^{\boldsymbol{\beta}}(X_t) + \nabla(W * \mathrm{P}_t^{\boldsymbol{\beta}})(X_t)\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^{\boldsymbol{\beta}}, \qquad t_0 \leqslant t \leqslant T \tag{3.21}$$

starting at time $t_0 \geqslant 0$, with X_{t_0} having initial distribution $\mathrm{P}_{t_0}^{\boldsymbol{\beta}} = \mathrm{P}_{t_0}$. Therefore, by analogy with Section 2.1, we can construct a probability measure $\mathbb{P}^{\boldsymbol{\beta}}$ on $\Omega \coloneqq C([t_0,T];\mathbb{R}^n)$, under which the canonical process $(X_t)_{t_0 \leqslant t \leqslant T}$ satisfies the stochastic differential equation (3.21), with $(B_t^{\boldsymbol{\beta}})_{t_0 \leqslant t \leqslant T}$ being a $\mathbb{P}^{\boldsymbol{\beta}}$-Brownian motion.

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

$$\partial_t p_t^{\boldsymbol{\beta}}(x) = \operatorname{div}\Big(\nabla p_t^{\boldsymbol{\beta}}(x) + p_t^{\boldsymbol{\beta}}(x)\nabla V^{\boldsymbol{\beta}}(x) + p_t^{\boldsymbol{\beta}}(x)\nabla(W * p_t^{\boldsymbol{\beta}})(x)\Big), \tag{3.22}$$

for $(t,x) \in (t_0,T) \times \mathbb{R}^n$, with initial condition $p_{t_0}^{\boldsymbol{\beta}}(\,\cdot\,) = p_{t_0}(\,\cdot\,)$.

By analogy with (2.3), we define the perturbed potentials

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

for $(x,\mu) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$. In parallel to (2.5), we introduce the *perturbed likelihood ratio functions*

$$\ell_t^{\beta}(x,\mu) := \frac{p_t^{\beta}(x)}{q(x,\mu)}\,, \qquad \ell_t^{\beta\uparrow}(x,\mu) := \frac{p_t^{\beta}(x)}{q^{\uparrow}(x,\mu)}\,, \qquad \ell_t^{\beta\downarrow}(x) := \frac{p_t^{\beta}(x)}{q^{\downarrow}(x)} \tag{3.23}$$

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

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

They are the perturbed versions of the measures Q_t and $\mathrm{Q}_t^{\uparrow}$ defined in (2.6). The relative entropy of P_t^{β} with respect to Q_t^{β} is then given by

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

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

$$I(\mathrm{P}_t^{\beta} \,|\, \mathrm{Q}_t^{\beta\uparrow}) = \mathbb{E}_{\mathbb{P}^{\beta}}\Big[\big|\nabla \log \ell_t^{\beta\uparrow}(X_t, \mathrm{P}_t^{\beta})\big|^2\Big] = \mathscr{D}(\mathrm{P}_t^{\beta}). \tag{3.24}$$

The following trajectorial result, Theorem 3.7 below, provides the semimartingale decomposition of the *perturbed relative entropy process*

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

In line with its unperturbed counterpart, Theorem 3.1, we shall formulate this result in the reverse direction of time. We first introduce the perturbed analogues of (3.2) and (3.3): the *time-reversed perturbed Fisher information process* is defined as

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

for all $0 \leqslant s \leqslant T - t_0$, where $(\overline{Y}_s)_{0\leqslant s\leqslant T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\leqslant s\leqslant T-t_0}$ 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

$$\overline{F}_s^{\beta} := \int_0^s \overline{I}_u^{\beta}\,\mathrm{d}u, \qquad 0 \leqslant s \leqslant T - t_0. \tag{3.25}$$

Theorem 3.7. *Suppose Assumptions 3.6 hold. On* $(\Omega, \mathbb{G}, \mathbb{P}^{\beta})$*, the time-reversed perturbed relative entropy process*

$$\log \overline{\ell}_s^{\beta}(\overline{X}_s, \overline{\mathrm{P}}_s^{\beta}) = \log \ell_{T-s}^{\beta}(X_{T-s}, \mathrm{P}_{T-s}^{\beta}), \qquad 0 \leqslant s \leqslant T - t_0 \tag{3.26}$$

admits the semimartingale decomposition

$$\log \overline{\ell}_s^{\beta}(\overline{X}_s, \overline{\mathrm{P}}_s^{\beta}) - \log \overline{\ell}_0^{\beta}(\overline{X}_0, \overline{\mathrm{P}}_0^{\beta}) = \overline{M}_s^{\beta} + \overline{F}_s^{\beta}. \tag{3.27}$$

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

$$\overline{M}_s^{\beta} \coloneqq \int_0^s \Big\langle \nabla \log \overline{\ell}_u^{\beta}(\overline{X}_u, \overline{\mathrm{P}}_u^{\beta}) \,,\, \sqrt{2}\, \mathrm{d}\overline{B}_u^{\beta} \Big\rangle, \tag{3.28}$$

with $(\overline{B}_s^{\beta})_{0 \leqslant s \leqslant T-t_0}$ *a* $\mathbb{P}^{\beta}$*-Brownian motion of the backward filtration* $\mathbb{G}$*, and where the compensator (3.25) satisfies*

$$\mathbb{E}_{\mathbb{P}^{\beta}}[\overline{F}_s^{\beta}] = \int_0^s \Big(I(\overline{\mathrm{P}}_u^{\beta} \,|\, \overline{\mathrm{Q}}_u^{\beta\uparrow}) + \mathbb{E}_{\mathbb{P}^{\beta}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_u^{\beta}) \,, \nabla \boldsymbol{\beta} \Big\rangle - \Delta \boldsymbol{\beta} \Big)(\overline{X}_u) \Big] \Big) \, \mathrm{d}u < \infty. \tag{3.29}$$

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 (3.30), and applying the Lebesgue differentiation theorem gives the perturbed rate of relative entropy dissipation (3.31).

Corollary 3.8. *Suppose Assumptions 3.6 hold. For all* $0 \leqslant t_0 \leqslant t \leqslant T$*, we have the perturbed relative entropy identity*

$$\begin{aligned} &H(\mathrm{P}_t^{\beta} \,|\, \mathrm{Q}_t^{\beta}) - H(\mathrm{P}_{t_0}^{\beta} \,|\, \mathrm{Q}_{t_0}^{\beta}) = \\ &\quad = -\int_{t_0}^t \Big(I(\mathrm{P}_u^{\beta} \,|\, \mathrm{Q}_u^{\beta\uparrow}) + \mathbb{E}_{\mathbb{P}^{\beta}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_u^{\beta}) \,, \nabla \boldsymbol{\beta} \Big\rangle - \Delta \boldsymbol{\beta} \Big)(X_u) \Big] \Big) \, \mathrm{d}u, \end{aligned} \tag{3.30}$$

and, for Lebesgue-a.e. $t_0 \in [0, T]$*, the perturbed rate of relative entropy dissipation*

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_0}^{+} H(\mathrm{P}_t^{\beta} \,|\, \mathrm{Q}_t^{\beta}) = \\ &\quad = -\Big(I(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0}) + \mathbb{E}_{\mathbb{P}}\Big[\Big(\Big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0}) \,, \nabla \boldsymbol{\beta} \Big\rangle - \Delta \boldsymbol{\beta} \Big)(X_{t_0}) \Big] \Big). \end{aligned} \tag{3.31}$$

Proof. The proof follows by analogy with Corollary 3.4. □

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

Corollary 3.9. *Suppose* Assumptions 3.6 *hold and*

$$\int_0^{T-t_0} \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[|\overline{I}_u^{\boldsymbol{\beta}}|\big]\,\mathrm{d}u < \infty.$$

For $\mathbb{P}^{\boldsymbol{\beta}}$-a.e. $\omega \in \Omega$ there exists a Lebesgue null set $N_\omega^{\boldsymbol{\beta}} \subseteq [0, T-t_0]$ such that for any $s_0 \in [0, T-t_0] \setminus N_\omega^{\boldsymbol{\beta}}$ we have

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

Proof. The proof proceeds almost verbatim as the proof of Corollary 3.5. The only difference is that we now use the semimartingale decomposition (3.27) and the $\mathbb{P}^{\boldsymbol{\beta}}$-martingale property of the process (3.28) in Theorem 3.7. □

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

$$v_t^{\boldsymbol{\beta}}(x) \coloneqq -(\nabla \log p_t^{\boldsymbol{\beta}} + \nabla V^{\boldsymbol{\beta}} + \nabla(W * p_t^{\boldsymbol{\beta}}))(x), \qquad (t,x) \in [t_0, T] \times \mathbb{R}^n, \tag{3.32}$$

so that the perturbed granular media equation (3.22) can be viewed as a *continuity equation*

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

with $v_t^{\boldsymbol{\beta}}(\,\cdot\,)$ as the corresponding *velocity field*. We recall the definition of the tangent space (see Definition 8.4.1 in [AGS08])

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

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

Assumptions 3.10. In addition to Assumptions 3.6, we suppose that

$$v_t(\,\cdot\,) \in \operatorname{Tan}_{\mathrm{P}_t} \mathscr{P}_2(\mathbb{R}^n) \qquad \text{for Lebesgue-a.e. } t \in [0,T], \tag{3.33}$$

where $v_t(\,\cdot\,)$ is obtained by taking $\boldsymbol{\beta} \equiv 0$ and $t_0 = 0$ in (3.32).

For example, we know from [AGS08, Theorem 10.4.13] that the condition (3.33) is satisfied if, in addition to Assumptions 3.6, V is uniformly convex, i.e., $\operatorname{Hess}(V) \geqslant \kappa_V I_n$ for some real constant κ_V, and W is a convex function satisfying the *doubling condition*

$$\exists\, C_W > 0 \text{ such that } \forall\, x, y \in \mathbb{R}^n \colon \qquad W(x+y) \leqslant C_W(1 + W(x) + W(y)).$$

The proof of the following result is based on the general theory of Wasserstein metric derivatives of absolutely continuous curves in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$; for a thorough discussion, we refer to Chapter 8 in [AGS08].

Lemma 3.11. *Suppose Assumptions 3.10 hold. For Lebesgue-a.e. $t_0 \in [0,T]$, the Wasserstein metric derivative of the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \leqslant t \leqslant T}$ is equal to*

$$\lim_{t \downarrow t_0} \frac{W_2(\mathrm{P}_t^{\boldsymbol{\beta}}, \mathrm{P}_{t_0}^{\boldsymbol{\beta}})}{t - t_0} = \|v_{t_0}^{\boldsymbol{\beta}}(X_{t_0})\|_{L^2(\mathbb{P})} = \|\nabla \log \ell_{t_0}^{\uparrow}(X_{t_0}, \mathrm{P}_{t_0}) + \nabla \boldsymbol{\beta}(X_{t_0})\|_{L^2(\mathbb{P})}. \tag{3.34}$$

Proof. Without loss of generality we can set $\boldsymbol{\beta} \equiv 0$. Note that from (3.7) we have

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

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 [AGS08] to the absolutely continuous curve $(\mathrm{P}_t)_{0 \leqslant t \leqslant T}$, which yields (3.34). □

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 $(\mathrm{P}_t)_{t_0 \leqslant t \leqslant T}$ is defined as

$$|\partial \mathscr{F}|_{W_2}(\mathrm{P}_{t_0}) := \lim_{t \downarrow t_0} \frac{H(\mathrm{P}_t \,|\, \mathrm{Q}_t) - H(\mathrm{P}_{t_0} \,|\, \mathrm{Q}_{t_0})}{W_2(\mathrm{P}_t, \mathrm{P}_{t_0})}.$$

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

$$|\partial \mathscr{F}^{\boldsymbol{\beta}}|_{W_2}(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}) := \lim_{t \downarrow t_0} \frac{H(\mathrm{P}_t^{\boldsymbol{\beta}} \,|\, \mathrm{Q}_t^{\boldsymbol{\beta}}) - H(\mathrm{P}_{t_0}^{\boldsymbol{\beta}} \,|\, \mathrm{Q}_{t_0}^{\boldsymbol{\beta}})}{W_2(\mathrm{P}_t^{\boldsymbol{\beta}}, \mathrm{P}_{t_0}^{\boldsymbol{\beta}})}$$

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

Theorem 3.12. *Suppose Assumptions 3.10 hold. The following assertions hold for Lebesgue-a.e. $t_0 \in [0,T]$: The random variables*

$$\mathcal{L}_{t_0}^{\uparrow} := \nabla \log \ell_{t_0}^{\uparrow}(X_{t_0}, \mathrm{P}_{t_0}) \qquad \text{and} \qquad \mathcal{B}_{t_0} := \nabla \boldsymbol{\beta}(X_{t_0}) \tag{3.35}$$

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

$$|\partial \mathscr{F}|_{W_2}(\mathrm{P}_{t_0}) = -\|\mathcal{L}_{t_0}^{\uparrow}\|_{L^2(\mathbb{P})}. \tag{3.36}$$

If $\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0} \neq 0$, the metric slope along the perturbed curve $(\mathrm{P}_t^{\boldsymbol{\beta}})_{t_0 \leqslant t \leqslant T}$ is equal to

$$|\partial \mathscr{F}^{\boldsymbol{\beta}}|_{W_2}(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}) = -\left\langle \mathcal{L}_{t_0}^{\uparrow}, \frac{\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}}{\|\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \tag{3.37}$$

In particular,

$$|\partial \mathscr{F}|_{W_2}(\mathrm{P}_{t_0}) \leqslant |\partial \mathscr{F}^{\boldsymbol{\beta}}|_{W_2}(\mathrm{P}_{t_0}^{\boldsymbol{\beta}}) \tag{3.38}$$

with equality if and only if $\mathcal{L}_{t_0}^{\uparrow} + \mathcal{B}_{t_0}$ is a positive multiple of $\mathcal{L}_{t_0}^{\uparrow}$.

Proof. The equality (3.36) follows from (3.17) and by taking $\boldsymbol{\beta} \equiv 0$ in (3.34). For the proof of (3.37), we first observe that from (3.31) and (3.34) we obtain the equality

$$|\partial \mathscr{F}^{\boldsymbol{\beta}}|_{W_2}(\mathrm{P}^{\boldsymbol{\beta}}_{t_0}) = -\frac{\|\mathcal{L}^{\uparrow}_{t_0}\|^2_{L^2(\mathbb{P})} + \mathbb{E}_{\mathbb{P}}\Big[\Big(\big\langle \nabla V + \nabla(W * \mathrm{P}_{t_0})\,, \nabla\boldsymbol{\beta}\big\rangle - \Delta\boldsymbol{\beta}\Big)(X_{t_0})\Big]}{\|\mathcal{L}^{\uparrow}_{t_0} + \mathcal{B}_{t_0}\|_{L^2(\mathbb{P})}}, \tag{3.39}$$

for Lebesgue-a.e. $t_0 \in [0,T]$. Integrating by parts and recalling the notations in (2.3) – (2.5), we find that the expectation in the numerator of (3.39) is equal to

$$\int_{\mathbb{R}^n} \Big\langle \log \nabla \ell^{\uparrow}_{t_0}(x)\,, \nabla\boldsymbol{\beta}(x)\Big\rangle\, p_{t_0}(x)\,\mathrm{d}x = \Big\langle \mathcal{L}^{\uparrow}_{t_0}\,,\, \mathcal{B}_{t_0}\Big\rangle_{L^2(\mathbb{P})}.$$

Now (3.38) follows by the Cauchy–Schwarz inequality. □

3.3. 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 this work. In particular, we shall impose convexity assumptions on the potentials V, W.

We fix two probability measures ν_0 and ν_1 in $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$. By Brenier's theorem [Bre91], there exists a convex function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$W_2^2(\nu_0,\nu_1) = \int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2\,\mathrm{d}\nu_0(x). \tag{3.40}$$

The displacement interpolation of McCann [McC97] between ν_0 and ν_1 is given by

$$\nu_t \coloneqq (T_t)_{\#}\nu_0, \qquad T_t(x) \coloneqq (1-t)x + t\nabla\varphi(x), \qquad 0 \leqslant t \leqslant 1. \tag{3.41}$$

In particular, since the endpoints ν_0 and ν_1 belong to $\mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$, each ν_t has a probability density function ρ_t; see, e.g., [Vil03, 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 (2.6), the σ-finite measures

$$\mu_t(A) \coloneqq \int_A q(x,\nu_t)\,\mathrm{d}x, \qquad \mu^{\uparrow}_t(A) \coloneqq \int_A q^{\uparrow}(x,\nu_t)\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n),$$

where we recall the definitions of the density functions q and $q^{\uparrow}$ in (2.4). In parallel to the likelihood ratio functions in (2.5), we define

$$r_t(x,\nu) \coloneqq \frac{\rho_t(x)}{q(x,\nu)}, \qquad r^{\uparrow}_t(x,\nu) \coloneqq \frac{\rho_t(x)}{q^{\uparrow}(x,\nu)} \tag{3.42}$$

for $(t,x,\nu) \in [0,1] \times \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$. Then the relative entropy of ν_t with respect to μ_t is given by

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

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

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

We impose the following regularity conditions for Proposition 3.14, noting that the strong assumptions placed on ρ_0 and ρ_1 are only temporary and will be removed in Assumptions 3.15 of Theorem 3.16.

Assumptions 3.13. The functions $V, W : \mathbb{R}^n \to [0, \infty)$ 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 3.14. *Suppose Assumptions 3.13 hold. Along the displacement interpolation $(\nu_t)_{0 \leqslant t \leqslant 1}$, the rate of relative entropy dissipation at time $t = 0$, with respect to the "reference curve of probability measures" $(\mu_t)_{0 \leqslant t \leqslant 1}$, is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}^{+} H(\nu_t \,|\, \mu_t) = \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \,, \nabla \varphi(x) - x \Big\rangle \, \rho_0(x) \, \mathrm{d}x. \tag{3.43}$$

Combining (3.43) of Proposition 3.14 with the displacement convexity results of McCann [McC97], we obtain the following generalization of the HWBI inequality. Equivalent versions of this inequality can be found in [CEGH04, Theorem 4.1] and [FK06, Theorem D.50].

Assumptions 3.15. The functions $V, W : \mathbb{R}^n \to [0, \infty)$ 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

$$\mathrm{Hess}(V) \geqslant \kappa_V I_n, \qquad \mathrm{Hess}(W) \geqslant \kappa_W I_n. \tag{3.44}$$

Theorem 3.16. *Suppose Assumptions 3.15 hold and the relative entropy $H(\nu_1 \,|\, \mu_1)$ is finite. Then*

$$H(\nu_0 \,|\, \mu_0) - H(\nu_1 \,|\, \mu_1) \leqslant - \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x, \nu_0) \,, \nabla \varphi(x) - x \Big\rangle \, \rho_0(x) \, \mathrm{d}x \tag{3.45}$$

$$- \tfrac{\kappa_V + \kappa_W}{2} \, W_2^2(\nu_0, \nu_1) + \tfrac{\kappa_W}{2} \, |b(\nu_0) - b(\nu_1)|^2. \tag{3.46}$$

By the Cauchy–Schwarz inequality, the right-hand side of (3.45) can be bounded from above by

$$\sqrt{\int_{\mathbb{R}^n} |\nabla \log r_0^{\uparrow}(x, \nu_0)|^2 \, \rho_0(x) \, \mathrm{d}x} \, \sqrt{\int_{\mathbb{R}^n} |\nabla \varphi(x) - x|^2 \, \rho_0(x) \, \mathrm{d}x} = \sqrt{I(\nu_0 \,|\, \mu_0^{\uparrow})} \, W_2(\nu_0, \nu_1),$$

and we obtain the usual form of the HWBI inequality (1.16); see also Theorem 4.2 in [AGK04].

4. Proofs of the main results

This chapter is devoted to the proofs of the results stated in Chapter 3. We shall first prove the main trajectorial results: Theorem 3.1 and its "perturbed" counterpart, Theorem 3.7.

4.1. The proofs of Theorem 3.1 and Theorem 3.7

Since Theorem 3.1 follows immediately from Theorem 3.7 by setting the perturbation $\boldsymbol{\beta}\colon \mathbb{R}^n \to \mathbb{R}$ to be the zero function, we start with the general setting of Theorem 3.7. We first recall a classical result concerning the time reversal of diffusions. For proof and references we refer to Theorems 4.1 and 5.1 of Part V.

Lemma 4.1. *Suppose Assumptions 3.6 hold. On* $(\Omega, \mathbb{G}, \mathbb{P}^{\boldsymbol{\beta}})$*, the process*

$$\overline{B}_s^{\boldsymbol{\beta}} := B_{T-s}^{\boldsymbol{\beta}} - B_T^{\boldsymbol{\beta}} - \sqrt{2}\int_0^s \nabla \log \overline{p}_u^{\boldsymbol{\beta}}(\overline{X}_u^{\boldsymbol{\beta}})\,\mathrm{d}u, \qquad 0 \leqslant s \leqslant T - t_0$$

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

$$\mathrm{d}\overline{X}_s = \Big(2\nabla \log \overline{\ell}_s^{\boldsymbol{\beta}\downarrow} - \nabla V + \nabla(W * \overline{\mathrm{P}}_s^{\boldsymbol{\beta}}) + \nabla\boldsymbol{\beta}\Big)(\overline{X}_s)\,\mathrm{d}s + \sqrt{2}\,\mathrm{d}\overline{B}_s^{\boldsymbol{\beta}}. \tag{4.1}$$

By means of Lemma 4.1, the first step in the proof of Theorem 3.7 is to compute the dynamics of the time-reversed perturbed relative entropy process (3.26). For the reader's convenience, we recall the following characterization of the L-derivative in [CD18, pp. 383].

Definition 4.2. Let $f\colon \mathscr{P}_2(\mathbb{R}^n) \to \mathbb{R}$ and $\mu_0 \in \mathscr{P}_2(\mathbb{R}^n)$. On a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, let X_0 be a random variable with distribution μ_0. We define

$$\partial_\mu f(\mu_0)\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

as the L-*derivative* of f at μ_0, if for any $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ and any random variable X with distribution μ, it holds that

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

Remark 4.3. 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 [CD18] for the details. ◇

Proposition 4.4. *Suppose Assumptions 3.6 hold. On $(\Omega, \mathbb{G}, \mathbb{P}^\beta)$, the time-reversed perturbed relative entropy process (3.26) satisfies*

$$\mathrm{d}\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s) = \Big\langle \nabla\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)\,,\,\sqrt{2}\,\mathrm{d}\overline{B}^\beta_s\Big\rangle \tag{4.2}$$

$$+\Big(|\nabla\log\overline{\ell}^{\beta\downarrow}_s|^2 + \tfrac{1}{2}|\nabla(W*\overline{\mathrm{P}}^\beta_s)|^2\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.3}$$

$$+\Big(\Big\langle \tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}^\beta_s)\,,\,2\nabla\log\overline{\ell}^{\beta\downarrow}_s + \nabla V^\beta\Big\rangle + \langle\nabla V\,,\nabla\boldsymbol{\beta}\rangle - \Delta\boldsymbol{\beta}\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.4}$$

$$-\mathbb{E}_{\tilde{\mathbb{P}}^\beta}\Big[\Big\langle \tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s)\,,\,\Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}^\beta_s) + \nabla\boldsymbol{\beta}\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s, \tag{4.5}$$

where $(\overline{Y}_s)_{0\leqslant s\leqslant T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\leqslant s\leqslant T-t_0}$ 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 [CD18, Proposition 5.102] and using the backward dynamics in (4.1), we obtain

$$\begin{aligned}\mathrm{d}\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s) &= \Big\langle \nabla\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)\,,\,\sqrt{2}\,\mathrm{d}\overline{B}^\beta_s\Big\rangle + \Big(\partial_s\log\overline{\ell}^\beta_s + \Delta\log\overline{\ell}^\beta_s\Big)(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)\,\mathrm{d}s\\
&+\Big\langle \nabla\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)\,,\,\Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}^\beta_s) + \nabla\boldsymbol{\beta}\Big)(\overline{X}_s)\Big\rangle\,\mathrm{d}s\\
&+\mathbb{E}_{\tilde{\mathbb{P}}^\beta}\Big[\Big\langle\Big(\partial_\mu\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)\Big)\,,\,\Big(2\nabla\log\overline{\ell}^{\beta\downarrow}_s - \nabla V + \nabla(W*\overline{\mathrm{P}}^\beta_s) + \nabla\boldsymbol{\beta}\Big)\Big\rangle(\overline{Y}_s)\Big]\,\mathrm{d}s \qquad (4.6)\\
&+\mathbb{E}_{\tilde{\mathbb{P}}^\beta}\Big[\mathrm{trace}\Big(\partial_y\partial_\mu\log\overline{\ell}^\beta_s(\overline{X}_s,\overline{\mathrm{P}}^\beta_s)(\overline{Y}_s)\Big)\Big]\,\mathrm{d}s, \qquad (4.7)\end{aligned}$$

where $(\overline{Y}_s)_{0\leqslant s\leqslant T-t_0}$ is a copy of the process $(\overline{X}_s)_{0\leqslant s\leqslant T-t_0}$ 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 (4.6) and (4.7) is calculated to be

$$\Big(\partial_\mu\log\overline{\ell}^\beta_s(x,\mu)\Big)(y) = \tfrac{1}{2}\big(\partial_\mu\,(W*\mu)\,(x)\big)(y) = -\tfrac{1}{2}\nabla W(x-y) \tag{4.8}$$

for $(x,\mu,y)\in\mathbb{R}^n\times\mathscr{P}_2(\mathbb{R}^n)\times\mathbb{R}^n$, see [CD18, 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

$$\mathrm{trace}\Big(\partial_y\partial_\mu\log\overline{\ell}^\beta_s(x,\mu)(y)\Big) = -\tfrac{1}{2}\mathrm{trace}\big(\partial_y\nabla W(x-y)\big) = \tfrac{1}{2}\Delta W(x-y). \tag{4.9}$$

Putting (4.8) and (4.9) into (4.6) and (4.7), respectively, as well as using the identities

$$\begin{aligned}\partial_s\log\overline{\ell}^\beta_s(x,\mu) &= \partial_s\log\overline{\ell}^{\beta\downarrow}_s(x),\\
\nabla\log\overline{\ell}^\beta_s(x,\mu) &= \nabla\log\overline{\ell}^{\beta\downarrow}_s(x) + \tfrac{1}{2}\nabla(W*\mu)(x),\\
\Delta\log\overline{\ell}^\beta_s(x,\mu) &= \Delta\log\overline{\ell}^{\beta\downarrow}_s(x) + \tfrac{1}{2}\Delta(W*\mu)(x),\end{aligned}$$

we obtain

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

$$+\Big(\partial_s\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}+\Delta\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.10}$$

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

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

$$+\tfrac{1}{2}\Big(\Delta(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})(\overline{X}_s)+\mathbb{E}_{\tilde{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Delta W(\overline{X}_s-\overline{Y}_s)\Big]\Big)\,\mathrm{d}s. \tag{4.11}$$

Regarding the expression of (4.11), we observe that

$$\Delta(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})(\overline{X}_s)=\mathbb{E}_{\tilde{\mathbb{P}}^{\boldsymbol{\beta}}}\Big[\Delta W(\overline{X}_s-\overline{Y}_s)\Big].$$

Finally, elementary computations based on (2.4), (3.22) and (3.23) show that the perturbed log-likelihood ratio function $(s,x)\mapsto\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}(x)$ of (3.23) satisfies

$$\begin{aligned}\partial_s\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}=\Big\langle\nabla\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}\,,\nabla V-\nabla(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})-\nabla\boldsymbol{\beta}\Big\rangle-\big|\nabla\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}\big|^2-\Delta\log\overline{\ell}_s^{\boldsymbol{\beta}\downarrow}\\+\Big\langle\nabla V\,,\nabla(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})+\nabla\boldsymbol{\beta}\Big\rangle-\Delta(W*\overline{\mathrm{P}}_s^{\boldsymbol{\beta}})-\Delta\boldsymbol{\beta}\end{aligned} \tag{4.12}$$

on $(0,T-t_0)\times\mathbb{R}^n$, with terminal condition $\log\overline{\ell}_{T-t_0}^{\boldsymbol{\beta}\downarrow}=\log\overline{\ell}_{T-t_0}^{\downarrow}$. Finally, inserting (4.12) into (4.10), we obtain (4.2) – (4.5). □

Setting the perturbation $\boldsymbol{\beta}$ to be the zero function, we obtain the following result.

Corollary 4.5. *Suppose Assumptions 2.1 hold. On $(\Omega,\mathbb{G},\mathbb{P})$, the time-reversed relative entropy process (3.4) satisfies*

$$\mathrm{d}\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s)=\Big\langle\nabla\log\overline{\ell}_s(\overline{X}_s,\overline{\mathrm{P}}_s)\,,\,\sqrt{2}\,\mathrm{d}\overline{B}_s\Big\rangle \tag{4.13}$$

$$+\Big(\big|\nabla\log\overline{\ell}_s^{\downarrow}\big|^2+\tfrac{1}{2}\big|\nabla(W*\overline{\mathrm{P}}_s)\big|^2+\Big\langle\tfrac{1}{2}\nabla(W*\overline{\mathrm{P}}_s)\,,\,2\nabla\log\overline{\ell}_s^{\downarrow}+\nabla V\Big\rangle\Big)(\overline{X}_s)\,\mathrm{d}s \tag{4.14}$$

$$-\mathbb{E}_{\tilde{\mathbb{P}}}\Big[\Big\langle\tfrac{1}{2}\nabla W(\overline{X}_s-\overline{Y}_s)\,,\,\Big(2\nabla\log\overline{\ell}_s^{\downarrow}-\nabla V+\nabla(W*\overline{\mathrm{P}}_s)\Big)(\overline{Y}_s)\Big\rangle\Big]\,\mathrm{d}s. \tag{4.15}$$

Here, the process

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

is a $\mathbb{P}$-Brownian motion with respect to the backward filtration $\mathbb{G}$, and $(\overline{Y}_s)_{0\leqslant s\leqslant T}$ is a copy of the process $(\overline{X}_s)_{0\leqslant s\leqslant T}$ 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 3.1, we state a classical result based on the general theory of the Cameron–Martin–Maruyama–Girsanov transformation [LS01]. The connection between relative entropy (the left-hand side of (4.16) below) and energy (the right-hand side of (4.16)) is the foundation of Föllmers entropy approach to the time reversal of diffusion processes on Wiener space [Föl85, Föl86, Föl88]. We denote by $\mathbb{W}_x$ the Wiener measure on $\Omega = C([0,T];\mathbb{R}^n)$ with starting point $x \in \mathbb{R}^n$, and define by

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

the Wiener measure with starting point x and variance 2.

Lemma 4.6. *The relative entropy of* $\mathbb{P}$ *with respect to the Wiener measure*

$$\mathbb{W}_{\mathrm{P}_0,2} \coloneqq \int_{\mathbb{R}^n} \mathbb{W}_{x,2}\,\mathrm{d}\mathrm{P}_0(x),$$

with initial distribution P_0 *and variance* 2*, is given by*

$$H(\mathbb{P} \,|\, \mathbb{W}_{\mathrm{P}_0,2}) = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \big|\nabla V(X_t) + \nabla(W * \mathrm{P}_t)(X_t)\big|^2\,\mathrm{d}t\bigg] < \infty. \tag{4.16}$$

Proof. Recalling (2.3), the drift of the McKean–Vlasov dynamics (1.1) can be expressed as

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

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

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

Using the linear growth condition (2.1) from Assumptions 2.1, (i), we find

$$\mathbb{E}_{\mathbb{P}}\Big[\big|\nabla V(X_t)\big|^2\Big] \leqslant 2C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \leqslant 2C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leqslant t\leqslant T}|X_t|^2\Big]\Big). \tag{4.17}$$

Similarly, by Jensen's inequality and (2.1), we obtain

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\Big[\big|\nabla(W * \mathrm{P}_t)(X_t)\big|^2\Big] &\leqslant \int_{\mathbb{R}^n\times\mathbb{R}^n} |\nabla W(x-y)|^2\, p_t(y)\, p_t(x)\,\mathrm{d}y\,\mathrm{d}x && (4.18)\\
&\leqslant 2C^2\bigg(1 + \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\, p_t(y)\, p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) && (4.19)\\
&\leqslant 2C^2\bigg(1 + 2\int_{\mathbb{R}^n\times\mathbb{R}^n} (|x|^2 + |y|^2)\, p_t(y)\, p_t(x)\,\mathrm{d}y\,\mathrm{d}x\bigg) && (4.20)\\
&= 2C^2\Big(1 + 4\,\mathbb{E}_{\mathbb{P}}\big[|X_t|^2\big]\Big) \leqslant 8C^2\Big(1 + \mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leqslant t\leqslant T}|X_t|^2\Big]\Big). && (4.21)
\end{aligned}$$

Altogether, we get

$$\mathbb{E}_{\mathbb{P}}\Big[\int_0^T |\nabla\Psi^{\uparrow}(X_t,\mathrm{P}_t)|^2\,\mathrm{d}t\Big] \leqslant 20T\Big(1+\mathbb{E}_{\mathbb{P}}\Big[\sup_{0\leqslant t\leqslant T}|X_t|^2\Big]\Big) < \infty, \tag{4.22}$$

where the finiteness of this expression follows from the uniform second moment property (2.2) of Lemma 2.2. From [LS01, 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

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}_{\mathrm{P}_0,2}}\bigg|_{\mathcal{F}_t} = \exp\bigg(-2\int_0^t \big\langle\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u)\,,\,\sqrt{2}\,\mathrm{d}B_u\big\rangle + \int_0^t |\nabla\Psi^{\uparrow}(X_u,\mathrm{P}_u)|^2\,\mathrm{d}u\bigg), \tag{4.23}$$

for $0\leqslant t\leqslant T$. The integrability property (4.22) implies that the $\mathbb{P}$-expectation of the stochastic integral in (4.23) vanishes, and we obtain

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

which shows (4.16). □

Denoting n-dimensional Lebesgue measure by λ, we consider on $\Omega = C([0,T];\mathbb{R}^n)$ the σ-finite measure

$$\mathbb{W}_{\lambda,2} := \int_{\mathbb{R}^n} \mathbb{W}_{x,2}\,\mathrm{d}\lambda(x),$$

which is known as the law of the *reversible Brownian motion* on $\mathbb{R}^n$ with variance 2; see [Léo14a, Léo14b]. 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 $\mathbb{W}_{\mathrm{P}_T,2} := \int_{\mathbb{R}^n}\mathbb{W}_{x,2}\,\mathrm{dP}_T(x)$ and its time reversal given by $\overline{\mathbb{W}}_{\mathrm{P}_T,2} = \int_{\mathbb{R}^n}\overline{\mathbb{W}}_{x,2}\,\mathrm{dP}_T(x)$. Then, as already noted in [Föl85, Remarks 3.7], we have the following result.

Lemma 4.7. *We have the relative entropy relations*

$$H(\mathbb{P}\,|\,\mathbb{W}_{\lambda,2}) = H(\mathrm{P}_0\,|\,\lambda) + H(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}) \tag{4.24}$$

and

$$H(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\lambda,2}) = H(\mathrm{P}_T\,|\,\lambda) + H(\mathbb{P}\,|\,\overline{\mathbb{W}}_{\mathrm{P}_T,2}). \tag{4.25}$$

Furthermore, all these relative entropies are finite.

Proof. For any $x\in\mathbb{R}^n$, 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 [Léo14b, Theorem 2] we have

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

and at the same time

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

implying the first identity (4.24). Regarding the finite entropy assertions, we recall (1.2), (1.3) and observe that

$$H(\mathrm{P}_0\,|\,\lambda) = \int_{\mathbb{R}^n} p_0(x)\log p_0(x)\,\mathrm{d}x = \mathcal{U}(\mathrm{P}_0) \leqslant \mathscr{F}(\mathrm{P}_0) < \infty,$$

where the finiteness follows from Assumptions 2.1, (ii). Furthermore, from Lemma 4.6 we know that $H(\mathbb{P}\,|\,\mathbb{W}_{\mathrm{P}_0,2}) < \infty$.

By the same arguments as above, (4.25) 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 $\mathrm{P}_T \in \mathscr{P}_{\mathrm{ac},2}(\mathbb{R}^n)$ by Lemma 2.2. 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. 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 3.1.

Proof of Theorem 3.1. Recalling the stochastic integral process $(\overline{M}_s)_{0\leqslant s\leqslant T}$ of (3.6) and the definition of the cumulative Fisher information process $(\overline{F}_s)_{0\leqslant s\leqslant T}$ in (3.3), we see that the stochastic differential of (4.13) – (4.15) can be expressed as claimed in (3.5).

Since $(\overline{M}_s)_{0\leqslant s\leqslant T}$ 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

$$\mathbb{E}_{\mathbb{P}}\Big[\langle \overline{M}, \overline{M}\rangle_T\Big] = \mathbb{E}_{\mathbb{P}}\bigg[2\int_0^T \big|\nabla \log \overline{\ell}_u(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2\,\mathrm{d}u\bigg] < \infty; \tag{4.26}$$

see, e.g. [RY99, Corollary IV.1.25]. On $(\Omega, \mathbb{G}, \mathbb{P})$, the time-reversed canonical process $(\overline{X}_s)_{0\leqslant s\leqslant T}$ has backward dynamics

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

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

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

for $(s, x) \in [0, T] \times \mathbb{R}^n$. Therefore, in order to prove (4.26), it suffices to show the two integrability conditions

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

The first condition is a direct consequence of (4.17). From [Föl85, 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 4.7.

In order to complete the proof of Theorem 3.1, it remains to show (3.7). To begin with, we take expectation with respect to $\mathbb{P}$ in (3.3) 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 (3.2). Next, we recall Assumptions 2.1, (i) and use the symmetry of the interaction potential, which implies that $\nabla W(-x) = -\nabla W(x)$ for all $x \in \mathbb{R}^n$. 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

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

for $0 \leqslant s \leqslant T$. Recalling the definitions in (2.3) – (2.5), we obtain

$$\mathbb{E}_{\mathbb{P}}\big[\overline{F}_s\big] = \int_0^s \mathbb{E}_{\mathbb{P}}\Big[\big|\nabla \log \overline{\ell}_u^{\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u)\big|^2\Big]\,\mathrm{d}u = \int_0^s I\big(\overline{\mathrm{P}}_u \,|\, \overline{\mathrm{Q}}_u^{\uparrow}\big)\,\mathrm{d}u < \infty, \tag{4.27}$$

for $0 \leqslant s \leqslant T$, where the second equality is immediate from (2.7), and the finiteness of the expression in (4.27) is justified as follows. Again, from (2.3) – (2.5) we find

$$\begin{aligned}
\big|\nabla \log \overline{\ell}_s^{\uparrow}(\overline{X}_s, \overline{\mathrm{P}}_s)\big|^2 &= \big|\nabla \log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s) + \tfrac{1}{2}\nabla(W * \overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2 \\
&\leqslant 2\big|\nabla \log \overline{\ell}_s(\overline{X}_s, \overline{\mathrm{P}}_s)\big|^2 + \tfrac{1}{2}\big|\nabla(W * \overline{\mathrm{P}}_s)(\overline{X}_s)\big|^2.
\end{aligned}$$

In light of (4.26) and (4.18) – (4.21), we see that the expression in (4.27) is finite, which in turn justifies a posteriori the former applications of Fubini's theorem. □

Remark 4.8. In the above proofs of Lemmas 4.6, 4.7 and Theorem 3.1, the linear growth condition (2.1) from Assumptions 2.1, (i) was crucial. For a different proof in the (linear) setting without interaction (i.e., $W \equiv 0$), we recall the proof of Theorem 4.1 in Part I, where the confinement potential V (which is called Ψ in Part I) is not necessarily of linear growth, but instead satisfies a weaker coercivity condition; see (2.7) in Chapter 2 of Part I. The key in that setting is to apply Lemma 2.48 in [KK21]; recall the proof of Theorem 4.1 in Chapter 5 of Part I and in particular Proposition 6.4 in Chapter 6 of Part I. ◇

The proof of Theorem 3.7 is now an easy consequence.

Proof of Theorem 3.7. Recalling the definition of the process $(\overline{M}_s^{\beta})_{0 \leqslant s \leqslant T - t_0}$ in (3.28) and of the perturbed cumulative Fisher information process $(\overline{F}_s^{\beta})_{0 \leqslant s \leqslant T - t_0}$ in (3.25), we see that the stochastic differential of (4.2) – (4.5) can be expressed as claimed in (3.27).

As in the proof of Theorem 3.1 we will now argue that

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\langle \overline{M}^{\boldsymbol{\beta}}, \overline{M}^{\boldsymbol{\beta}}\rangle_{T-t_0}\Big] = \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta}}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 \,\mathrm{d}u\bigg] < \infty, \tag{4.28}$$

which will then imply that the stochastic integral process $\overline{M}^{\boldsymbol{\beta}}$ is an $L^2(\mathbb{P}^{\boldsymbol{\beta}})$-bounded martingale. To this end, we define the density

$$q^{\boldsymbol{\beta}}(x,\mu) \coloneqq \mathrm{e}^{-\Psi^{\boldsymbol{\beta}}(x,\mu)}, \qquad (x,\mu) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$$

and consider the "doubly perturbed" likelihood ratio function

$$\ell_t^{\boldsymbol{\beta},\boldsymbol{\beta}}(x,\mu) \coloneqq \frac{p_t^{\boldsymbol{\beta}}(x)}{q^{\boldsymbol{\beta}}(x,\mu)}, \qquad (t,x) \in [t_0,T] \times \mathbb{R}^n.$$

As the Assumptions 2.1 are invariant under the passage from the potential V to the perturbed potential $V^{\boldsymbol{\beta}} = V + \boldsymbol{\beta}$, we can apply Theorem 3.1 to the potential $V^{\boldsymbol{\beta}}$ and obtain

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\bigg[2\int_0^{T-t_0} \big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta},\boldsymbol{\beta}}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 \,\mathrm{d}u\bigg] < \infty. \tag{4.29}$$

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

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

is a bounded function. Together with (4.29), this implies (4.28).

It remains to check (3.29). A similar calculation as in the proof of Theorem 3.1 leads to the identity

$$\mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\big[\overline{F}_s^{\boldsymbol{\beta}}\big] = \int_0^s \mathbb{E}_{\mathbb{P}^{\boldsymbol{\beta}}}\Big[\big|\nabla \log \overline{\ell}_u^{\boldsymbol{\beta}\uparrow}(\overline{X}_u, \overline{\mathrm{P}}_u^{\boldsymbol{\beta}})\big|^2 + \Big(\Big\langle \nabla V + \nabla(W * \overline{\mathrm{P}}_u^{\boldsymbol{\beta}}), \nabla \boldsymbol{\beta}\Big\rangle - \Delta \boldsymbol{\beta}\Big)(\overline{X}_u)\Big] \,\mathrm{d}u \tag{4.30}$$

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

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

Since the function

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

is bounded, we conclude that the quantity of (4.30) is finite. Finally, recalling the definition (3.24), we arrive at (3.29). □

4.2. The proofs of Proposition 3.14 and Theorem 3.16

Proof of Proposition 3.14. The first step is to view the probability density functions $(\rho_t)_{0\leqslant t\leqslant 1}$, corresponding to the displacement interpolation $(\nu_t)_{0\leqslant t\leqslant 1}$ of (3.41), as a solution to a continuity equation. Recalling the convex function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ of (3.40), we define a function $\mathsf{u}_0\colon \mathbb{R}^n \to \mathbb{R}$ by $\mathsf{u}_0(x) := \varphi(x) - |x|^2/2$; and for each $t \in (0,1]$, we let the function $\mathsf{u}_t\colon \mathbb{R}^n \to \mathbb{R}$ be defined by the Hopf–Lax formula

$$\mathsf{u}_t(x) := \inf_{y\in\mathbb{R}^n} \Big(\mathsf{u}_0(y) + \frac{|x-y|^2}{2t}\Big).$$

For all $t \in [0,1)$, we denote the gradient of u_t by $\mathsf{v}_t := \nabla \mathsf{u}_t$. 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 [Vil03, Theorem 5.51 (i)], and

$$\mathsf{v}_t(x) = \nabla \mathsf{u}_0 \circ (T_t)^{-1}(x), \qquad \text{for all } x \in T_t(\mathbb{R}^n), \tag{4.31}$$

where T_t is defined in (3.41). Note that the inverse of T_t is well-defined because T_t is injective; see [Vil03, Section 5.4.8]. From (4.31) we see that $(\mathsf{v}_t)_{0\leqslant t<1}$ is the *velocity field* associated with the trajectories $(T_t)_{0\leqslant t<1}$, i.e.,

$$T_t(x) = x + \int_0^t \mathsf{v}_s(T_s(x))\,\mathrm{d}s, \qquad 0 \leqslant t < 1. \tag{4.32}$$

By [Vil03, Theorem 5.51 (ii)], the curve of probability density functions $(\rho_t)_{0<t<1}$ satisfies the continuity equation

$$\partial_t \rho_t(x) + \operatorname{div}\big(\rho_t(x)\,\mathsf{v}_t(x)\big) = 0, \qquad (t,x) \in (0,1)\times\mathbb{R}^n. \tag{4.33}$$

On a sufficiently rich probability space $(S, \mathcal{S}, \mathsf{P})$, we let $Z_0\colon S \to \mathbb{R}^n$ be a random variable with probability distribution ν_0. For each $0 < t \leqslant 1$, we let $Z_t := T_t(Z_0)$. From (3.41) we see that the random variable Z_t has distribution ν_t, and (4.32) yields the representation

$$Z_t = Z_0 + \int_0^t \mathsf{v}_s(Z_s)\,\mathrm{d}s, \qquad 0 \leqslant t < 1. \tag{4.34}$$

In conjunction with (4.33), we deduce

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

and thus

$$\mathrm{d}\log\rho_t(Z_t) = -\operatorname{div}\big(\mathsf{v}_t(Z_t)\big)\,\mathrm{d}t. \tag{4.35}$$

Recalling the definition of the density function q in (2.4), a similar argument as in (4.8) shows that

$$\big(\partial_\nu \log q(x,\nu)\big)(y) = \tfrac{1}{2}\big(\partial_\nu\,(W*\nu)\,(x)\big)(y) = -\tfrac{1}{2}\nabla W(x-y) \tag{4.36}$$

for $(x, \nu, y) \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n) \times \mathbb{R}^n$. Applying a generalized version of Itô's formula for McKean–Vlasov diffusions [CD18, Proposition 5.102], and using the dynamics (4.34) as well as the L-derivative (4.36), we obtain

$$\mathrm{d}\log q(Z_t, \nu_t) = -\Big\langle \nabla V + \tfrac{1}{2}\nabla(W * \nu_t)\,,\, \mathsf{v}_t \Big\rangle(Z_t)\,\mathrm{d}t + \tfrac{1}{2}\mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t)\,,\, \mathsf{v}_t(\tilde{Z}_t)\Big\rangle\Big]\,\mathrm{d}t \tag{4.37}$$

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 (4.35) and (4.37) gives the dynamics

$$\log r_t(Z_t, \nu_t) - \log r_0(Z_0, \nu_0) \tag{4.38}$$

$$= \int_0^t \Big(\Big\langle \nabla V + \tfrac{1}{2}\nabla(W * \nu_s)\,,\, \mathsf{v}_s \Big\rangle(Z_s) - \operatorname{div}(\mathsf{v}_s(Z_s))\Big)\,\mathrm{d}s \tag{4.39}$$

$$- \tfrac{1}{2}\int_0^t \mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_s - \tilde{Z}_s)\,,\, \mathsf{v}_s(\tilde{Z}_s)\Big\rangle\Big]\,\mathrm{d}s \tag{4.40}$$

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

$$\mathbb{E}_{\mathsf{P}}\Big[\operatorname{div}(\mathsf{v}_t(Z_t))\Big] = -\mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla \log \rho_t(Z_t)\,,\, \mathsf{v}_t(Z_t)\Big\rangle\Big]. \tag{4.41}$$

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

$$\mathbb{E}_{\mathsf{P}}\Big[\mathbb{E}_{\tilde{\mathsf{P}}}\Big[\Big\langle \nabla W(Z_t - \tilde{Z}_t)\,,\, \mathsf{v}_t(\tilde{Z}_t)\Big\rangle\Big]\Big] = -\mathbb{E}_{\mathsf{P}}\Big[\Big\langle \nabla(W * \nu_t)(Z_t)\,,\, \mathsf{v}_t(Z_t)\Big\rangle\Big]. \tag{4.42}$$

Returning to (4.38) – (4.40), we take P-expectations and use (4.41), (4.42) to obtain

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

where for the second equality we recall the notations in (3.42) and (2.4). Finally, letting $t \downarrow 0$, we get

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

and since $\mathsf{v}_0 = \nabla\varphi - \mathrm{Id}$, we arrive at (3.43). □

Proof of Theorem 3.16. Without loss of generality, we assume that the probability density functions ρ_0 and ρ_1 satisfy the strong regularity Assumptions 3.13. The general case then follows by a density argument. We will not provide the details here, but refer

to [Vil03, Chapter 9.4], where this regularization is carried out in the simpler setting of the HWI inequality.

Let us recall the energy functionals $\mathcal{U}$, $\mathcal{V}$, $\mathcal{W}$ defined in (1.3) – (1.5), and introduce the functions

$$f(t) := \mathcal{U}(\rho_t), \qquad g(t) := \mathcal{V}(\rho_t), \qquad h(t) := \mathcal{W}(\rho_t), \qquad 0 \leqslant t \leqslant 1,$$

where $(\rho_t)_{0\leqslant t\leqslant 1}$ is the curve of probability density functions corresponding to the displacement interpolation $(\nu_t)_{0\leqslant t\leqslant 1}$ of (3.41). Then the sum $F := f + g + h$ of these functions satisfies the relation $F(t) = H(\nu_t \,|\, \mu_t)$. In light of [Vil03, Theorem 5.15 (i)], the internal energy functional $\mathcal{U}$ is displacement convex, i.e.,

$$f''(t) \geqslant 0, \qquad 0 \leqslant t \leqslant 1. \tag{4.43}$$

By Assumptions 3.15, the confinement potential $V\colon \mathbb{R}^n \to [0,\infty)$ is κ_V-uniformly convex. Therefore, [Vil03, Theorem 5.15 (ii)] implies that the potential energy functional $\mathcal{V}$ is κ_V-uniformly displacement convex. In other words,

$$g''(t) \geqslant \kappa_V\, W_2^2(\nu_0,\nu_1), \qquad 0 \leqslant t \leqslant 1. \tag{4.44}$$

Again from Assumptions 3.15, the interaction potential $W\colon \mathbb{R}^n \to [0,\infty)$ is assumed to be symmetric and κ_W-uniformly convex. Therefore, a similar argument as in the proof of [Vil03, Theorem 5.15 (iii)] leads to the $\kappa_W(W_2^2(\nu_0,\nu_1) - |b(\nu_0) - b(\nu_1)|^2)$-uniform convexity of h, so

$$h''(t) \geqslant \kappa_W \Big(W_2^2(\nu_0,\nu_1) - |b(\nu_0) - b(\nu_1)|^2 \Big), \qquad 0 \leqslant t \leqslant 1. \tag{4.45}$$

The details of the proof of (4.45) are postponed to Section 5.2. By combining the estimates (4.43) – (4.45), we deduce that the relative entropy function

$$[0,1] \ni t \longmapsto F(t) = H(\nu_t \,|\, \mu_t)$$

satisfies the inequality

$$F''(t) \geqslant (\kappa_V + \kappa_W)\, W_2^2(\nu_0,\nu_1) - \kappa_W\, |b(\nu_0) - b(\nu_1)|^2. \tag{4.46}$$

Furthermore, from Proposition 3.14 we have

$$F'(0^+) = \int_{\mathbb{R}^n} \Big\langle \nabla \log r_0^{\uparrow}(x,\nu_0)\,, \nabla\varphi(x) - x \Big\rangle\, \rho_0(x)\,\mathrm{d}x. \tag{4.47}$$

In conjunction with (4.46) and (4.47), the Taylor formula

$$F(1) = F(0) + F'(0^+) + \int_0^1 (1-t) F''(t)\,\mathrm{d}t$$

now yields the inequality (3.45), (3.46). □

5. Proofs of auxiliary results

5.1. The proof of Lemma 2.2

Proof of Lemma 2.2. The generalized potential $\Psi^{\uparrow}(\,\cdot\,,\,\cdot\,)$ of (2.3) allows us to cast the McKean–Vlasov dynamics of (1.1) in the more compact form

$$\mathrm{d}X_t = -\nabla\Psi^{\uparrow}(X_t, \mathrm{P}_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \qquad 0 \leqslant t \leqslant T. \tag{5.1}$$

Then, for any two pairs $(x,\mu), (x',\mu') \in \mathbb{R}^n \times \mathscr{P}_2(\mathbb{R}^n)$, using the Lipschitz continuity of ∇V in Assumptions 2.1, (i) yields

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

For the convolution term, using Jensen's inequality and the Lipschitz continuity of ∇W in Assumptions 2.1, (i) leads to

$$|\nabla(W*\mu)(x) - \nabla(W*\mu')(x')| \leqslant \|\nabla W\|_{\mathrm{Lip}}\,|x-x'| + |\nabla(W*\mu)(x') - \nabla(W*\mu')(x')|.$$

For the last term above, by the Kantorovich–Rubinstein theorem [Vil03, Theorem 1.14], we have

$$\begin{aligned}
&\biggl|\int_{\mathbb{R}^n} \nabla W(x'-\,\cdot\,)\,\mathrm{d}(\mu-\mu')\biggr| \\
&\qquad \leqslant \|\nabla W\|_{\mathrm{Lip}}\,\sup\biggl\{\int_{\mathbb{R}^n}\varphi\,\mathrm{d}(\mu-\mu') \colon \varphi \in L^1(|\mu-\mu'|), \|\varphi\|_{\mathrm{Lip}} \leqslant 1\biggr\} \\
&\qquad = \|\nabla W\|_{\mathrm{Lip}}\,W_1(\mu,\mu') \leqslant \|\nabla W\|_{\mathrm{Lip}}\,W_2(\mu,\mu'),
\end{aligned} \tag{5.2}$$

where

$$W_1(\mu,\mu') \coloneqq \inf_{Y\sim\mu, Z\sim\mu'} \mathbb{E}\bigl[|Y-Z|\bigr], \qquad \mu,\mu' \in \mathscr{P}_1(\mathbb{R}^n)$$

denotes the 1-Wasserstein-distance, and the estimate in (5.2) follows from Jensen's inequality. Altogether, we obtain

$$|\nabla\Psi^{\uparrow}(x,\mu) - \nabla\Psi^{\uparrow}(x',\mu')| \leqslant (\|\nabla V\|_{\mathrm{Lip}} + \|\nabla W\|_{\mathrm{Lip}})\,|x-x'| + \|\nabla W\|_{\mathrm{Lip}}\,W_2(\mu,\mu'). \tag{5.3}$$

In particular, this shows that the function $-\nabla\Psi^{\uparrow}(\,\cdot\,,\,\cdot\,)$ is Lipschitz continuous on the product metric space $(\mathbb{R}^n, |\,\cdot\,|) \times (\mathscr{P}_2(\mathbb{R}^n), W_2)$. In conjunction with Assumptions 2.1, (ii), [CD18, Theorem 4.21] implies that the McKean–Vlasov stochastic differential equation (5.1) has a pathwise unique, strong solution satisfying the uniform second moment

condition (2.2). Now we can linearize (5.1) by fixing the time-marginals $(\mathrm{P}_t)_{0\leqslant t\leqslant T}$, so that the drift term can be viewed as a function $(t,x)\mapsto \nabla\Psi^{\uparrow}(x,\mathrm{P}_t)$, and (5.1) becomes an ordinary stochastic differential equation with a time-inhomogeneous drift coefficient.

The absolute continuity of the time-marginals $(\mathrm{P}_t)_{0\leqslant t\leqslant T}$ is immediate from Lemma 4.6. A standard argument using the classical Itô's formula shows that the curve of probability density functions $(p_t)_{0\leqslant t\leqslant T}$ is a weak solution of the granular media equation (1.12). Finally, we turn to the regularity of this solution. From (5.3), 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 2.1, (i) implies that the drift is also of linear growth. The desired smoothness of $(p_t)_{0\leqslant t\leqslant T}$ now follows from a straightforward adaptation of the theorem in [Rog85], see also Remarks (i) – (ii) therein. □

5.2. The proof of (4.45)

Proof of (4.45): We first rewrite the interaction energy functional $\mathcal{W}$ of (1.5) along the displacement interpolation $(\nu_t)_{0\leqslant t\leqslant 1}$. Using (3.41), for any $t\in[0,1]$, we have

$$\begin{aligned}
h(t) &= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W(x-y)\,\nu_t(\mathrm{d}x)\,\nu_t(\mathrm{d}y)\\
&= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\big(T_t(x)-T_t(y)\big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y)\\
&= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} W\Big(x-y-t\big(\theta(x)-\theta(y)\big)\Big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y),
\end{aligned}$$

where $\theta\colon\mathbb{R}^n\to\mathbb{R}^n$ 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 3.15, we obtain

$$\begin{aligned}
&\sigma h(t_1)+(1-\sigma)h(t_2)-h(\sigma t_1+(1-\sigma)t_2)\\
&\quad= \tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n}\Big(\sigma W\Big(x-y-t_1\big(\theta(x)-\theta(y)\big)\Big)+(1-\sigma)W\Big(x-y-t_2\big(\theta(x)-\theta(y)\big)\Big)\\
&\qquad- W\Big(x-y-\big(\sigma t_1+(1-\sigma)t_2\big)\big(\theta(x)-\theta(y)\big)\Big)\Big)\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y)\\
&\quad\geqslant \tfrac{1}{4}\kappa_W\sigma(1-\sigma)(t_1-t_2)^2\int_{\mathbb{R}^n\times\mathbb{R}^n}|\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y). \qquad (5.4)
\end{aligned}$$

Next, we express the integral in (5.4) as

$$\begin{aligned}
&\tfrac{1}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} |\theta(x)-\theta(y)|^2\,\nu_0(\mathrm{d}x)\,\nu_0(\mathrm{d}y)\\
&\quad= \int_{\mathbb{R}^n} |\theta(x)|^2\,\nu_0(\mathrm{d}x) - \Big|\int_{\mathbb{R}^n} \theta(x)\,\nu_0(\mathrm{d}x)\Big|^2\\
&\quad= \int_{\mathbb{R}^n} |x-\nabla\varphi(x)|^2\,\nu_0(\mathrm{d}x) - \Big|\int_{\mathbb{R}^n} x\,\nu_0(\mathrm{d}x) - \int_{\mathbb{R}^n} x\,\nu_1(\mathrm{d}x)\Big|^2\\
&\quad= W_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2.
\end{aligned}$$

Putting this back into (5.4), we deduce that h is uniformly convex, with constant

$$\kappa_W\Big(W_2^2(\nu_0,\nu_1) - |b(\nu_0)-b(\nu_1)|^2\Big),$$

which shows (4.45). □

Part IV.

Diffusion processes as Wasserstein gradient flows via stochastic control of the volatility matrix

Abstract

We consider a class of time-homogeneous diffusion processes on $\mathbb{R}^n$ with common invariant measure but varying volatility matrices. In Euclidean space, we show that the corresponding flow of time-marginal distributions admits an entropic gradient flow formulation in the quadratic Wasserstein space if the volatility matrix of the diffusion is the identity. Our approach is based on stochastic control of the diffusion coefficient and time-reversal techniques. Equipping $\mathbb{R}^n$ with a Riemannian metric, we prove that the diffusion process can be viewed as a gradient flow in the induced Wasserstein space if the volatility matrix is the inverse of the underlying metric tensor. In the Euclidean case, our probabilistic result corresponds to the gradient flow characterization of the Fokker–Planck equation, first discovered in a seminal paper by Jordan, Kinderlehrer, and Otto. In the Riemannian setting, the corresponding result on the level of partial differential equations was established by Lisini, building on the metric theory developed by Ambrosio, Gigli, and Savaré. This setting was also used by Fathi to develop a gradient flow approach to large deviations for diffusion processes.

1. Introduction

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant T}$, we consider a class of time-homogeneous diffusion processes $(X_t^\Sigma)_{0 \leqslant t \leqslant T}$ on $\mathbb{R}^n$ of the form

$$\mathrm{d}X_t^\Sigma = \Big(\operatorname{div} \Sigma(X_t^\Sigma) - \Sigma(X_t^\Sigma) \nabla V(X_t^\Sigma) \Big) \,\mathrm{d}t + \sqrt{2\Sigma(X_t^\Sigma)} \,\mathrm{d}B_t, \tag{1.1}$$

for $0 \leqslant t \leqslant T < \infty$; with initial condition $X_0^\Sigma = Y$, a square-integrable random vector with continuous probability density function $p_0^\Sigma \colon \mathbb{R}^n \to (0, \infty)$, which is independent of the standard n-dimensional Brownian motion $(B_t)_{0 \leqslant t \leqslant T}$. We denote by $P_t^\Sigma \coloneqq (X_t^\Sigma)_\# \mathbb{P}$ the distribution of the random vector X_t^Σ and by $p_t^\Sigma \colon \mathbb{R}^n \to (0, \infty)$ the corresponding probability density function. The functions $V \colon \mathbb{R}^n \to \mathbb{R}$ and $\Sigma \colon \mathbb{R}^n \to \mathbb{R}^{n \times n}$ are smooth and sufficiently well-behaved to guarantee the existence of a pathwise unique, strong solution of the stochastic differential equation (1.1) in $L^2(\mathbb{P})$, as well as a classical solution of the corresponding Fokker–Planck equation [Fri75, Gar09, Ris96, Sch80]

$$\partial_t p_t^\Sigma(x) = \operatorname{div} \Big(\Sigma(x) \left(\nabla \log p_t^\Sigma(x) + \nabla V(x) \right) p_t^\Sigma(x) \Big), \qquad (t, x) \in (0, T) \times \mathbb{R}^n, \tag{1.2}$$

which is also known as Kolmogorov forward equation [Kol31].

The function $\Sigma \colon \mathbb{R}^n \to \mathbb{R}^{n \times n}$ takes values in the set S_{++}^n of symmetric and positive definite matrices and we denote by $\sqrt{\Sigma(x)} \in S_{++}^n$ the unique square root of $\Sigma(x) \in S_{++}^n$. The divergence of the matrix-valued function Σ is the vector field $\operatorname{div} \Sigma \colon \mathbb{R}^n \to \mathbb{R}^n$ with components $(\operatorname{div} \Sigma)_i = \sum_{j=1}^n \partial_j \Sigma_{ij}$.

The aim of this work is to study entropy dissipation and gradient flow properties of the diffusion equation (1.1). The novelty of our stochastic analysis approach is that it is based on stochastic control of the volatility matrix. Our primary inspiration derives from [KST22], where a similar endeavor has been established in the case when the volatility matrix Σ is equal to the identity matrix.

1.1. Invariant measure

The stationary version

$$\operatorname{div} \Big(\Sigma(x) \left(\nabla \log q(x) + \nabla V(x) \right) q(x) \Big) = 0, \qquad x \in \mathbb{R}^n \tag{1.3}$$

of the Fokker–Planck equation (1.2) has a strictly positive solution $q(x) \coloneqq \mathrm{e}^{-V(x)}$. We denote by Q the σ-finite measure on the Borel sets $\mathscr{B}(\mathbb{R}^n)$ having the function q as

density with respect to Lebesgue measure. This measure Q is invariant for the diffusion process $(X_t^\Sigma)_{0\leqslant t\leqslant T}$. If

$$\mathrm{Q}(\mathbb{R}^n) = \int_{\mathbb{R}^n} \mathrm{e}^{-V(x)}\,\mathrm{d}x < \infty, \tag{1.4}$$

the measure Q can be normalized to become a probability measure (the so-called *Gibbs measure* from statistical mechanics). We emphasize that we do not impose the condition (1.4) and admit the possibility that the total mass $\mathrm{Q}(\mathbb{R}^n)$ is infinite. This allows for the important case when the function V vanishes identically and Q is equal to Lebesgue measure on $\mathbb{R}^n$. Instead, we require the weaker integrability condition

$$\mathbb{E}_{\mathrm{Q}}\Big[(x \mapsto \mathrm{e}^{-|x|^2})\Big] = \int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2 - V(x)}\,\mathrm{d}x < \infty. \tag{1.5}$$

If, for example, V is nonnegative, this condition is automatically satisfied. Furthermore, we stress that no convexity assumptions are imposed on the function V.

We remark that, if $\operatorname{div}\Sigma = \Sigma\,\nabla V$, the process $(X_t^\Sigma)_{0\leqslant t\leqslant T}$ is a martingale. In this case, if $n = 1$, the invariant measure Q has density $q(x) = \frac{1}{\Sigma(x)}$.

1.2. Relative entropy and free energy

Given the flow of probability measures $(P_t^\Sigma)_{0\leqslant t\leqslant T}$ defined by the diffusion process of (1.1), our goal is to study the *relative entropy function*

$$[0,T] \ni t \longmapsto H(P_t^\Sigma \,|\, \mathrm{Q}). \tag{1.6}$$

The *relative entropy* or *Kullback–Leibler divergence* of $P_t^\Sigma = \mathrm{Law}(X_t^\Sigma)$ with respect to the invariant measure Q is equal to

$$H(P_t^\Sigma \,|\, \mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\big[\log \ell_t^\Sigma(X_t^\Sigma)\big] = \int_{\mathbb{R}^n} \log\left(\frac{p_t^\Sigma(x)}{q(x)}\right) p_t^\Sigma(x)\,\mathrm{d}x, \tag{1.7}$$

where the *likelihood ratio function* $(t,x) \mapsto \ell_t^\Sigma(x)$ is given by

$$\ell_t^\Sigma(x) := \frac{p_t^\Sigma(x)}{q(x)} = p_t^\Sigma(x)\,\mathrm{e}^{V(x)}, \qquad (t,x) \in [0,T]\times\mathbb{R}^n. \tag{1.8}$$

The relative entropy (1.7) can also be interpreted as a *free energy*. Defining the *free energy functional* as the sum of the internal and potential energy functionals via

$$p \longmapsto \mathscr{F}(p) := \int_{\mathbb{R}^n} p(x)\log p(x)\,\mathrm{d}x + \int_{\mathbb{R}^n} V(x)\,p(x)\,\mathrm{d}x, \tag{1.9}$$

for probability densities $p(\,\cdot\,)$ on $\mathbb{R}^n$, we have the relation $H(P_t^\Sigma \,|\, \mathrm{Q}) = \mathscr{F}(p_t^\Sigma)$ between entropy and energy.

Since the diffusion process $(X_t^\Sigma)_{0\leqslant t\leqslant T}$ lies in $L^2(\mathbb{P})$, the corresponding time-marginals $(P_t^\Sigma)_{0\leqslant t\leqslant T}$ are elements of $\mathscr{P}_2(\mathbb{R}^n)$, the collection of probability measures on $\mathscr{B}(\mathbb{R}^n)$ with finite second moments. Together with the integrability condition (1.5), this ensures that the relative entropy (1.7) with respect to the σ-finite reference measure Q is well-defined and takes values in the interval $(-\infty,\infty]$; we refer to Section 6.1 of Part I or [Léo14b, Section 3] for the details.

1.3. Riemannian structure

We fix a function $A \in \mathcal{S}$, which then induces a smooth and symmetric metric tensor

$$G\colon \mathbb{R}^n \longrightarrow \mathbb{R}^{n\times n}\colon x \longmapsto G(x) \coloneqq A^{-1}(x).$$

This metric tensor G defines a Riemannian inner product $\langle\,\cdot\,,\cdot\,\rangle_G$ on $\mathbb{R}^n$, given by

$$\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}\colon (v,w) \longmapsto \langle v,w\rangle_G(x) \coloneqq \langle G(x)v, w\rangle$$

for $x \in \mathbb{R}^n$. We denote the induced norm by $|\cdot|_G \coloneqq \sqrt{\langle\,\cdot\,,\cdot\,\rangle_G}$. Equipped with this Riemannian structure, we can regard the pair $(\mathbb{R}^n, G)$ as a Riemannian manifold; we refer to [Lee18] for an introduction to Riemannian manifolds. The metric tensor G naturally induces a Riemannian metric d_G on $(\mathbb{R}^n, G)$ defined by

$$d_G(x,y) \coloneqq \inf_{\gamma\in\Gamma(x,y)} \left\{ \int_0^1 \sqrt{\Big\langle G(\gamma(t))\dot\gamma(t)\,,\dot\gamma(t)\Big\rangle}\,\mathrm{d}t \right\}, \qquad x,y\in\mathbb{R}^n,$$

where $\Gamma(x,y)$ denotes the collection of absolutely continuous curves $\gamma\colon [0,1]\to\mathbb{R}^n$ with starting point $\gamma(0)=x$ and endpoint $\gamma(1)=y$. We write $\mathbb{R}^n_G$ for the metric space $\mathbb{R}^n$ equipped with this metric d_G. As a consequence of (1.13), G satisfies a uniform ellipticity condition of the form

$$\forall x,\xi\in\mathbb{R}^n\colon \quad C_A^{-1}|\xi|^2 \leqslant \langle G(x)\xi,\xi\rangle \leqslant c_A^{-1}|\xi|^2. \tag{1.10}$$

Note that $\mathscr{P}_2(\mathbb{R}^n_G) = \mathscr{P}_2(\mathbb{R}^n)$ on account of (1.10). In other words, for a probability measure on $\mathbb{R}^n$, it is immaterial whether the finite second moment property is tested with respect to the metric d_G or the Euclidean metric, and by equivalence of norms the topology and the Borel sets remain the same.

1.4. Relative Fisher information

For a probability measure $P\in\mathscr{P}(\mathbb{R}^n)$ and vector fields $\boldsymbol{v},\boldsymbol{w}\colon\mathbb{R}^n\to\mathbb{R}^n$ we define the scalar product

$$\langle \boldsymbol{v},\boldsymbol{w}\rangle_{L^2_G(P)} \coloneqq \mathbb{E}_P\big[\langle \boldsymbol{v},\boldsymbol{w}\rangle_G\big] = \int_{\mathbb{R}^n} \langle G\boldsymbol{v},\boldsymbol{w}\rangle\,\mathrm{d}P$$

and the induced norm $\|\boldsymbol{v}\|_{L^2_G(P)} \coloneqq \sqrt{\langle \boldsymbol{v},\boldsymbol{v}\rangle_{L^2_G(P)}}$.

The gradient of a smooth function $f\colon\mathbb{R}^n\to\mathbb{R}$ on the Riemannian manifold $(\mathbb{R}^n,G)$ is given by $\nabla_G f \coloneqq A\nabla f$ (see, e.g., [Lee18]).

In our Riemannian setting, the *relative Fisher information* (see, e.g., [OV00, Vil03]) of $P_t^\Sigma = \mathrm{Law}(X_t^\Sigma)$ with respect to the invariant measure Q is equal to

$$I_G(P_t^\Sigma\,|\,\mathrm{Q}) \coloneqq \|\nabla_G \log \ell_t^\Sigma\|^2_{L^2_G(P_t^\Sigma)} = \int_{\mathbb{R}^n} \langle \nabla\log\ell_t^\Sigma, A\,\nabla\log\ell_t^\Sigma\rangle(x)\,p_t^\Sigma(x)\,\mathrm{d}x.$$

We can also express the relative Fisher information as a $\mathbb{P}$-expectation, namely

$$I_G(P_t^\Sigma\,|\,\mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\big[|\nabla_G\log\ell_t^\Sigma(X_t^\Sigma)|_G^2\big]. \tag{1.11}$$

For the classical definition of Fisher information in the information theory literature we refer to the books [CL98, CT06].

1.5. Wasserstein distance

The quadratic Wasserstein distance (see, e.g., [AGS08, Vil03]) between two probability measures P_1 and P_2 in $\mathscr{P}_2(\mathbb{R}^n)$, with $\mathbb{R}^n$ being equipped with the Riemannian metric d_G, is defined by

$$W_{2,G}(P_1, P_2) \coloneqq \sqrt{\inf_{\pi \in \Pi(P_1,P_2)} \int_{\mathbb{R}^n \times \mathbb{R}^n} d_G(x,y)^2 \,\mathrm{d}\pi(x,y)}\,, \tag{1.12}$$

where $\Pi(P_1, P_2)$ denotes the set of all couplings $\pi \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^n)$ between P_1 and P_2, i.e., probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with first marginal P_1 and second marginal P_2.

1.6. Class of admissible volatilities

Definition 1.1. We define the *class of admissible volatilities* $\mathcal{S}$ as the collection of all matrix-valued functions $\Sigma\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$ such that

(i) $\Sigma(x)$ is a symmetric and positive definite matrix, for every $x \in \mathbb{R}^n$,

(ii) the component functions $\mathbb{R}^n \ni x \mapsto \Sigma_{ij}(x) \in \mathbb{R}$ are smooth, for all $i, j \in \{1, \dots, n\}$,

(iii) for some constants $c, C > 0$ the uniform ellipticity condition

$$\forall x, \xi \in \mathbb{R}^n\colon \quad c|\xi|^2 \leqslant \langle \Sigma(x)\xi, \xi\rangle \leqslant C|\xi|^2 \tag{1.13}$$

is satisfied,

(iv) the stochastic differential equation (1.1) admits a pathwise unique, strong solution $(X_t^\Sigma)_{0\leqslant t\leqslant T}$, which is bounded in $L^2(\mathbb{P})$,

(v) the Fokker–Planck equation (1.2) has a classical solution $(p_t^\Sigma)_{0\leqslant t\leqslant T}$,

(vi) the curve of probability measures $(P_t^\Sigma)_{0\leqslant t\leqslant T}$ given by $P_t^\Sigma = \mathrm{Law}(X_t^\Sigma)$ is absolutely continuous on the quadratic Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_{2,G})$, i.e., there exists $m \in L^1([0,T])$ such that

$$W_{2,G}(P_{t_0}^\Sigma, P_t^\Sigma) \leqslant \int_{t_0}^t m(u)\,\mathrm{d}u$$

for all $0 \leqslant t_0 \leqslant t \leqslant T$,

(vii) and the initial relative entropy $H(P_0^\Sigma \,|\, \mathrm{Q})$ is finite.

1.7. Main results

Under the assumptions of this chapter, we can state our first main result.

Theorem 1.2. *For every $\Sigma \in \mathcal{S}$ and $0 \leqslant t_0 \leqslant t \leqslant T$ we have*

$$|H(P_t^\Sigma \,|\, \mathrm{Q}) - H(P_{t_0}^\Sigma \,|\, \mathrm{Q})| \leqslant \int_{t_0}^t \sqrt{I_G(P_u^\Sigma \,|\, \mathrm{Q})} \lim_{h \to 0} \frac{W_{2,G}(P_{u+h}^\Sigma, P_u^\Sigma)}{|h|} \,\mathrm{d}u, \tag{1.14}$$

with equality if the diffusion matrix Σ is a constant multiple of A. In the case $\Sigma = A$ we have

$$H(P_{t_0}^A \,|\, \mathrm{Q}) - H(P_t^A \,|\, \mathrm{Q}) = \tfrac{1}{2} \int_{t_0}^t I_G(P_u^A \,|\, \mathrm{Q}) \,\mathrm{d}u + \tfrac{1}{2} \int_{t_0}^t \Big(\lim_{h \to 0} \frac{W_{2,G}(P_{u+h}^A, P_u^A)}{|h|} \Big)^2 \,\mathrm{d}u. \tag{1.15}$$

The message of Theorem 1.2 is that the "flow of probability measures" $(P_t^A)_{0 \leqslant t \leqslant T}$, defined by the stochastic differential equation

$$\mathrm{d}X_t^A = \Big(\operatorname{div} A(X_t^A) - A(X_t^A) \, \nabla V(X_t^A) \Big) \,\mathrm{d}t + \sqrt{2A(X_t^A)} \,\mathrm{d}B_t, \tag{1.16}$$

is the *gradient flow* of the *relative entropy functional*

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P \,|\, \mathrm{Q}) \tag{1.17}$$

with respect to the quadratic Wasserstein distance $W_{2,G}$. The gradient flow property in this Riemannian setting was established by Lisini in [Lis09], where nonlinear diffusion equations with variable coefficients are identified as gradient flows in Wasserstein spaces with respect to a Riemannian metric. The approach of [Lis09] is based on the theory of minimizing movements and the theory of curves of maximal slope in metric spaces, taking as a reference [AGS08]. While in [Lis09] it is crucial to study the Fokker–Planck equation

$$\partial_t p_t^A(x) = \operatorname{div} \Big(A(x) \, (\nabla \log p_t^A(x) + \nabla V(x)) \, p_t^A(x) \Big), \tag{1.18}$$

our approach is based on the corresponding stochastic differential equation (1.16) via stochastic control of the diffusion coefficient. More precisely, we fix $A \in \mathcal{S}$, and for every Σ in the class of admissible volatilities $\mathcal{S}$, we vary the diffusion coefficient by considering a stochastic differential equation of the form (1.1). After some stochastic analysis and taking expectations this yields the inequality (1.14). In the language of analysis in metric spaces, the inequality (1.14) has the interpretation that the square root of the *relative Fisher information functional*

$$P \longmapsto \sqrt{I_G(P \,|\, \mathrm{Q})} = \|\nabla_G \log(\tfrac{p}{q})\|_{L^2_G(P)} = \sqrt{\int_{\mathbb{R}^n} \langle \nabla \log(\tfrac{p}{q}), A \, \nabla \log(\tfrac{p}{q}) \rangle(x) \, p(x) \,\mathrm{d}x} \tag{1.19}$$

is a *strong upper gradient* (see [AGS08, Definition 1.2.1], [Lis09, Definition 3.2]) for the relative entropy functional (1.17). The domain of these two functionals consists of the

collection of absolutely continuous probability measures $P \in \mathscr{P}_2(\mathbb{R}^n)$, having a probability density function $p(\,\cdot\,)$ with respect to n-dimensional Lebesgue measure. In the case $\Sigma = A$, we obtain the so-called *energy identity* (1.15) (see [AGS08, Remark 1.3.3], [Lis09, Definition 3.3]). This shows that $(P_t^A)_{0 \leqslant t \leqslant T}$ is a *curve of maximal slope* (see [AGS08, Definition 1.3.2], [Lis09, Definition 3.3]) for the relative entropy functional (1.17) with respect to its strong upper gradient (1.19) in the Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_{2,G})$. According to [AGS08, Theorem 11.1.3], *curves of maximal slope coincide with gradient flows*, which verifies the stated gradient flow property of $(P_t^A)_{0 \leqslant t \leqslant T}$.

We leave now the Riemannian setting and return to the standard Euclidean structure on $\mathbb{R}^n$. In other words, we fix the identity matrix $A = I_n \in \mathcal{S}$ and consider the quadratic Wasserstein distance

$$W_2(P_1, P_2) = \sqrt{\inf_{\pi \in \Pi(P_1, P_2)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \,\mathrm{d}\pi(x, y)}\,, \qquad P_1, P_2 \in \mathscr{P}_2(\mathbb{R}^n), \tag{1.20}$$

together with the relative Fisher information functional

$$P \longmapsto I(P \,|\, \mathrm{Q}) = \|\nabla \log(\tfrac{p}{q})\|^2_{L^2(P)} = \int_{\mathbb{R}^n} |\nabla \log(\tfrac{p(x)}{q(x)})|^2 \, p(x) \,\mathrm{d}x.$$

As an immediate consequence of Theorem 1.2, we obtain the following result.

Corollary 1.3. *For every $\Sigma \in \mathcal{S}$ and $0 \leqslant t_0 \leqslant t \leqslant T$ we have*

$$\big|H(P_t^\Sigma \,|\, \mathrm{Q}) - H(P_{t_0}^\Sigma \,|\, \mathrm{Q})\big| \leqslant \int_{t_0}^t \sqrt{I(P_u^\Sigma \,|\, \mathrm{Q})} \, \lim_{h \to 0} \frac{W_2(P_{u+h}^\Sigma, P_u^\Sigma)}{|h|} \,\mathrm{d}u,$$

with equality if the diffusion matrix Σ is a constant multiple of the identity matrix. In the case $\Sigma = I_n$ we have

$$H(P_{t_0} \,|\, \mathrm{Q}) - H(P_t \,|\, \mathrm{Q}) = \tfrac{1}{2} \int_{t_0}^t I(P_u \,|\, \mathrm{Q}) \,\mathrm{d}u + \tfrac{1}{2} \int_{t_0}^t \bigg(\lim_{h \to 0} \frac{W_2(P_{u+h}, P_u)}{|h|} \bigg)^2 \mathrm{d}u.$$

Here, the probability measures $(P_t)_{0 \leqslant t \leqslant T}$ denote the time-marginal distributions of the stochastic differential equation of Langevin–Smoluchowski type

$$\mathrm{d}X_t = -\nabla V(X_t) \,\mathrm{d}t + \sqrt{2} \,\mathrm{d}B_t. \tag{1.21}$$

The corresponding probability density functions $(p_t)_{0 \leqslant t \leqslant T}$ satisfy the Fokker–Planck equation

$$\partial_t p_t(x) = \operatorname{div}\Big(\big(\nabla \log p_t(x) + \nabla V(x)\big) \, p_t(x) \Big). \tag{1.22}$$

Clearly the equations (1.21) and (1.22) are obtained by taking A to be the n-dimensional identity matrix in (1.16) and (1.18), respectively. By analogy with the Riemannian setting considered above, Corollary 1.3 has a gradient flow interpretation in Euclidean space. Namely, it establishes that the flow of probability measures $(P_t)_{0 \leqslant t \leqslant T}$ is the gradient flow of the relative entropy functional (1.17) with respect to the quadratic

Wasserstein distance W_2 on the space $\mathscr{P}_2(\mathbb{R}^n)$. The gradient flow property in this context was first established in [JKO98].

The special feature of our approach is that it is mainly based on stochastic analysis. In fact, most of Theorem 1.2, which is a deterministic result, will be a consequence of a stronger trajectorial result, Theorem 2.1. This constitutes the second main result of this work, it is stated and proved in Chapter 2. Trajectorial approaches of a similar kind were initiated by [FJ16, KST22]. Several important consequences and ramifications of Theorem 2.1 are developed in Section 2.2.

In Chapter 3 we prove the gradient flow property as stated in Theorem 1.2. Most of the proof will be a consequence of Theorem 2.1. As a final ingredient, we need to compute the derivative of the Wasserstein distance

$$f(u) := \lim_{h \to 0} \frac{W_{2,G}(P^{\Sigma}_{u+h}, P^{\Sigma}_{u})}{|h|}, \tag{1.23}$$

a quantity appearing in (1.14) as well as (1.15). The derivative (1.23) is known as a *metric derivative* in [AGS08, Lis09]. By relying on the metric theory of [AGS08, Lis09] we show in Section 3.1 that

$$f(u) = \sqrt{\mathbb{E}_{\mathbb{P}}\Big[\big|(\Sigma\, G\, \nabla_G \log \ell^{\Sigma}_u)(X^{\Sigma}_u)\big|^2_G\Big]}.$$

Finally, in Section 3.2 we merge our findings and provide the proof of Theorem 1.2.

1.8. Literature review

In the seminal paper [JKO98], Jordan, Kinderlehrer, and Otto established a new relationship between a particular class of Fokker–Planck equations, the associated free energy functionals, and the quadratic Wasserstein distance. More precisely, they considered the flow of probability densities described by a stochastic differential equation of Langevin–Smoluchowski type as in (1.21), for which the drift term is given by the gradient of a potential and the diffusion coefficient is constant. The authors of [JKO98] constructed a time discrete, iterative variational scheme, whose solutions converge to the solution of the corresponding Fokker–Planck equation (1.22). Associated with this Fokker–Planck equation is a free energy functional, modeled as the sum of internal and potential energy as in (1.9), and defined on the set of probability densities with finite second moments. The path-breaking result shown in [JKO98] is that the slope of this free energy functional along the Fokker–Planck probability density flow is the steepest possible, if one chooses the quadratic Wasserstein distance (1.20) as metric. In other words, this flow of probability densities can be regarded as a "curve of steepest descent", or a "gradient flux", for the free energy functional with respect to the Wasserstein distance.

The point of view taken in [JKO98] was further developed in the paper [Ott01], where Otto showed the Wasserstein gradient flow structure of the porous medium equation and introduced a formal Riemannian calculus on the space of probability density functions,

which was later dubbed "Otto calculus" in Chapter 15 of [Vil09]; see also Chapter 8 of [Vil03]. As just one example, the results by Otto and Villani on transport inequalities in [OV00] show the power of this heuristic method. For rigorous constructions we refer to the research papers [CG03, CMV06, Lot08, LV09, Oht09, Stu06a, Stu06b] and the textbook [AGS08] by Ambrosio, Gigli, and Savaré.

For the general and rigorous definition of gradient flows in metric spaces we again refer to the book [AGS08], which is based on the theory of minimizing movements and the theory of curves of maximal slope in metric spaces, building on previous work like [DG93, DGMT80, DMT85]. For literature on optimal transport and gradient flows we also refer to [Vil03, Vil09, AG13, San15].

The stochastic analysis approach of this work stems from [FJ16] and [KST22]. Such trajectorial approaches, including applications of time-reversal, were also carried forward in [KT22, KMS21, TY23, KY22, CL22, JK23]. For the connection of gradient flows with large deviations we mention [ADPZ13, Fat16]; further recent related work includes [BCGL20, HRŠS21]. Entropy dissipation for stochastic differential equations in a Riemannian context was also studied in the articles [FL23a, FL23b, BFL24].

2. Dynamics of the relative entropy process

In this chapter we present and prove our second main result, Theorem 2.1. In contrast to Theorem 1.2 it is not of a deterministic but of a trajectorial nature. Most of Theorem 1.2 will then follow from Theorem 2.1 by taking expectations.

Our eventual goal is to study the time evolution of the relative entropy function (1.6). We will do this in a trajectorial fashion by modelling the relative entropy function as a stochastic process. This idea goes back to [FJ16] and was utilized in the context of gradient flows in [KST22]. In the latter work a stochastic analysis approach to the characterization of diffusions of Langevin–Smoluchowski type (1.21) as entropic gradient flows in Wasserstein space was developed. The main difference to our current work is that [KST22] is formulated in the setting of [JKO98], where the diffusion coefficient is constant and the underlying metric is Euclidean. Along the lines of [Lis09], we allow here for a variable diffusion coefficient as well as a Riemannian geometry (determined by $A \in \mathcal{S}$ as in Section 1.3). As a consequence, while in [KST22] the drift coefficient ∇V is controlled (or, put another way, perturbed), we vary the diffusion coefficient Σ in the class of admissible volatilities $\mathcal{S}$. In particular, we provide probabilistic interpretations of the corresponding results in [Lis09] by considering the associated diffusion process (1.16) rather than the underlying Fokker–Planck equation (1.18). From this perspective, our work can be viewed as a generalization of [KST22], in a similar way as [Lis09] generalizes the setting of [JKO98] and some of the results in [AGS08]. Clearly, by choosing the underlying metric to be Euclidean (i.e., by fixing the identity matrix $A = I_n \in \mathcal{S}$ as a reference), we recover the results of [JKO98, KST22]; recall Corollary 1.3 in this context. We also remark that the setting of [Lis09] allows for more general nonlinear partial differential equations, as opposed to the linear case (1.18).

We fix $\Sigma \in \mathcal{S}$ and consider the diffusion process $(X_t^\Sigma)_{0\leqslant t\leqslant T}$, governed by the stochastic differential equation (1.1), with time-marginal distributions $P_t^\Sigma = \mathrm{Law}(X_t^\Sigma)$. With the above motivation in mind, we consider the likelihood ratio function (1.8) and its logarithm as stochastic processes. The resulting random objects are the *likelihood ratio process* (or *Radon–Nikodým derivative process*)

$$\frac{\mathrm{d}P_t^\Sigma}{\mathrm{dQ}}(X_t^\Sigma) = \ell_t^\Sigma(X_t^\Sigma) = \frac{p_t^\Sigma(X_t^\Sigma)}{q(X_t^\Sigma)} = p_t^\Sigma(X_t^\Sigma)\,\mathrm{e}^{V(X_t^\Sigma)}, \qquad 0 \leqslant t \leqslant T \tag{2.1}$$

and the *relative entropy process*

$$\log \ell_t^\Sigma(X_t^\Sigma) = \log\left(\frac{p_t^\Sigma(X_t^\Sigma)}{q(X_t^\Sigma)}\right) = \log p_t^\Sigma(X_t^\Sigma) + V(X_t^\Sigma), \qquad 0 \leqslant t \leqslant T. \tag{2.2}$$

Since taking the $\mathbb{P}$-expectation in (2.2) leads to the relative entropy as in (1.7), the idea is to compute the dynamics of the relative entropy process. More precisely, we will

determine the stochastic differential of the process (2.2) under the probability measure $\mathbb{P}$. Motivated from results like [DPP89, Pav89], and most notably the work [FJ16] by Fontbona and Jourdain, as well as the stochastic approach of [KST22], we expect that these dynamics will be most transparent when regarded in the backward direction of time. Thus we consider the time-reversed process

$$\overline{X}_s^\Sigma := X_{T-s}^\Sigma, \qquad 0 \leqslant s \leqslant T. \tag{2.3}$$

Due to the strong solvability of the stochastic differential equation (1.1), we can assume that the filtration $\mathbb{F} = (\mathcal{F}_t)_{0\leqslant t\leqslant T}$ is given by

$$\mathcal{F}_t = \sigma(X_u^\Sigma, B_u \colon\ 0 \leqslant u \leqslant t)$$

modulo $\mathbb{P}$-augmentation. We define the *backwards filtration* $\overline{\mathbb{F}} = (\overline{\mathcal{F}}_s)_{0\leqslant s\leqslant T}$ by

$$\overline{\mathcal{F}}_s := \sigma(\overline{X}_u^\Sigma, B_{T-u} - B_{T-s} \colon\ 0 \leqslant u \leqslant s)$$

modulo $\mathbb{P}$-augmentation.

The following Theorem 2.1 describes the semimartingale dynamics of the time-reversed relative entropy process

$$\log \overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma) := \log \ell_{T-s}^\Sigma(X_{T-s}^\Sigma), \qquad 0 \leqslant s \leqslant T \tag{2.4}$$

in terms of the *cumulative relative Fisher information process*

$$\overline{F}_s^\Sigma := \int_0^s \Big\langle \nabla \log \overline{\ell}_u^\Sigma \,,\, \Sigma\, \nabla \log \overline{\ell}_u^\Sigma \Big\rangle (\overline{X}_u^\Sigma)\, \mathrm{d}u, \tag{2.5}$$

for $0 \leqslant s \leqslant T$. In particular, we will see that the difference between the processes of (2.4) and (2.5) is a martingale. The proof of the theorem below is given in Section 2.1.

Theorem 2.1. *Let $\Sigma \in \mathcal{S}$, fix $A \in \mathcal{S}$ and set $G = A^{-1}$. On the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$ we have*

$$\mathbb{E}_\mathbb{P}\big[\, \overline{F}_T^\Sigma \,\big] = \int_0^T \mathbb{E}_\mathbb{P}\Big[\Big\langle \nabla \log \ell_t^\Sigma \,,\, \Sigma\, \nabla \log \ell_t^\Sigma \Big\rangle (X_t^\Sigma) \Big]\, \mathrm{d}t < \infty \tag{2.6}$$

and the process

$$\overline{M}_s^\Sigma := \big(\log \overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma) - \log \ell_T^\Sigma(X_T^\Sigma) \big) - \overline{F}_s^\Sigma, \qquad 0 \leqslant s \leqslant T \tag{2.7}$$

is a martingale bounded in $L^2(\mathbb{P})$, with representation as a stochastic integral

$$\overline{M}_s^\Sigma = \int_0^s \Big\langle \nabla \log \overline{\ell}_u^\Sigma(\overline{X}_u^\Sigma) \,,\, \sqrt{2\Sigma(\overline{X}_u^\Sigma)}\, \mathrm{d}\overline{B}_u \Big\rangle, \tag{2.8}$$

where the process $(\overline{B}_s)_{0\leqslant s\leqslant T}$ is $\overline{\mathbb{F}}$-Brownian motion (see Lemma 2.2).

We now consider the important special case $\Sigma = A$. In this situation, the cumulative relative Fisher information process (2.5) takes the form

$$\overline{F}_s^A = \int_0^s \Big\langle \nabla \log \overline{\ell}_u^A \,,\, A\,\nabla \log \overline{\ell}_u^A \Big\rangle (\overline{X}_u^A)\,\mathrm{d}u = \int_0^s |\nabla_G \log \overline{\ell}_u^A(\overline{X}_u^A)|_G^2\,\mathrm{d}u$$

and in light of (1.11) we have

$$\mathbb{E}_{\mathbb{P}}\big[\,\overline{F}_T^A\,\big] = \int_0^T \mathbb{E}_{\mathbb{P}}\Big[|\nabla_G \log \ell_t^A(X_t^A)|_G^2\Big]\,\mathrm{d}t = \int_0^T I_G(P_t^A \,|\, \mathrm{Q})\,\mathrm{d}t < \infty.$$

Again, the process

$$\overline{M}_s^A = \big(\log \overline{\ell}_s^A(\overline{X}_s^A) - \log \ell_T^A(X_T^A)\big) - \overline{F}_s^A, \qquad 0 \leqslant s \leqslant T$$

is a martingale bounded in $L^2(\mathbb{P})$, with representation as a stochastic integral

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

2.1. The proof of Theorem 2.1

We place ourselves on the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$. As already pointed out, our goal is to calculate the dynamics of the time-reversed relative entropy process (2.4). In order to do this, the first question is how the forward dynamics (1.1) of the diffusion process $(X_t^\Sigma)_{0\leqslant t\leqslant T}$ change when regarded in the backward direction of time. In other words, we need to know the stochastic differential of the time-reversed process (2.3). This requires the classical theory of time-reversal; see, e.g., [Föl85, Föl86], [HP86], [Mey94], [Nel01], and [Par86]. We summarize the results relevant for us in the following lemma. For its proof we refer to Theorems 4.1 and 5.1 of Part V.

Lemma 2.2. *The stochastic process $(\overline{B}_s)_{0\leqslant s\leqslant T}$ given by*

$$\overline{B}_s := B_{T-s} - B_T - \int_0^s \mathrm{div}\,\Big(\overline{p}_u^\Sigma(\overline{X}_u^\Sigma)\,\sqrt{2\Sigma(\overline{X}_u^\Sigma)}\,\Big)\big(\overline{p}_u^\Sigma(\overline{X}_u^\Sigma)\big)^{-1}\,\mathrm{d}u$$

is a Brownian motion on the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$, and the time-reversed process (2.3) satisfies the stochastic differential equation

$$\mathrm{d}\overline{X}_s^\Sigma = \big(\,\mathrm{div}\,\Sigma - \Sigma\nabla V + 2\Sigma\,\nabla \log \overline{\ell}_s^\Sigma\big)(\overline{X}_s^\Sigma)\,\mathrm{d}s + \sqrt{2\Sigma(\overline{X}_s^\Sigma)}\,\mathrm{d}\overline{B}_s, \tag{2.9}$$

for $0 \leqslant s \leqslant T$.

Here, the bar over a deterministic function means that time is reversed, i.e., the time parameter $s \in [0,T]$ is replaced by $T - s$. In the lemma above, this conventions means that $\overline{p}_u^\Sigma = p_{T-u}^\Sigma$ and $\overline{\ell}_s^\Sigma = \ell_{T-s}^\Sigma$.

The following lemma describes the partial differential equation satisfied by the likelihood ratio function (1.8). It will be crucial in order to compute the stochastic differential of the time-reversed likelihood ratio process

$$\overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma) = \ell_{T-s}^\Sigma(X_{T-s}^\Sigma), \qquad 0 \leqslant s \leqslant T \tag{2.10}$$

and of its logarithm (2.4).

Lemma 2.3. *The likelihood ratio function (1.8) satisfies the partial differential equation*

$$\partial_t \ell_t^\Sigma = \sum_{i,j=1}^n \Sigma_{ij}\, \partial_{ij}^2 \ell_t^\Sigma + \langle\, \mathrm{div}\, \Sigma - \Sigma\, \nabla V\,, \nabla \ell_t^\Sigma \rangle. \tag{2.11}$$

Reversing time, we derive from (2.11) the equation

$$-\partial_s \overline{\ell}_s^\Sigma = \sum_{i,j=1}^n \Sigma_{ij}\, \partial_{ij}^2 \overline{\ell}_s^\Sigma + \langle\, \mathrm{div}\, \Sigma - \Sigma\, \nabla V\,, \nabla \overline{\ell}_s^\Sigma \rangle. \tag{2.12}$$

This is exactly the *Kolmogorov backward equation* corresponding to the stochastic differential equation (1.1), a circumstance indicating that the dynamics of the likelihood ratio process (2.1) are most descriptive under time reversal.

Proof of Lemma 2.3. Differentiating with respect to the temporal and spatial variables the product

$$p_t^\Sigma(x) = q(x)\, \ell_t^\Sigma(x) = \mathrm{e}^{-V(x)}\, \ell_t^\Sigma(x),$$

substituting these derivatives into the Fokker–Planck equation (1.2), and using the stationary version (1.3) of this equation, one obtains the partial differential equation (2.11) for the likelihood ratio function. □

Lemma 2.4. *On the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$, the time-reversed likelihood ratio process (2.10) satisfies the stochastic differential equation*

$$\frac{\mathrm{d}\overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma)}{\overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma)} = \Big\langle \nabla \log \overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma)\,, \sqrt{2\Sigma(\overline{X}_s^\Sigma)}\, \mathrm{d}\overline{B}_s \Big\rangle + \Big\langle \nabla \log \overline{\ell}_s^\Sigma\,, 2\Sigma\, \nabla \log \overline{\ell}_s^\Sigma \Big\rangle (\overline{X}_s^\Sigma)\, \mathrm{d}s, \tag{2.13}$$

for $0 \leqslant s \leqslant T$.

Proof. Applying Itô's formula to the process (2.10) and using the backward dynamics (2.9) in conjunction with the Kolmogorov backward equation (2.12), we obtain the stochastic differential equation (2.13). □

Lemma 2.5. *On the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$, the time-reversed relative entropy process (2.4) satisfies the stochastic differential equation*

$$\mathrm{d} \log \overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma) = \Big\langle \nabla \log \overline{\ell}_s^\Sigma(\overline{X}_s^\Sigma)\,, \sqrt{2\Sigma(\overline{X}_s^\Sigma)}\, \mathrm{d}\overline{B}_s \Big\rangle + \Big\langle \nabla \log \overline{\ell}_s^\Sigma\,, \Sigma\, \nabla \log \overline{\ell}_s^\Sigma \Big\rangle (\overline{X}_s^\Sigma)\, \mathrm{d}s, \tag{2.14}$$

for $0 \leqslant s \leqslant T$.

Proof. We just have to apply Itô's formula to the process (2.4) and use the dynamics (2.13). □

With Lemma 2.5 we have prepared the main ingredient for the proof of Theorem 2.1. In order to prove the integrability condition (2.6) we follow a similar strategy as in the proof of Theorem 4.1 in Part I.

Proof of Theorem 2.1. The fact that the process $\overline{M}^\Sigma = (\overline{M}^\Sigma_s)_{0\leqslant s\leqslant T}$ of (2.7) is a local martingale with stochastic integral representation (2.8) follows from the stochastic differential (2.14) established in Lemma 2.5. It remains to show the integrability condition (2.6), which in particular implies that $\overline{M}^\Sigma$ is a martingale bounded in $L^2(\mathbb{P})$.

To this end, let us consider the inverse of the process (2.10), i.e.,

$$(\overline{\ell}^\Sigma_s)^{-1}(\overline{X}^\Sigma_s) = \frac{q(\overline{X}^\Sigma_s)}{\overline{p}^\Sigma_s(\overline{X}^\Sigma_s)}, \qquad 0 \leqslant s \leqslant T. \tag{2.15}$$

Applying Itô's formula and using (2.13), we obtain the stochastic differential

$$\mathrm{d}(\overline{\ell}^\Sigma_s)^{-1}(\overline{X}^\Sigma_s) = -(\overline{\ell}^\Sigma_s)^{-1}(\overline{X}^\Sigma_s)\left\langle \nabla \log \overline{\ell}^\Sigma_s(\overline{X}^\Sigma_s)\,,\sqrt{2\Sigma(\overline{X}^\Sigma_s)}\,\mathrm{d}\overline{B}_s\right\rangle \tag{2.16}$$

and conclude that the process (2.15) is a local martingale. From (2.16), or directly from (2.14), we obtain the dynamics of the logarithm of the process (2.15), namely

$$\mathrm{d}\log(\overline{\ell}^\Sigma_s)^{-1}(\overline{X}^\Sigma_s) = -\left\langle \nabla \log \overline{\ell}^\Sigma_s(\overline{X}^\Sigma_s)\,,\sqrt{2\Sigma(\overline{X}^\Sigma_s)}\,\mathrm{d}\overline{B}_s\right\rangle - \left\langle \nabla \log \overline{\ell}^\Sigma_s\,,\,\Sigma\,\nabla \log \overline{\ell}^\Sigma_s\right\rangle(\overline{X}^\Sigma_s)\,\mathrm{d}s.$$

At this point, let us recall the relative entropy (1.7) and point (vii) of Definition 1.1, which ensures that the initial relative entropy $H(P^\Sigma_0\,|\,\mathrm{Q})$ is finite. This implies that the terminal value $\log(\overline{\ell}^\Sigma_T)^{-1}(\overline{X}^\Sigma_T)$ is integrable, i.e.,

$$\mathbb{E}_{\mathbb{P}}\big[\log(\overline{\ell}^\Sigma_T)^{-1}(\overline{X}^\Sigma_T)\big] = -H(P^\Sigma_0\,|\,\mathrm{Q}) \in (-\infty,\infty).$$

On the other hand, for the initial value we have that

$$\mathbb{E}_{\mathbb{P}}\big[\log(\overline{\ell}^\Sigma_0)^{-1}(\overline{X}^\Sigma_0)\big] = -H(P^\Sigma_T\,|\,\mathrm{Q}) \in [-\infty,\infty). \tag{2.17}$$

Thus we can apply Proposition 6.4 of Part I to the local martingale (2.15) and the deterministic stopping time $\tau = T$. We conclude that the expectation in (2.17) is finite and we have the identity

$$\mathbb{E}_{\mathbb{P}}\big[\log(\overline{\ell}^\Sigma_T)^{-1}(\overline{X}^\Sigma_T)\big] - \mathbb{E}_{\mathbb{P}}\big[\log(\overline{\ell}^\Sigma_0)^{-1}(\overline{X}^\Sigma_0)\big] = -\mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \left\langle \nabla \log \overline{\ell}^\Sigma_u\,,\,\Sigma\,\nabla \log \overline{\ell}^\Sigma_u\right\rangle(\overline{X}^\Sigma_u)\,\mathrm{d}u\bigg].$$

This implies that the local martingale $\overline{M}^\Sigma$ is bounded in $L^2(\mathbb{P})$. In fact, by the Itô isometry, the $L^2(\mathbb{P})$-norm of the stochastic integral

$$\overline{M}^\Sigma_T = \int_0^T \left\langle \nabla \log \overline{\ell}^\Sigma_u(\overline{X}^\Sigma_u)\,,\sqrt{2\Sigma(\overline{X}^\Sigma_u)}\,\mathrm{d}\overline{B}_u\right\rangle$$

is finite and can be computed as

$$\tfrac{1}{2}\|\overline{M}^\Sigma_T\|^2_{L^2(\mathbb{P})} = \mathbb{E}_{\mathbb{P}}\bigg[\int_0^T \left\langle \nabla \log \ell^\Sigma_t\,,\,\Sigma\,\nabla \log \ell^\Sigma_t\right\rangle(X^\Sigma_t)\,\mathrm{d}t\bigg] = H(P^\Sigma_0\,|\,\mathrm{Q}) - H(P^\Sigma_T\,|\,\mathrm{Q}).$$

Recalling (2.5), this shows the integrability condition (2.6) and finishes the proof of Theorem 2.1. □

2.2. Consequences of Theorem 2.1

In this section we state several important consequences of our basic trajectorial result, Theorem 2.1.

Corollary 2.6. *Let $\Sigma \in \mathcal{S}$, fix $A \in \mathcal{S}$ and set $G = A^{-1}$. For all $0 \leqslant t_0, t \leqslant T$ we have the relative entropy identity*

$$H(P^\Sigma_{t_0} \,|\, \mathrm{Q}) - H(P^\Sigma_t \,|\, \mathrm{Q}) = \int_{t_0}^t \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \nabla \log \ell^\Sigma_u \,,\, \Sigma\, \nabla \log \ell^\Sigma_u \Big\rangle (X^\Sigma_u)\Big]\, \mathrm{d}u. \tag{2.18}$$

Furthermore, for Lebesgue-a.e. $0 \leqslant t \leqslant T$, the dissipation of relative entropy is equal to

$$\frac{\mathrm{d}}{\mathrm{d}t} H(P^\Sigma_t \,|\, \mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \nabla \log \ell^\Sigma_t \,,\, \Sigma\, \nabla \log \ell^\Sigma_t \Big\rangle (X^\Sigma_t)\Big]. \tag{2.19}$$

Proof. The identity (2.18) follows directly from Theorem 2.1 by taking expectations and using that the process $\overline{M}^\Sigma$ of (2.7) is a martingale. Applying the Lebesgue differentiation theorem yields (2.19). □

As a consequence of (2.18), we observe that the relative entropy function (recall (1.6))

$$[0, T] \ni t \longmapsto H(P^\Sigma_t \,|\, \mathrm{Q})$$

is nonincreasing. As the initial relative entropy $H(P^\Sigma_0 \,|\, \mathrm{Q})$ is finite by Definition 1.1, (vii), this implies that the relative entropy $H(P^\Sigma_t \,|\, \mathrm{Q})$ is finite, for all times $0 \leqslant t \leqslant T$.

Choosing $\Sigma = A$ and recalling the relative Fisher information of (1.11), we immediately obtain from Corollary 2.6 the following result.

Corollary 2.7. *Let $A \in \mathcal{S}$ and set $G = A^{-1}$. For all $0 \leqslant t_0, t \leqslant T$ we have the relative entropy identity*

$$\begin{aligned} H(P^A_{t_0} \,|\, \mathrm{Q}) - H(P^A_t \,|\, \mathrm{Q}) &= \int_{t_0}^t \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \nabla \log \ell^A_u \,,\, A\, \nabla \log \ell^A_u \Big\rangle (X^A_u)\Big]\, \mathrm{d}u \\ &= \int_{t_0}^t I_G(P^A_u \,|\, \mathrm{Q})\, \mathrm{d}u. \end{aligned} \tag{2.20}$$

Furthermore, for Lebesgue-a.e. $0 \leqslant t \leqslant T$, the dissipation of relative entropy is equal to

$$\frac{\mathrm{d}}{\mathrm{d}t} H(P^A_t \,|\, \mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\Big[\Big\langle \nabla \log \ell^A_t \,,\, A\, \nabla \log \ell^A_t \Big\rangle (X^A_t)\Big] = I_G(P^A_t \,|\, \mathrm{Q}). \tag{2.21}$$

The *entropy dissipation* (2.21) is also known as an *entropy production equality*. Various authors have studied this identity in different settings, let us refer to [BÉ85, OV00, AMTU01, CMV03, Vil03, AGS08, Lis09, KST22]. The relation

$$\frac{\mathrm{d}}{\mathrm{d}t} H(P^A_t \,|\, \mathrm{Q}) = I_G(P^A_t \,|\, \mathrm{Q})$$

between entropy dissipation and Fisher information also takes the familiar form of a *de Bruijn type identity* from information theory (see, e.g. [CT06]).

We now return to the probabilistic setting and place ourselves on the filtered probability space $(\Omega, \overline{\mathcal{F}}_T, \overline{\mathbb{F}}, \mathbb{P})$ as in Theorem 2.1. The following result provides the trajectorial analogue of (2.19).

Corollary 2.8. *Let $\Sigma \in \mathcal{S}$, fix $A \in \mathcal{S}$ and set $G = A^{-1}$. For Lebesgue-a.e. $0 \leqslant t \leqslant T$, the trajectorial rate of relative entropy dissipation is equal to*

$$\lim_{s \uparrow T-t} \frac{\log \overline{\ell}_s^{\Sigma}(\overline{X}_s^{\Sigma}) - \mathbb{E}_{\mathbb{P}}\big[\log \ell_t^{\Sigma}(X_t^{\Sigma}) \,|\, \overline{\mathcal{F}}_s\big]}{T-t-s} = -\Big\langle \nabla \log \ell_t^{\Sigma} \,,\, \Sigma\, \nabla \log \ell_t^{\Sigma} \Big\rangle (X_t^{\Sigma}), \tag{2.22}$$

$$\lim_{s \downarrow T-t} \frac{\log \ell_t^{\Sigma}(X_t^{\Sigma}) - \mathbb{E}_{\mathbb{P}}\big[\log \overline{\ell}_s^{\Sigma}(\overline{X}_s^{\Sigma}) \,|\, \overline{\mathcal{F}}_{T-t}\big]}{s-(T-t)} = -\Big\langle \nabla \log \ell_t^{\Sigma} \,,\, \Sigma\, \nabla \log \ell_t^{\Sigma} \Big\rangle (X_t^{\Sigma}), \tag{2.23}$$

where both limits exist in $L^1(\mathbb{P})$.

Proof. It follows from Theorem 2.1 that the numerator of the fraction on the left-hand side of (2.22) is equal to

$$-\mathbb{E}_{\mathbb{P}}\Big[\overline{F}_{T-t}^{\Sigma} - \overline{F}_s^{\Sigma} \,|\, \overline{\mathcal{F}}_s\Big] = -\mathbb{E}_{\mathbb{P}}\bigg[\int_s^{T-t} \Big\langle \nabla \log \overline{\ell}_u^{\Sigma} \,,\, \Sigma\, \nabla \log \overline{\ell}_u^{\Sigma} \Big\rangle (\overline{X}_u^{\Sigma})\, \mathrm{d}u \;\Big|\; \overline{\mathcal{F}}_s\bigg]. \tag{2.24}$$

After dividing by $T-t-s$, the $L^1(\mathbb{P})$-convergence to the right-hand side of (2.22), as $s \uparrow T-t$, follows from the Lebesgue differentiation theorem and its generalization to conditional expectations (see, e.g., Proposition 6.3 in Section 6.3 of Part I). The proof of (2.23) follows a similar line of reasoning. □

3. Proof of the gradient flow property

In this chapter we prove the gradient flow property as formulated in Theorem 1.2. In a nutshell, we show that the flow of probability measures $(P_t^A)_{0\leqslant t\leqslant T}$, defined by the stochastic differential equation (1.16), is the gradient flow of the relative entropy functional (1.17) with respect to the quadratic Wasserstein distance $W_{2,G}$ as defined in (1.12).

3.1. Derivative of the Wasserstein distance

We define the vector field

$$[0,T]\times\mathbb{R}^n \longrightarrow \mathbb{R}^n\colon (t,x)\longmapsto \boldsymbol{v}_t^\Sigma(x) \coloneqq -\Sigma(x)\left(\nabla \log p_t^\Sigma(x) + \nabla V(x)\right).$$

Then we can write the Fokker–Planck equation (1.2) as a *continuity equation*, to wit

$$\partial_t p_t^\Sigma(x) + \operatorname{div}\left(\boldsymbol{v}_t^\Sigma(x)\, p_t^\Sigma(x)\right) = 0, \qquad (t,x)\in(0,T)\times\mathbb{R}^n.$$

The derivative of the Wasserstein distance along an absolutely continuous "curve of probability measures" satisfying a continuity equation is a well known result; see, e.g. [AGS08, Lis09, San15].

Lemma 3.1. *Let $\Sigma\in\mathcal{S}$, fix $A\in\mathcal{S}$ and set $G=A^{-1}$. For Lebesgue-a.e. $t\in[0,T]$, we have that*

$$\lim_{h\to 0}\frac{W_{2,G}(P_{t+h}^\Sigma, P_t^\Sigma)}{|h|} = \|\boldsymbol{v}_t^\Sigma\|_{L^2_G(P_t^\Sigma)} = \sqrt{\mathbb{E}_{\mathbb{P}}\left[\left|(\Sigma\, G\,\nabla_G \log \ell_t^\Sigma)(X_t^\Sigma)\right|_G^2\right]}. \tag{3.1}$$

Proof. For a proof of this result we refer to [Lis09, Theorem 2.4], which is an extension of [AGS08, Theorem 8.3.1] to the current Riemannian setting induced by G. Alternatively, the reader may consult [San15, Theorem 5.14] and adapt the arguments by means of [McC01], so that they apply also in our Riemannian setting. □

In the case $\Sigma = A$, the derivative (3.1) takes the form

$$\lim_{h\to 0}\frac{W_{2,G}(P_{t+h}^A, P_t^A)}{|h|} = \|\boldsymbol{v}_t^A\|_{L^2_G(P_t^A)} = \sqrt{\mathbb{E}_{\mathbb{P}}\left[|\nabla_G \log \ell_t^A(X_t^A)|_G^2\right]} = \sqrt{I_G(P_t^A\,|\,\mathrm{Q})}; \tag{3.2}$$

for the last equality in (3.2), recall the definition of relative Fisher information (1.11).

3.2. The proof of Theorem 1.2

Proof of Theorem 1.2. Let $\Sigma \in \mathcal{S}$, fix $A \in \mathcal{S}$ and set $G = A^{-1}$. Let $0 \leqslant t_0 \leqslant t \leqslant T$. We write the relative entropy identity (2.18) of Corollary 2.6 as

$$H(P_{t_0}^{\Sigma} \,|\, \mathrm{Q}) - H(P_t^{\Sigma} \,|\, \mathrm{Q}) = \int_{t_0}^{t} \mathbb{E}_{\mathbb{P}} \Big[\Big\langle \nabla_G \log \ell_u^{\Sigma} \,,\, \Sigma\, G\, \nabla_G \log \ell_u^{\Sigma} \Big\rangle_G (X_u^{\Sigma}) \Big] \, \mathrm{d}u.$$

Applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} &\big| H(P_t^{\Sigma} \,|\, \mathrm{Q}) - H(P_{t_0}^{\Sigma} \,|\, \mathrm{Q}) \big| \leqslant \\ &\quad \leqslant \int_{t_0}^{t} \sqrt{\mathbb{E}_{\mathbb{P}} \Big[\big| \nabla_G \log \ell_u^{\Sigma}(X_u^{\Sigma}) \big|_G^2 \Big]} \, \sqrt{\mathbb{E}_{\mathbb{P}} \Big[\big| (\Sigma\, G\, \nabla_G \log \ell_u^{\Sigma})(X_u^{\Sigma}) \big|_G^2 \Big]} \, \mathrm{d}u. \end{aligned}$$

Recalling the definition of relative Fisher information (1.11) and the derivative of the Wasserstein distance (3.1), this proves the inequality (1.14). Clearly, if Σ is a constant multiple of A, this inequality is an equality.

Now let us consider the case $\Sigma = A$. Recalling the relative entropy identity (2.20) of Corollary 2.7 we have

$$H(P_{t_0}^{A} \,|\, \mathrm{Q}) - H(P_t^{A} \,|\, \mathrm{Q}) = \int_{t_0}^{t} I_G(P_u^{A} \,|\, \mathrm{Q}) \, \mathrm{d}u.$$

According to (3.2), we have the derivative of the Wasserstein distance

$$\lim_{h \to 0} \frac{W_{2,G}(P_{t+h}^{A}, P_t^{A})}{|h|} = \sqrt{I_G(P_t^{A} \,|\, \mathrm{Q})},$$

so that we can write

$$I_G(P_u^{A} \,|\, \mathrm{Q}) = \tfrac{1}{2} I_G(P_u^{A} \,|\, \mathrm{Q}) + \tfrac{1}{2} \bigg(\lim_{h \to 0} \frac{W_{2,G}(P_{u+h}^{A}, P_u^{A})}{|h|} \bigg)^2.$$

We conclude the equality (1.15). □

Part V.

Time reversal of diffusions

Abstract

We review the theory of time reversal for diffusion processes developed by Föllmer [Föl85, Föl86], Haussmann and Pardoux [HP86], and Pardoux [Par86]. This part can be read independently of the rest of the thesis; it does not present novel results.

1. Introduction

It is very well known that the Markov property is invariant under time reversal. In other words, a Markov process remains a Markov process under time reversal (e.g., [RW00a, Exercise E60.41, p. 162]). On the other hand, it is also well known that the strong Markov property is not necessarily preserved under time reversal (e.g., [RW00a, p. 330]), and neither is the semimartingale property (e.g., [Wal82]). The reason for such failure is the same in both cases: after reversing time, "we may know too much". Thus, the following questions arise rather naturally:

Given a diffusion process (in particular, a strong Markov semimartingale with continuous paths) $X = (X(t))_{0 \leqslant t \leqslant T}$ with certain specific drift and dispersion characteristics, under what conditions might the time-reversed process

$$\widehat{X}(s) := X(T - s), \qquad 0 \leqslant s \leqslant T, \tag{1.1}$$

also be a diffusion? If it happens to be, what are the characteristics of the time-reversed diffusion?

Such questions go back at least to Boltzmann [Bol96, Bol98a, Bol98b], Schrödinger [Sch31, Sch32] and Kolmogorov [Kol37]; they were dealt with systematically by Nelson [Nel01] (see also Carlen [Car84]) in the context of Nelson's dynamical theories for Brownian motion and diffusion. There is now a rather complete theory that answers these questions and provides, as a kind of "bonus", some rather unexpected results as well. It was developed in the context of theories of filtering, interpolation and extrapolation, where such issues arise naturally — most notably Haussmann and Pardoux [HP86], and Pardoux [Par86]. Very interesting related results in a non-Markovian context, but with dispersion structure given by the identity matrix, have been obtained by Föllmer [Föl85, Föl86]. Let us refer also to the papers [Nag64, NM79] dealing with time reversal of Markov processes, and to the book [Nag93] on diffusion theory. In what follows, this theory is presented in the spirit of the expository paper by Meyer [Mey94].

2. The setting

We place ourselves on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = (\mathcal{F}(t))_{0 \leqslant t \leqslant T}$ rich enough to support an $\mathbb{R}^d$-valued Brownian motion $W = (W_1, \ldots, W_d)'$ adapted to $\mathbb{F}$, as well as an independent $\mathcal{F}(0)$-measurable random vector $\xi = (\xi_1, \ldots, \xi_n)' \colon \Omega \to \mathbb{R}^n$. In fact, we shall assume that $\mathbb{F}$ is the filtration generated by these two objects, in the sense that we

shall take

$$\mathcal{F}(t) = \sigma\big(\xi, W(\theta)\colon\ 0 \leqslant \theta \leqslant t\big), \qquad 0 \leqslant t \leqslant T,$$

modulo $\mathbb{P}$-augmentation. Next, we assume that the system of stochastic equations

$$X_i(t) = \xi_i + \int_0^t b_i\big(\theta, X(\theta)\big)\,\mathrm{d}\theta + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}\big(\theta, X(\theta)\big)\,\mathrm{d}W_\nu(\theta), \qquad 0 \leqslant t \leqslant T, \tag{2.1}$$

for $i = 1, \ldots, n$ admits a pathwise unique, strong solution. It is then well known that the resulting continuous process $X = (X_1, \ldots, X_n)'$ is $\mathbb{F}$-adapted (the strong solvability of the equation (2.1)), which implies that we have also

$$\begin{aligned} \mathcal{F}(t) &= \sigma\big(X(\theta), W(\theta)\colon\ 0 \leqslant \theta \leqslant t\big) \\ &= \sigma\big(X(0), W(t) - W(\theta)\colon\ 0 \leqslant \theta \leqslant t\big) \end{aligned} \tag{2.2}$$

modulo $\mathbb{P}$-augmentation, for $0 \leqslant t \leqslant T$; as well as that X has the strong Markov property, and is thus a diffusion process with drifts $b_i(\,\cdot\,,\,\cdot\,)$ and dispersions $\mathrm{s}_{i\nu}(\,\cdot\,,\,\cdot\,)$, for $i = 1, \ldots, n$ and $\nu = 1, \ldots, d$. We shall denote the (i,j)-entry of the covariance matrix $\mathrm{a}(t,x) := \mathrm{s}(t,x)\,\mathrm{s}'(t,x)$ by

$$\mathrm{a}_{ij}(t,x) := \sum_{\nu=1}^{d} \mathrm{s}_{i\nu}(t,x)\,\mathrm{s}_{j\nu}(t,x), \qquad i,j = 1, \ldots, n.$$

The drift and dispersion characteristics are given mappings from $[0,T] \times \mathbb{R}^n$ into $\mathbb{R}$ with sufficient smoothness; in particular, such that the probability density function $p(t,\,\cdot\,)\colon \mathbb{R}^n \to (0,\infty)$ in

$$\mathbb{P}[X(t) \in A] = \int_A p(t,x)\,\mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n),$$

is smooth. Sufficient conditions on the drift $b_i(\,\cdot\,,\,\cdot\,)$ and dispersion $\mathrm{s}_{i\nu}(\,\cdot\,,\,\cdot\,)$ characteristics that lead to such smoothness, are provided by the Hörmander hypoellipticity conditions; see for instance [Bel95], [Nua06] for this result, as well as [Rog85] for a very simple argument in the one-dimensional case ($n = d = 1$), which also applies to the Langevin type equation (1.3) for arbitrary $n \in \mathbb{N}$. We refer to [Fri75], [RW00b] or [KS98] for the basics of the theory of stochastic equations of the form (2.1).

The probability density function $p(t,\,\cdot\,)\colon \mathbb{R}^n \to (0,\infty)$ solves the forward Kolmogorov [Kol31] equation [Fri75, p. 149]

$$\partial_t p(t,x) = \tfrac{1}{2} \sum_{i,j=1}^{n} \partial_{ij}^2\big(\mathrm{a}_{ij}(t,x)\,p(t,x)\big) - \sum_{i=1}^{n} \partial_i\big(b_i(t,x)\,p(t,x)\big), \tag{2.3}$$

for $(t,x) \in (0,T] \times \mathbb{R}^n$. If the drift and dispersion characteristics do not depend on time, and an invariant probability measure exists for the diffusion process of (2.1), the density

function $p(\,\cdot\,)$ of this measure solves the stationary version of this forward Kolmogorov equation, to wit

$$\tfrac{1}{2}\sum_{i,j=1}^{n}\partial_{ij}^{2}\big(\mathrm{a}_{ij}(x)\,p(x)\big)=\sum_{i=1}^{n}\partial_{i}\big(b_{i}(x)\,p(x)\big),\qquad x\in\mathbb{R}^{n}.$$

3. Time reversal and the backwards filtration

Consider now the family of σ-algebras $(\widehat{\mathcal{F}}(t))_{0\leqslant t\leqslant T}$ given by

$$\widehat{\mathcal{F}}(t):=\sigma\big(X(\theta),W(\theta)-W(t)\colon\ t\leqslant\theta\leqslant T\big),\qquad 0\leqslant t\leqslant T. \tag{3.1}$$

It is not hard to see that the σ-algebra in (3.1) is expressed equivalently as

$$\begin{aligned}\widehat{\mathcal{F}}(t)&=\sigma\big(X(t),W(\theta)-W(t)\colon\ t\leqslant\theta\leqslant T\big)=\sigma\big(X(t),W(\theta)-W(T)\colon\ t\leqslant\theta\leqslant T\big)\\&=\sigma\big(X(T),W(\theta)-W(t)\colon\ t\leqslant\theta\leqslant T\big)=\sigma\big(X(T)\big)\vee\mathcal{H}(t).\end{aligned} \tag{3.2}$$

Here, the σ-algebra generated by the Brownian increments after time t, namely,

$$\mathcal{H}(t):=\sigma\big(W(\theta)-W(t)\colon\ t\leqslant\theta\leqslant T\big),\qquad 0\leqslant t\leqslant T,$$

is independent of the random vector $X(t)$. The time-reversed processes $\widehat{X}$ as in (1.1), as well as the process

$$\widetilde{W}(s):=W(T-s)-W(T),\qquad 0\leqslant s\leqslant T, \tag{3.3}$$

are both adapted to the *backwards filtration*

$$\widehat{\mathbb{F}}:=\big(\widehat{\mathcal{F}}(T-s)\big)_{0\leqslant s\leqslant T}, \tag{3.4}$$

where

$$\begin{aligned}\widehat{\mathcal{F}}(T-s)&=\sigma\big(X(T-u),W(T-u)-W(T-s)\colon\ 0\leqslant u\leqslant s\big)\\&=\sigma\big(\widehat{X}(u),\widetilde{W}(u)-\widetilde{W}(s)\colon\ 0\leqslant u\leqslant s\big)\end{aligned} \tag{3.5}$$

from (3.1). Note that, by complete analogy with (2.2), we have also

$$\begin{aligned}\widehat{\mathcal{F}}(T-s)&=\sigma\big(X(T),W(T-u)-W(T-s)\colon\ 0\leqslant u\leqslant s\big)\\&=\sigma\big(\widehat{X}(0)\big)\vee\mathcal{H}(T-s)\end{aligned} \tag{3.6}$$

on account of (3.2), where

$$\begin{aligned}\mathcal{H}(T-s)&=\sigma\big(W(T-u)-W(T-s)\colon\ 0\leqslant u\leqslant s\big)\\&=\sigma\big(\widetilde{W}(u)-\widetilde{W}(s)\colon\ 0\leqslant u\leqslant s\big).\end{aligned} \tag{3.7}$$

In words: the σ-algebra $\widehat{\mathcal{F}}(T-s)$ is generated by the terminal value $X(T)$ of the forward process (i.e., by the original value $\widehat{X}(0)$ of the backward process) and by the increments of the time-reversed process $\widetilde{W}$ on the interval $[0,s]$; see the expressions right above. Furthermore, the σ-algebra $\widehat{\mathcal{F}}(T-s)$ measures all the random variables $\widehat{X}(u)$, $0 \leqslant u \leqslant s$.

In fact, the time-reversed process $\widetilde{W}$ *is a Brownian motion of the backwards filtration*

$$\mathbb{H} := \big(\mathcal{H}(T-s)\big)_{0 \leqslant s \leqslant T} \subseteq \widehat{\mathbb{F}}$$

as in (3.7), generated by the increments of W after time $T-s$, $0 \leqslant s \leqslant T$. This is because it is a martingale of this filtration, has continuous paths, and its quadratic variation is that of Brownian motion (Lévy's theorem [KS98, Theorem 5.1]). In Section 4 we shall see that the process $\widetilde{W}$ *is only a semimartingale of the larger backwards filtration*

$$\widehat{\mathbb{F}} = \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T} \supseteq \mathbb{H},$$

and identify its semimartingale decomposition.

4. Some remarkable Brownian motions

Following the exposition and ideas in [Mey94], we start with a couple of observations. First, for every $0 \leqslant t \leqslant T$ and every integrable, $\widehat{\mathcal{F}}(t)$-measurable random variable $\mathcal{K}$, we have

$$\mathbb{E}\big[\mathcal{K} \,|\, \mathcal{F}(t)\big] = \mathbb{E}\big[\mathcal{K} \,|\, X(t)\big], \qquad \text{almost surely.} \tag{4.1}$$

Secondly, we fix a function $G \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ and a time-point $t \in (0,T]$, and define

$$g(\theta, x) := \mathbb{E}\big[G(X(t)) \,|\, X(\theta) = x\big], \qquad (\theta, x) \in [0,t] \times \mathbb{R}^n.$$

Invoking the Markov property of X, we deduce that the process

$$g(\theta, X(\theta)) = \mathbb{E}\big[G(X(t)) \,|\, X(\theta)\big] = \mathbb{E}\big[G(X(t)) \,|\, \mathcal{F}(\theta)\big], \qquad 0 \leqslant \theta \leqslant t$$

is an $\mathbb{F}$-martingale, and obtain

$$\begin{aligned} G(X(t)) - g(\theta, X(\theta)) &= g(t, X(t)) - g(\theta, X(\theta)) \\ &= \sum_{i=1}^{n} \sum_{\nu=1}^{d} \int_\theta^t \partial_i g(v, X(v))\, \mathrm{s}_{i\nu}(v, X(v))\, \mathrm{d}W_\nu(v). \end{aligned}$$

For every index $\nu = 1, \dots, d$ this gives, after integrating by parts,

$$\begin{aligned}
\mathbb{E}\big[(W_\nu(t) - W_\nu(\theta)) \cdot G(X(t))\big] &= \mathbb{E}\Big[(W_\nu(t) - W_\nu(\theta)) \cdot \Big(g(t, X(t)) - g(\theta, X(\theta))\Big)\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{n} \int_\theta^t \partial_i g(v, X(v))\, \mathrm{s}_{i\nu}(v, X(v))\, \mathrm{d}v\Big] \\
&= \sum_{i=1}^{n} \int_\theta^t \int_{\mathbb{R}^n} (\partial_i g \cdot \mathrm{s}_{i\nu})(v, x)\, p(v, x)\, \mathrm{d}x\, \mathrm{d}v \\
&= -\sum_{i=1}^{n} \int_\theta^t \int_{\mathbb{R}^n} g(v, x)\, \partial_i(p(v, x)\, \mathrm{s}_{i\nu}(v, x))\, \mathrm{d}x\, \mathrm{d}v \\
&= -\int_\theta^t \int_{\mathbb{R}^n} g(v, x)\ \mathrm{div}\,(p(v, x)\, \bar{\mathrm{s}}_\nu(v, x))\, \mathrm{d}x\, \mathrm{d}v \\
&= -\int_\theta^t \mathbb{E}\Big[g(v, X(v)) \cdot \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p}(v, X(v))\Big]\, \mathrm{d}v \\
&= -\mathbb{E}\Big[G(X(t)) \cdot \int_\theta^t \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v\Big].
\end{aligned}$$

Here $\bar{\mathrm{s}}_\nu(v, \cdot\,)$ is the νth column vector of the dispersion matrix. Comparing the first and last expressions in the above string of equalities, we see that with $0 \leqslant \theta \leqslant t$ we have

$$\mathbb{E}\Big[G(X(t)) \cdot \Big(W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v\Big)\Big] = 0 \tag{4.2}$$

for every $G \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and thus by extension for every bounded, measurable function $G\colon \mathbb{R}^n \to \mathbb{R}$.

Theorem 4.1. *The vector process $B = (B_1, \dots, B_d)'$ defined as*

$$B_\nu(s) := \widetilde{W}_\nu(s) - \int_0^s \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p}(T - u, \widehat{X}(u))\, \mathrm{d}u \tag{4.3}$$

$$= W_\nu(T - s) - W_\nu(T) - \int_{T-s}^{T} \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v, \qquad 0 \leqslant s \leqslant T, \tag{4.4}$$

for $\nu = 1, \dots, d$, is a Brownian motion of the backwards filtration $\widehat{\mathbb{F}}$ as in (3.4).

The Brownian motion process B is thus independent of $\widehat{\mathcal{F}}(T)$, and therefore also of the $\widehat{\mathcal{F}}(T)$-measurable random variable $X(T)$. A bit more generally,

$$\{B(T - \theta) - B(T - t)\colon\ 0 \leqslant \theta \leqslant t\}$$

is independent of

$$\widehat{\mathcal{F}}(t) \supseteq \sigma(X(v)\colon\ t \leqslant v \leqslant T).$$

Note also from (4.4) that

$$B_\nu(T-\theta) - B_\nu(T-t) = W_\nu(\theta) - W_\nu(t) - \int_\theta^t \frac{\operatorname{div}(p\,\bar{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v,$$

for $0 \leqslant \theta \leqslant t$. Reversing time once again, we obtain the following corollary of Theorem 4.1.

Corollary 4.2. *The $\mathbb{F}$-adapted vector process $V = (V_1, \dots, V_d)'$ with components*

$$\begin{aligned} V_\nu(t) &:= B_\nu(T-t) - B_\nu(T) \\ &= W_\nu(t) + \int_0^t \frac{\operatorname{div}(p\,\bar{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v, \qquad 0 \leqslant t \leqslant T, \end{aligned}$$

for $\nu = 1, \dots, d$, is yet another Brownian motion (with respect to its own filtration $\mathbb{F}^V \subseteq \mathbb{F}$). This process is independent of the random variable $X(T)$; and a bit more generally, for every $t \in (0, T]$, the σ-algebra

$$\mathcal{F}^V(t) := \sigma\big(V(\theta)\colon\ 0 \leqslant \theta \leqslant t\big)$$

generated by present-and-past values of V, is independent of $\sigma(X(v)\colon\ t \leqslant v \leqslant T)$, the σ-algebra generated by present-and-future values of X.

Proof of Theorem 4.1. It suffices to show that each component process B_ν is a martingale of the backwards filtration $\widehat{\mathbb{F}}$ of (3.4); because then, in view of the continuity of paths and the easily checked property $\langle B_\nu, B_\ell\rangle(s) = s\,\delta_{\nu\ell}$, we can deduce that each B_ν is a Brownian motion in the backwards filtration $\widehat{\mathbb{F}}$ (and of course also in its own filtration), and that B_ν, B_ℓ are independent for $\ell \neq \nu$, by appealing to Lévy's theorem once again.

Now we have to show

$$\mathbb{E}\big[(B_\nu(T-\theta) - B_\nu(T-t)) \cdot \mathcal{K}\big] = 0, \qquad 0 \leqslant \theta \leqslant t \leqslant T,$$

for every bounded, $\widehat{\mathcal{F}}(t)$-measurable random variable $\mathcal{K}$; equivalently,

$$\mathbb{E}\left[\mathbb{E}[\mathcal{K} \,|\, \mathcal{F}(t)] \cdot \left(W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\operatorname{div}(p\,\bar{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v\right)\right] = 0,$$

as the expression

$$W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\operatorname{div}(p\,\bar{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v$$

is $\mathcal{F}(t)$-measurable. But recalling (4.1) we have that

$$\mathbb{E}[\mathcal{K} \,|\, \mathcal{F}(t)] = \mathbb{E}[\mathcal{K} \,|\, X(t)], = G(X(t))$$

for some bounded, measurable $G\colon \mathbb{R}^n \to \mathbb{R}$, and the desired result follows from (4.2). □

5. The diffusion property under time reversal

Let us return now to the question, whether the time-reversed process $\widehat{X}$ of (1.1), (2.1) is a diffusion. We start by expressing X_i of (2.1) in terms of a backwards Itô integral (see Section 6) as

$$X_i(t) - \xi_i - \int_0^t b_i(\theta, X(\theta))\,\mathrm{d}\theta = \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta))\,\mathrm{d}W_\nu(\theta)$$

$$= \sum_{\nu=1}^{d} \Big(\int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta) - \langle \mathrm{s}_{i\nu}(\,\cdot\,, X), W_\nu \rangle(t) \Big).$$

From (2.1), we have by Itô's formula that the process

$$\mathrm{s}_{i\nu}(\,\cdot\,, X) - \mathrm{s}_{i\nu}(0, \xi) - \sum_{j=1}^{n} \sum_{\kappa=1}^{d} \int_0^{\cdot} \partial_j \mathrm{s}_{i\nu}(\theta, X(\theta)) \cdot \mathrm{s}_{j\kappa}(\theta, X(\theta))\,\mathrm{d}W_\kappa(\theta)$$

is of finite first variation, therefore

$$\langle \mathrm{s}_{i\nu}(\,\cdot\,, X), W_\nu \rangle(t) = \sum_{j=1}^{n} \int_0^t \mathrm{s}_{j\nu}(\theta, X(\theta))\,\partial_j \mathrm{s}_{i\nu}(\theta, X(\theta))\,\mathrm{d}\theta.$$

We conclude

$$X_i(t) = \xi_i - \int_0^t \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\,\partial_j \mathrm{s}_{i\nu} - b_i \Big)(\theta, X(\theta))\,\mathrm{d}\theta + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta).$$

Evaluating also at $t = T$, then subtracting, we obtain

$$X_i(t) = X_i(T) + \int_t^T \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\,\partial_j \mathrm{s}_{i\nu} - b_i \Big)(\theta, X(\theta))\,\mathrm{d}\theta - \sum_{\nu=1}^{d} \int_t^T \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta),$$

as well as

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\,\partial_j \mathrm{s}_{i\nu} - b_i \Big)(T - u, \widehat{X}(u))\,\mathrm{d}u$$

$$+ \sum_{\nu=1}^{d} \int_0^s \mathrm{s}_{i\nu}(T - u, \widehat{X}(u))\,\mathrm{d}\widetilde{W}_\nu(u)$$

by reversing time. It is important here to note that the backward Itô integral for W becomes a forward Itô integral for the process $\widetilde{W}$, the time reversal of W in the manner of (3.3).

But now let us recall (4.3), on the strength of which the above expression takes the form

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \sum_{\nu=1}^{d} \int_0^s \mathrm{s}_{i\nu}(T-u, \widehat{X}(u))\, \mathrm{d}B_\nu(u)$$
$$+ \int_0^s \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\, \partial_j \mathrm{s}_{i\nu} + \sum_{\nu=1}^{d} \mathrm{s}_{i\nu}\, \frac{\mathrm{div}(p\, \bar{\mathrm{s}}_\nu)}{p} - b_i \Big)(T-u, \widehat{X}(u))\, \mathrm{d}u,$$

for $0 \leqslant s \leqslant T$. But in conjunction with Theorem 4.1, this means that the time-reversed process $\widehat{X}$ of (1.1), (2.1) is a semimartingale of the backwards filtration $\widehat{\mathbb{F}}$ as in (3.4), with decomposition

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \widehat{b}_i(T-u, \widehat{X}(u))\, \mathrm{d}u + \sum_{\nu=1}^{d} \int_0^s \mathrm{s}_{i\nu}(T-u, \widehat{X}(u))\, \mathrm{d}B_\nu(u) \tag{5.1}$$

for $0 \leqslant s \leqslant T$, where, for each $i = 1, \dots, n$, the function $\widehat{b}_i(\,\cdot\,,\,\cdot\,)$ is specified by

$$\widehat{b}_i(t,x) + b_i(t,x) = \sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}(t,x)\, \partial_j \mathrm{s}_{i\nu}(t,x) + \sum_{\nu=1}^{d} \mathrm{s}_{i\nu}(t,x)\, \frac{\mathrm{div}\,(p(t,x)\, \bar{\mathrm{s}}_\nu(t,x))}{p(t,x)}$$
$$= \sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}(t,x)\, \partial_j \mathrm{s}_{i\nu}(t,x) + \sum_{\nu=1}^{d} \frac{\mathrm{s}_{i\nu}(t,x)}{p(t,x)} \Big(\sum_{j=1}^{n} \partial_j \big(p(t,x)\, \mathrm{s}_{j\nu}(t,x)\big) \Big)$$
$$= \sum_{j=1}^{n} \big(\partial_j \mathrm{a}_{ij}(t,x) + \mathrm{a}_{ij}(t,x) \cdot \partial_j \log p(t,x)\big).$$

Theorem 5.1. *Under the assumptions of Section 2, the time-reversed process $\widehat{X}$ of (1.1), (2.1) is a diffusion with respect to the backwards filtration $\widehat{\mathbb{F}}$ of (3.4), with characteristics as in (5.1), namely, dispersions $\mathrm{s}_{i\nu}(T-s,x)$ and drifts $\widehat{b}_i(T-s,x)$ given by the* generalized Nelson equation

$$\widehat{b}_i(t,x) + b_i(t,x) = \sum_{j=1}^{n} \Big(\partial_j \mathrm{a}_{ij}(t,x) + \mathrm{a}_{ij}(t,x) \cdot \partial_j \log p(t,x)\Big), \qquad i = 1, \dots, n.$$

Equivalently, and with $\mathrm{div}\,(\mathrm{a}(t,x)) := \big(\sum_{j=1}^{n} \partial_j \mathrm{a}_{ij}(t,x)\big)_{1 \leqslant i \leqslant n}$, we write

$$\widehat{b}(t,x) + b(t,x) = \mathrm{div}(\mathrm{a}(t,x)) + \mathrm{a}(t,x) \cdot \nabla \log p(t,x).$$

Theorem 5.1 can be extended to the more general case where the sums of the distributional derivatives

$$\sum_{j=1}^{n} \partial_j\big(\mathrm{a}_{ij}(t,x)\, p(t,x)\big), \qquad i = 1, \ldots, n,$$

are only assumed to be locally integrable functions of $x \in \mathbb{R}^n$; see [MNS89, RVW01].

Remark 5.2 (Some filtration comparisons). For an invertible dispersion matrix $\mathrm{s}(\,\cdot\,,\,\cdot\,)$, it follows from (5.1) that the Brownian motion B is adapted to the filtration generated by $\widehat{X}$; that is,

$$\begin{aligned} \mathcal{F}^B(s) \subseteq \mathcal{F}^{\widehat{X}}(s) &:= \sigma\big(\widehat{X}(u)\colon\ 0 \leqslant u \leqslant s\big) \\ &= \sigma\big(X(T-u)\colon\ 0 \leqslant u \leqslant s\big), \qquad 0 \leqslant s \leqslant T. \end{aligned} \tag{5.2}$$

Now recall (4.3); in its light, the filtration comparison in (5.2) implies the inclusion $\mathcal{F}^{\widetilde{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s)$, thus from (3.7) we derive

$$\mathcal{H}(T-s) \subseteq \mathcal{F}^{\widetilde{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s), \qquad 0 \leqslant s \leqslant T,$$

and from (3.6) also $\widehat{\mathcal{F}}(T-s) \subseteq \mathcal{F}^{\widehat{X}}(s)$.

But we have also the reverse inclusion $\mathcal{F}^{\widehat{X}}(s) \subseteq \widehat{\mathcal{F}}(T-s)$ on account of (3.5) and (5.2); therefore, $\mathcal{F}^{\widehat{X}}(s) = \widehat{\mathcal{F}}(T-s)$ holds for all $0 \leqslant s \leqslant T$ when $s(\,\cdot\,,\,\cdot\,)$ is invertible. These considerations inform our choice of backwards filtration

$$\mathcal{G}(T-s) = \mathcal{F}^{\widehat{X}}(s), \qquad 0 \leqslant s \leqslant T$$

in (4.2) of Part I. ◇

6. The backwards Itô integral

For two continuous semimartingales

$$X = X(0) + M + B \qquad \text{and} \qquad Y = Y(0) + N + C,$$

with M, N being continuous local martingales and B, C being continuous adapted processes of finite variation, let us recall the definition of the Fisk–Stratonovich integral in [KS98, Definition 3.3.13, p. 156], as well as its properties in [KS98, Problem 3.3.14] and [KS98, Problem 3.3.15].

By analogy with this definition, we introduce the *backwards Itô integral*

$$\int_0^{\cdot} Y(t) \bullet \mathrm{d}X(t) := \int_0^{\cdot} Y(t)\, \mathrm{d}M(t) + \int_0^{\cdot} Y(t)\, \mathrm{d}B(t) + \langle M, N\rangle(\,\cdot\,),$$

where the first (respectively, the second) integral on the right-hand side is to be interpreted in the Itô (respectively, the Lebesgue–Stieltjes) sense, and the process $\langle M, N \rangle$ is the cross-variation of M and N.

If $\Pi = \{t_0, t_1, \ldots, t_m\}$ is a partition of the interval $[0, T]$ with

$$0 = t_0 < t_1 < \ldots < t_m = T,$$

then the sums

$$\sum_{j=0}^{m-1} Y(t_{j+1})(X(t_{j+1}) - X(t_j))$$

converge in probability to $\int_0^T Y(t) \bullet \mathrm{d}X(t)$ as the mesh $\|\Pi\|$ of the partition tends to zero. Note that the increments of X here "stick backwards into the past", as opposed to "sticking forward into the future" as in the Itô integral.

For the backwards Itô integral we have the change of variable formula

$$f(X(\,\cdot\,)) = f(X(0)) + \sum_{i=1}^{n} \int_0^{\cdot} \partial_i f(X(t)) \bullet \mathrm{d}X_i(t) - \tfrac{1}{2} \sum_{i,j=1}^{n} \int_0^{\cdot} \partial_{ij}^2 f(X(t))\, \mathrm{d}\langle M_i, M_j \rangle(t),$$

where now $X = (X_1, \ldots, X_n)'$ is a vector of continuous semimartingales $X_1, \ldots, X_n$ of the form $X_i = X_i(0) + M_i + B_i$ as above, for $i = 1, \ldots, n$. Note the change of sign, from $(+)$ to $(-)$ in the last, stochastic correction term.

Part VI.

The structure of martingale Benamou–Brenier in $\mathbb{R}^d$

Joint work with Julio Backhoff-Veraguas, Mathias Beiglböck, and Walter Schachermayer

Abstract

In classical optimal transport, the contributions of Benamou–Brenier and McCann regarding the time-dependent version of the problem are cornerstones of the field and form the basis for a variety of applications in other mathematical areas.

Stretched Brownian motion provides an analogue for the martingale version of this problem. In this work we provide a characterization in terms of gradients of convex functions, similar to the characterization of optimizers in the classical transport problem for quadratic distance cost.

1. Introduction

Optimal transport as a field in mathematics goes back to Monge [Mon81] and Kantorovich [Kan42], who established its modern formulation. The seminal results of Benamou, Brenier, and McCann [Bre87, Bre91, McC94, BB99] form the basis of the modern theory, with striking applications in a variety of different areas, see the monographs [Vil03, Vil09, AG13, San15].

We are interested in transport problems where the transport plan satisfies an additional martingale constraint. This additional requirement arises naturally in finance, but is of independent mathematical interest. For example there are notable consequences for the study of martingale inequalities (e.g. [BN15, OST15, HLOST16]) and the Skorokhod embedding problem (e.g. [BCH17, KTT17, BNS22]). Early articles on this topic of *martingale optimal transport* include [HN12, BHLP13, TT13, DS14, GHLT14, CLM17].

In view of the central role taken by the results of Benamou–Brenier [BB99] and McCann [McC01] on optimal transport for the squared (Euclidean) distance, the related continuous-time transport problem and McCann's displacement interpolation, it is paramount to search for similar concepts in the martingale context. This is the main motivation of the present work. Before describing our results, we briefly recap the classical role models.

1.1. Benamou–Brenier transport problem and McCann interpolation in probabilistic terms

Given probabilities μ, ν in the space $\mathscr{P}_2(\mathbb{R}^d)$ of d-dimensional distributions with finite second moment, Brenier's theorem asserts that, if μ is absolutely continuous, the following are equivalent for a coupling $\pi \in \mathsf{Cpl}(\mu, \nu)$, i.e., a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν:

(1) π minimizes the transportation costs with respect to the squared distance between μ and ν.

(2) π is concentrated on the graph of a function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ which is *monotone* in the sense that $f = \nabla v$ for a convex function $v\colon \mathbb{R}^d \to (-\infty, +\infty]$.

Brenier's theorem gives a structural description of the optimal transport plan for the most widely used cost function. Another essential consequence of the result is that it provides a particularly natural way of moving probabilities: it implies that μ can be transported to ν via the gradient of a convex function. In fact, in many applications it is the mere existence of this monotone transport map that is required, irrespective of its optimality properties.

Ideally, a martingale counterpart of Brenier's theorem should mimic these aspects. That is, starting from a natural optimization problem our goal is to define a natural martingale that connects the probabilities μ and ν.

While [BJ16, HLT16, GKL19] have proposed martingale versions of Brenier's monotone transport map based on static transport problems, our starting point is a continuous formulation in the spirit of the Benamou–Brenier continuous-time transport problem given in [HT19, BBHK20].

We first recapitulate the continuous-time formulation of Brenier's theorem in probabilistic language. For $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, consider

$$T_2(\mu,\nu) \coloneqq \inf_{\substack{X_0\sim\mu,\, X_1\sim\nu,\\ X_t = X_0 + \int_0^t b_s\,ds}} \mathbb{E}\Big[\int_0^1 |b_t|^2\,dt\Big]. \tag{BB}$$

If μ is absolutely continuous with respect to Lebesgue measure, (BB) has a unique (in law) optimizer and the following are equivalent for processes $X = (X_t)_{0\leqslant t\leqslant 1}$ with $X_0 \sim \mu$ and $X_1 \sim \nu$:

(1) X solves (BB).

(2) $X_1 = f(X_0)$, where f is the gradient of a convex function $v\colon \mathbb{R}^d \to (-\infty, +\infty]$ and all particles move with constant speed, i.e., $X_t = X_0 + bt$, $t \in [0,1]$ for the random variable $b = X_1 - X_0$.

1.2. Martingale optimization problem

Assume that $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ are in convex order, in signs $\mu \preceq_{\mathrm{c}} \nu$, that is $\int \phi\,d\mu \leqslant \int \phi\,d\nu$ for all convex functions $\phi\colon \mathbb{R}^d \to \mathbb{R}$ with linear growth. We consider the optimization problem

$$MT(\mu,\nu) \coloneqq \inf_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s\,dB_s}} \mathbb{E}\Big[\int_0^1 |\sigma_t - I_d|_{\mathrm{HS}}^2\,dt\Big], \tag{MBB}$$

where B is Brownian motion on $\mathbb{R}^d$ and $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm.

The main result of [BBHK20] is that (MBB) admits a unique optimizer M^*. Moreover, M^* is a strong Markov martingale and has properties of time consistency analogous to the classical deterministic case. While (BB) entails that particles move along straight lines, the functional in (MBB) stipulates that M maximizes the correlation with Brownian motion subject to the given marginal constraints. It is also shown in [BBHK20] that M^* is the process whose evolution follows the movement of a Brownian particle as closely as possible with respect to an *adapted Wasserstein distance* (see e.g. [BBBE20, Föl22]) subject to the marginal conditions $M_0 \sim \mu$ and $M_1 \sim \nu$. These properties motivate to call the martingale M^* *stretched Brownian motion* between μ and ν as in [BBHK20].

1.3. The Bass martingale

Our main result relates the optimality property in the definition of stretched Brownian motion to a structural description. As motivation we recall a classical construction of Bass [Bas83] which provides a particularly appealing martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ that terminates in a fixed measure ν on the real line and starts at the barycenter of ν. This amounts to a solution of the Skorokhod embedding problem (modulo time change). Let $(B_t)_{0\leqslant t\leqslant 1}$ be Brownian motion (started in $B_0 = 0$), let $\gamma := \mathrm{Law}(B_1)$ denote the standard Gaussian and define $f\colon \mathbb{R} \to \mathbb{R}$ as the γ-a.e. unique increasing mapping that pushes γ to ν. Bass then defines the martingale

$$M_t := \mathbb{E}[f(B_1) \,|\, \sigma(B_s\colon s \leqslant t)] = \mathbb{E}[f(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1, \tag{1.1}$$

so that M_0 starts at the barycenter of ν and $\mathrm{Law}(M_1) = \nu$.

As we are interested in martingales in multiple dimensions and with possibly non-degenerate starting law μ, we consider the construction of Bass in the following generality:

Definition 1.1. Let $B = (B_t)_{0\leqslant t\leqslant 1}$ be Brownian motion on $\mathbb{R}^d$, where $B_0 \sim \alpha \in \mathscr{P}(\mathbb{R}^d)$, and let $v\colon \mathbb{R}^d \to \mathbb{R}$ be convex such that $\nabla v(B_1)$ is square-integrable. Then we call

$$M_t := \mathbb{E}[\nabla v(B_1) \,|\, \sigma(B_s\colon s \leqslant t)] = \mathbb{E}[\nabla v(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1$$

a *Bass martingale* from $\mu := \mathrm{Law}(M_0)$ to $\nu := \mathrm{Law}(M_1)$.

It turns out that a Bass martingale from μ to ν is a stretched Brownian motion, but not vice versa. While a stretched Brownian motion from μ to ν exists uniquely, the following question arises: under which conditions on $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ is there a Bass martingale M from μ to ν, i.e. satisfying $M_0 \sim \mu$ and $M_1 \sim \nu$, for μ and ν prescribed in advance?

1.4. Structure of stretched Brownian motion

We need a connectivity assumption on the marginals, known under the name of *irreducibility* in the martingale transport literature.

Definition 1.2. For $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ we say that the pair (μ, ν) is *irreducible* if for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$ there is a martingale $X = (X_t)_{0\leqslant t\leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

With this definition in hand we can announce our first main result:

Theorem 1.3. *Let $\mu \preceq_{\mathrm{c}} \nu$ be probabilities on $\mathbb{R}^d$ with finite second moments and suppose that (μ, ν) is irreducible. Then the following are equivalent for a martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ with $M_0 \sim \mu$ and $M_1 \sim \nu$:*

(1) M is stretched Brownian motion, i.e., the optimizer of (MBB).

(2) M is a Bass martingale.

We briefly provide some context for the irreducibility assumption in Theorem 1.3. In classical optimal transport, the product coupling of the marginals guarantees that mass can be transported from an arbitrary starting position to an arbitrary target position. However, this is not necessarily true in the case of martingale transport where it may happen that $\mathbb{R}^d$ decomposes into disjoint (convex) regions that do not communicate with each other. The irreducibility assumption excludes this as it guarantees that for any sets A and B that are charged by μ and ν, respectively, there exists a martingale X connecting them (in the sense that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$). For $d > 1$, the appearance of more than one irreducible component leads to intricate phenomena, analyzed in the remarkable contributions [OS17, DMT19, GKL19]. This problem is revisited in the follow-up paper [ST24] in terms of Bass martingales.

In Theorem 1.4 below we give equivalent characterizations of irreducibility for a pair (μ, ν), as introduced in Definition 1.2. In particular, it is equivalent to consider continuous- or discrete-time martingales in Definition 1.2.

Theorem 1.4. *Let μ, ν in $\mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then the following are equivalent.*

(1) The pair (μ, ν) is De March–Touzi irreducible in the sense of Definition 7.2.

(2) There exists $\pi \in \mathsf{MT}(\mu, \nu)$ such that $\pi_x \sim \nu$, for μ-a.e. $x \in \mathbb{R}^d$.

(3) The pair (μ, ν) is irreducible in the sense of Definition 1.2, i.e., for all Borel sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$ there is a martingale $(X_t)_{0 \leqslant t \leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

(4) For all Borel sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$ there exists $\pi \in \mathsf{MT}(\mu, \nu)$ such that $\pi(A \times B) > 0$.

(5) For all compact sets $A \subseteq \mathbb{R}^d$ and open halfspaces H with $\mu(A), \nu(H) > 0$ there exists $\pi \in \mathsf{MT}(\mu, \nu)$ such that $\pi(A \times H) > 0$.

Proof. For the proof of the implication "(1) $\Rightarrow$ (2)" we can take $\pi = \pi^{\mathrm{SBM}}$ in (2) by Corollary 7.7. We turn to the proof of the implication "(2) $\Rightarrow$ (3)": Clearly condition (2) implies (4). Thus, in order to show (3), it suffices to construct a continuous-time martingale $(X_t)_{0 \leqslant t \leqslant 1}$ with $\mathrm{Law}(X_0, X_1) = \pi$. This can be achieved as in the proof of [BBHK20, Theorem 2.2]. For the proof of "(3) $\Rightarrow$ (4)" we can take $\pi = \mathrm{Law}(X_0, X_1)$ in (4). The implication "(4) $\Rightarrow$ (5)" is trivial. It remains to show that "(5) $\Rightarrow$ (1)": We fix an open halfspace H satisfying $\nu(H) > 0$ and set

$$m \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \Big\{ \mu\big(\{x \in \mathbb{R}^d \colon \pi_x(H) > 0\}\big) \Big\}.$$

Considering countable convex combinations of elements of $\mathsf{MT}(\mu, \nu)$ it follows that the supremum is attained by some $\bar{\pi} \in \mathsf{MT}(\mu, \nu)$. If the set

$$\{x \in \mathbb{R}^d \colon \bar{\pi}_x(H) = 0\}$$

had positive μ-measure, it would contain a compact set A with positive μ-measure in contradiction to condition (5). Hence $m = 1$.

Observe that there is a countable family of open halfspaces $\{H_n\}_{n\geqslant 1}$ with $\nu(H_n) > 0$ such that, for every open halfspace H with $\nu(H) > 0$, there is some $n \geqslant 1$ satisfying $H_n \subseteq H$.

Next, for each $n \geqslant 1$, we pick $\pi^{(n)} \in \mathsf{MT}(\mu,\nu)$ such that $\pi_x^{(n)}(H_n) > 0$, for μ-a.e. $x \in \mathbb{R}^d$. Set $\hat{\pi} \coloneqq \sum_{n\geqslant 1} 2^{-n}\pi^{(n)} \in \mathsf{MT}(\mu,\nu)$. Then by Lemma 1.5 below it follows that $\widehat{\operatorname{supp}}(\nu) \subseteq \widehat{\operatorname{supp}}(\hat{\pi}_x)$, for μ-a.e. $x \in \mathbb{R}^d$. □

Lemma 1.5. *Let $\rho \in \mathscr{P}(\mathbb{R}^d)$. Then $y \in \widehat{\operatorname{supp}}(\rho)$ if and only if $\rho(H) > 0$ for every open halfspace H such that $y \in H$.*

Proof. This is a simple consequence of Hahn–Banach. □

Remark 1.6. We note that irreducibility is not only a sufficient assumption for Theorem 1.3, but it is in fact also necessary. Indeed, as in the proof of Corollary 7.7 below one sees that the Bass martingale does the job of connecting any two sets which are charged by μ and ν, implying the irreducibility of (μ,ν) by Theorem 1.4. ◇

An important consequence of Theorem 1.3 is that for any irreducible pair $\mu \preceq_{\mathrm{c}} \nu$ there exists a unique Bass martingale

$$M_t = \mathbb{E}[\nabla v(B_1) \,|\, \sigma(B_s \colon s \leqslant t)] = \mathbb{E}[\nabla v(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1, \tag{1.2}$$

with $M_0 \sim \mu$, $M_1 \sim \nu$ and it is worthwhile to comment on the properties of M. We write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and $v_t \coloneqq v * \gamma^{1-t} \colon \mathbb{R}^d \to \mathbb{R}$ for the convolution of the function v and the measure γ^{1-t}. In these terms, (1.2) amounts to

$$M_t = \nabla v_t(B_t), \qquad 0 \leqslant t \leqslant 1, \tag{1.3}$$

which emphasizes that the Bass martingale is obtained as a monotone transformation of Brownian motion at each time point. Finally, as a Brownian martingale, M is a diffusion which connects μ and ν. Indeed, applying Itô's formula to (1.3) we obtain

$$dM_t = \nabla^2 v_t \circ \nabla v_t^*(M_t)\, dB_t, \qquad 0 \leqslant t \leqslant 1,$$

where v_t^* denotes the convex conjugate of v_t, i.e., $v_t^*(y) \coloneqq \sup_{x\in\mathbb{R}^d}\{\langle x, y\rangle - v_t(x)\}$.

1.5. Dual viewpoint

Our arguments rely on a novel dual viewpoint on stretched Brownian motion. In classical transport minimization of the squared distance is equivalent to maximization of correlation and the latter formulation yields a dual problem which is simpler to interpret. A

similar fact holds true in the present martingale setting: problem (MBB) is equivalent to maximizing the covariance with Brownian motion

$$P(\mu,\nu) \coloneqq \sup_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s\, dB_s}} \mathbb{E}\Big[\int_0^1 \operatorname{tr}(\sigma_t)\, dt\Big], \tag{1.4}$$

in the sense that both problems have the same optimizer and the values are related via $MT(\mu,\nu) = d + \int |y|^2\, d\nu(y) - \int |x|^2\, d\mu(x) - 2P(\mu,\nu)$.

As in the classical case, the advantage of the formulation (1.4) is that its dual problem admits a simpler formulation.

Theorem 1.7. *Assume that $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ are in convex order. The value $P(\mu,\nu)$ of the problem (1.4) is equal to*

$$D(\mu,\nu) \coloneqq \inf_{\substack{\psi\in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi\, d\nu - \int (\psi^* * \gamma)^*\, d\mu \Big). \tag{1.5}$$

The infimum is uniquely attained (modulo adding affine functions) by a lower semicontinuous convex function $\psi_{\mathrm{opt}}\colon \mathbb{R}^d \to (-\infty,+\infty]$ satisfying $\mu(\mathrm{ri}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$ if and only if (μ,ν) is irreducible. In this case the (unique) optimizer to (MBB) is given by the Bass martingale

$$M_t \coloneqq \mathbb{E}[\nabla v(B_1) \,|\, \sigma(B_s\colon s\leqslant t)] = \mathbb{E}[\nabla v(B_1)\,|\,B_t], \qquad 0\leqslant t\leqslant 1,$$

where $v = \psi_{\mathrm{opt}}^*$ and $B_0 \sim \nabla(\psi_{\mathrm{opt}}^* * \gamma)^*(\mu)$.

Here, $\mathrm{ri}(\operatorname{dom}\psi_{\mathrm{opt}})$ denotes the relative interior of the domain of ψ_{opt}, i.e. the set on which ψ_{opt} is finite. The symbol $*$ used as a superscript denotes the convex conjugate of a function, otherwise it is the standard convolution operator.

It turns out that the optimizer $\psi_{\mathrm{opt}}\colon \mathbb{R}^d \to (-\infty,+\infty]$ is not necessarily ν-integrable. Therefore, in order for the difference of the integrals in (1.5) to be well-defined for $\psi = \psi_{\mathrm{opt}}$, attainment of $D(\mu,\nu)$ has to be understood in a "relaxed" sense frequently encountered in martingale transport problems; see [BJ16, BNT17, BNS22] and Propositions 4.1, 4.2 below.

1.6. Literature

The first article that provides structural results for the martingale transport problem in general dimensions is the work [GKL19] of Ghoussoub–Kim–Lim. They obtain descriptions for the minimizers and maximizers for the cost function $c(x,y) = |x-y|$, when marginals are supported on $\mathbb{R}^2$, as well as for marginals on higher-dimensional state spaces that are in subharmonic order. Given a specific martingale X, Ghoussoub–Kim–Lim also define a finest paving of the source space into cells that are invariant under the martingale X.

The contributions of De March–Touzi [DMT19] and Obłój–Siorpaes [OS17] put the theme of irreducible decompositions center stage. In contrast to the work of Ghoussoub–Kim–Lim, their interest lies in pavings that are invariant under *all* martingales which start in μ and terminate in ν. Specifically it is shown in [DMT19] that there exists a unique finest paving with this property. This De March–Touzi paving is revisited in the follow-up paper [ST24], where it is characterized in terms of Bass martingales.

Huesmann–Trevisan [HT19] investigate Benamou–Brenier-type formulations for the martingale transport problem on $\mathbb{R}^d$. In particular, they provide equivalent PDE-formulations and establish existence and duality results. In the context of market impact in finance, Loeper [Loe18] arrives independently to problem (MBB).

For further contributions to the martingale transport in continuous time, we mention [DS14, BHLT17, COT19, GKP19, GLW19, GKL20, CKPS21, GL21] among many others.

Finally, [BBHK20] is a predecessor of the present work in which it is established that the martingale transport problem (MBB) admits a unique solution, solves further related optimization problems and has properties of time consistency in the spirit of classical optimal transport for the squared distance cost function. Furthermore, the counterpart of our main result is established in dimension one (without reference to duality). Among other differences, we highlight a particular property which distinguishes martingale transport in dimension one from the higher-dimensional case: on the real line the convex pavings of Ghoussoub–Kim–Lim [GKL19] and De March–Touzi [DMT19] / Obłój–Siorpaes [OS17] coincide (and their existence is relatively trivial, see [BJ16]). This is no longer true in higher dimensions which renders Theorem 1.3 substantially more intricate in the present multi-dimensional case.

1.7. Preview

In Chapter 2 we introduce some definitions and frequently used notation. Duality results for stretched Brownian motion and the important role of convexity are discussed in Chapter 3. The proof of the first part of Theorem 1.7, namely that there is no duality gap between the primal problem (1.4) and the dual problem (1.5), is given in Chapter 4. In Chapter 5 we outline and prepare the proof of the second part of Theorem 1.7, which gives a necessary and sufficient condition for dual attainment in terms of the irreducibility assumption. To this end, we analyze the connection between the existence of dual optimizers and Bass martingales, which is the content of Chapter 6. In Chapter 7 we show that the irreducibility assumption implies the existence of a dual optimizer. After these preparations we are in a position to prove Theorem 1.3 in Section 7.2 and complete the proof of Theorem 1.7 in Section 7.3.

2. Definitions and Notation

- We write $\mathscr{P}(\mathbb{R}^d)$ for the probability measures on $\mathbb{R}^d$, $\mathscr{P}_p(\mathbb{R}^d)$ for the subset of probability measures satisfying $\int |x|^p \, d\mu < +\infty$, and $\mathscr{P}_p^x(\mathbb{R}^d)$ for the elements of $\mathscr{P}_p(\mathbb{R}^d)$ with barycenter $x \in \mathbb{R}^d$, for $p \in [1, +\infty)$.

- For $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, we denote by $\mathsf{Cpl}(\mu, \nu)$ the set of all couplings $\pi \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ between μ and ν, i.e., probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν.

- We say that $\mu \in \mathscr{P}_1(\mathbb{R}^d)$ is dominated by $\nu \in \mathscr{P}_1(\mathbb{R}^d)$ in convex order and write $\mu \preceq_{\mathrm{c}} \nu$, if for all convex functions $\phi \colon \mathbb{R}^d \to \mathbb{R}$ with linear growth we have $\int \phi \, d\mu \leqslant \int \phi \, d\nu$.

- For $\mu, \nu \in \mathscr{P}_1(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ we define the collection of martingale transports $\mathsf{MT}(\mu, \nu)$ as those couplings $\pi \in \mathsf{Cpl}(\mu, \nu)$ satisfying $\mathrm{bary}(\pi_x) \coloneqq \int y \, \pi_x(dy) = x$, for μ-a.e. $x \in \mathbb{R}^d$. Here, the family of probability measures $\{\pi_x\}_{x \in \mathbb{R}^d} \subseteq \mathscr{P}(\mathbb{R}^d)$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e., $\pi(dx, dy) = \pi_x(dy) \, \mu(dx)$.

- The d-dimensional Gaussian distribution with barycenter $x \in \mathbb{R}^d$ and covariance matrix tI_d is denoted by γ_x^t and we set $\gamma_x \coloneqq \gamma_x^1$ as well as $\gamma \coloneqq \gamma_0$.

- We denote by $C_{\mathrm{q}}(\mathbb{R}^d)$ the set of continuous functions $\psi \colon \mathbb{R}^d \to \mathbb{R}$ with quadratic growth, meaning that there are constants $a, k, \ell \in \mathbb{R}$ with

$$\ell + \tfrac{|\cdot|^2}{2} \leqslant \psi(\,\cdot\,) \leqslant a + k|\cdot|^2.$$

  We also introduce the set

$$C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d) \coloneqq \{\psi(\,\cdot\,) + \mathrm{aff}(\,\cdot\,) \colon \psi \in C_{\mathrm{q}}(\mathbb{R}^d), \mathrm{aff} \colon \mathbb{R}^d \to \mathbb{R} \text{ affine}\}.$$

- For two measures ϱ and ρ we write $\varrho * \rho$ for their convolution. If f is a function, the convolution of f and ρ is defined as

$$(f * \rho)(x) \coloneqq \int f(x - y) \, \rho(dy).$$

  In particular, $(f * \gamma)(x) = \int f(x + y) \, \gamma(dy)$.

- For a function $f \colon \mathbb{R}^d \to (-\infty, +\infty]$, its convex conjugate is given by

$$f^*(y) \coloneqq \sup_{x \in \mathbb{R}^d} \big(\langle x, y \rangle - f(x)\big)$$

and we write

$$\operatorname{dom} f \coloneqq \{x \in \mathbb{R}^d \colon f(x) < +\infty\}$$

for the domain of f. If f is convex and $\operatorname{dom} f \neq \varnothing$, we say that f is a proper convex function. If f is additionally lower semicontinuous, then $f = f^{**}$ by the Fenchel–Moreau theorem, a fact that we will use repeatedly in this work.

- The operators int and ri denote the interior and relative interior of a set, respectively. We write $\overline{A}$ for the closure of a set $A \subseteq \mathbb{R}^d$. The operator conv applied to a function or a set denotes the convex hull.
- The support and the closed convex hull of the support of a measure ρ are denoted by $\operatorname{supp}(\rho)$ and $\widehat{\operatorname{supp}}(\rho)$, respectively.
- The symbol ∂, applied to a convex function, denotes its subdifferential or — by abuse of notation — also a subgradient.

3. Duality for stretched Brownian motion

We fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and consider the *primal problem*

$$\tilde{P}(\mu,\nu) \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx), \tag{3.1}$$

where the *maximal covariance* (MCov) between two probability measures $p_1 \in \mathscr{P}_2(\mathbb{R}^d)$ and $p_2 \in \mathscr{P}_2(\mathbb{R}^d)$ is defined as

$$\mathrm{MCov}(p_1,p_2) \coloneqq \sup_{q \in \mathsf{Cpl}(p_1,p_2)} \int \langle x_1, x_2 \rangle\, q(dx_1, dx_2).$$

Since μ and ν are in convex order, a celebrated result by Strassen [Str65] shows that $\mathsf{MT}(\mu,\nu) \neq \varnothing$. The optimization problem (3.1) is a *weak* optimal transport problem in the sense of [GRST17], since the function $\pi \mapsto \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx)$ is nonlinear, as opposed to classical optimal transport, where linear problems of the form $\pi \mapsto \int c(x,y)\, \pi(dx,dy)$ are studied. The problem (3.1) can also be viewed as a discrete-time version of the continuous-time optimization problem (1.4). In fact, by [BBHK20, Theorem 2.2] the value $\tilde{P}(\mu,\nu)$ of (3.1) is finite and equals $P(\mu,\nu)$, as defined in (1.4). Furthermore, there exists a unique optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ of (3.1) and if $(M_t)_{0 \leqslant t \leqslant 1}$ is the stretched Brownian motion from μ to ν, then the law of (M_0, M_1) equals π^{SBM}. From now on, we will focus on the discrete-time formulation (3.1).

To formulate a *dual problem* to (3.1), we consider the set of continuous functions with quadratic growth

$$C_{\mathrm{q}}(\mathbb{R}^d) \coloneqq \{\psi \colon \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \exists\, a,k,\ell \in \mathbb{R} \text{ with } \ell + \tfrac{|\cdot|^2}{2} \leqslant \psi(\,\cdot\,) \leqslant a + k|\cdot|^2\}.$$

Then we define the dual problem

$$\tilde{D}(\mu,\nu) \coloneqq \inf_{\psi \in C_{\mathrm{q}}(\mathbb{R}^d)} \Big(\int \psi\, d\nu - \int \varphi^{\psi}\, d\mu \Big), \tag{3.2}$$

where the function $\mathbb{R}^d \ni x \mapsto \varphi^{\psi}(x)$ is given by

$$\varphi^{\psi}(x) \coloneqq \inf_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int \psi\, dp - \mathrm{MCov}(p,\gamma) \Big). \tag{3.3}$$

Theorem 3.1. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. There is no duality gap between the primal problem (3.1) and the dual problem (3.2), i.e., $\tilde{P}(\mu,\nu) = \tilde{D}(\mu,\nu)$. Moreover, the primal problem is uniquely attained and has a finite value, i.e., there exists a unique $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ such that*

$$\tilde{P}(\mu,\nu) = \int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, \mu(dx) < +\infty. \tag{3.4}$$

Proof. Existence and uniqueness of the optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ as well as finiteness of $\tilde{P}(\mu,\nu)$ were proved in [BBHK20, Theorem 2.2]. In order to show that there is no duality gap, we apply [BBP19, Theorem 1.3] with the cost function

$$C(x,p) \coloneqq \begin{cases} -\mathrm{MCov}(p,\gamma) + \frac{1}{2}\int |y|^2\,dp, & \text{if } \mathrm{bary}(p) = x, \\ +\infty, & \text{if } \mathrm{bary}(p) \neq x, \end{cases}$$

for $x \in \mathbb{R}^d$ and $p \in \mathscr{P}_2(\mathbb{R}^d)$. This function is bounded from below and convex in the second argument. If we equip $\mathscr{P}_2(\mathbb{R}^d)$ with the topology induced by the quadratic Wasserstein distance one verifies that $C(x,p)$ is also jointly lower semicontinuous with respect to the product topology on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$. We introduce the space of continuous functions which are bounded from below and have at most quadratic growth

$$C_{\mathrm{b,q}}(\mathbb{R}^d) \coloneqq \{\tilde{\psi}\colon \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \exists\, a,k,\ell \in \mathbb{R} \text{ with } \ell \leqslant \tilde{\psi}(\,\cdot\,) \leqslant a + k|\cdot|^2\}.$$

Then by [BBP19, Theorem 1.3] the value $\tilde{P}(\mu,\nu)$ of the primal problem equals

$$\tilde{D}_{\mathrm{b,q}}(\mu,\nu) \coloneqq \inf_{\tilde{\psi}\in C_{\mathrm{b,q}}(\mathbb{R}^d)} \Big(\int (\tilde{\psi}(\,\cdot\,) + \tfrac{|\cdot|^2}{2})\,d\nu - \int \tilde{\varphi}^{\tilde{\psi}}\,d\mu \Big),$$

where

$$\tilde{\varphi}^{\tilde{\psi}}(x) \coloneqq \inf_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int (\tilde{\psi}(\,\cdot\,) + \tfrac{|\cdot|^2}{2})\,dp - \mathrm{MCov}(p,\gamma) \Big).$$

Finally, passing from the functions $\tilde{\psi} \in C_{\mathrm{b,q}}(\mathbb{R}^d)$ to $\psi(\,\cdot\,) \coloneqq \tilde{\psi}(\,\cdot\,) + \frac{|\cdot|^2}{2} \in C_{\mathrm{q}}(\mathbb{R}^d)$, we see that $\tilde{D}_{\mathrm{b,q}}(\mu,\nu) = \tilde{D}(\mu,\nu)$. □

To motivate (3.2) as a plausible dual formulation of (3.1) we provide some heuristics. Considering

$$\chi(\pi) \coloneqq \inf_{\psi\in C_{\mathrm{q}}(\mathbb{R}^d)} \Big(\int \psi\,d\nu - \int \psi\,d\pi_x\,d\mu(x) \Big) = \begin{cases} 0, & \text{if } \int \pi_x\,d\mu(x) = \nu, \\ -\infty, & \text{else,} \end{cases}$$

we *formally* obtain the desired duality relation by interchanging inf and sup:

$$\begin{aligned} \tilde{P}(\mu,\nu) &= \sup_{\substack{\pi(dx,dy)=\pi_x(dy)\,\mu(dx),\\ \pi_x\in\mathscr{P}_2^x(\mathbb{R}^d)}} \Big(\int \mathrm{MCov}(\pi_x,\gamma)\,d\mu(x) + \chi(\pi) \Big) \\ &= \inf_{\psi\in C_{\mathrm{q}}(\mathbb{R}^d)} \Big(\int \sup_{\pi_x\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(\pi_x,\gamma) + \int \psi\,d(\nu-\pi_x) \Big)\,d\mu(x) \Big) = \tilde{D}(\mu,\nu). \end{aligned}$$

In the remainder of this chapter we make the important observation that in the dual problem (3.2) it suffices to optimize over the class of functions $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ which are *convex*. We refer to [GJ20, Theorem 1.1] for an analogous result (compare also [GRST17]). We then show that it is also equivalent to optimize over all convex functions $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ which are only μ-a.s. finite, but not necessarily of quadratic growth.

Given some $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$, we denote by $\operatorname{conv}\psi$ the convex hull of ψ, i.e., the greatest convex function smaller or equal to ψ. It will be convenient to have an explicit representation for the convex hull of a function, as in (3.5) below. This identity is usually stated in the more specific form

$$(\operatorname{conv}\psi)(y) = \inf\Big\{\sum_{i=1}^{d+1}\lambda_i\psi(y_i)\colon \sum_{i=1}^{d+1}\lambda_i y_i = y\Big\},$$

where the infimum is taken over all expressions of y as a convex combination of $d+1$ points, see [Roc70, Corollary 17.1.5].

Lemma 3.2. *Let $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. Then the convex hull $\operatorname{conv}\psi$ satisfies*

$$(\operatorname{conv}\psi)(y) = \inf_{p\in\mathscr{P}_2^y(\mathbb{R}^d)}\int \psi\,dp, \qquad y\in\mathbb{R}^d \tag{3.5}$$

and again $\operatorname{conv}\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$.

Recalling the dual problem (3.2), we define the dual function

$$\mathcal{D}(\psi) \coloneqq \int \psi\,d\nu - \int \varphi^\psi\,d\mu, \tag{3.6}$$

for $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. Now we can prove our crucial observation, that it suffices to optimize the dual function over the class of functions $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ which are *convex*.

Proposition 3.3. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then $\mathcal{D}(\operatorname{conv}\psi) \leqslant \mathcal{D}(\psi)$ for all $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ and consequently*

$$\tilde{D}(\mu,\nu) = \inf_{\psi\in C_{\mathrm{q}}(\mathbb{R}^d)}\mathcal{D}(\psi) = \inf_{\substack{\psi\in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}}\mathcal{D}(\psi). \tag{3.7}$$

Proof. Let $\varepsilon > 0$, $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ and $\{p_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ be a measurable collection of probability measures with $\operatorname{bary}(p_x) = x$. To show the claim, it is sufficient to construct a measurable family $\{\bar{p}_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ with $\operatorname{bary}(\bar{p}_x) = x$ such that

$$\operatorname{MCov}(p_x,\gamma) + \int \operatorname{conv}\psi\,d(\nu - p_x) \leqslant \operatorname{MCov}(\bar{p}_x,\gamma) + \int \psi\,d(\nu-\bar{p}_x) + \varepsilon. \tag{3.8}$$

Let us construct appropriate probability measures $\{\bar{p}_x\}_{x\in\mathbb{R}^d}$. By Lemma 3.2 and a measurable selection argument we can choose a measurable collection of probability measures $\{\tilde{p}_y\}_{y\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ with $\operatorname{bary}(\tilde{p}_y) = y$ such that

$$\int \psi\,d\tilde{p}_y \leqslant (\operatorname{conv}\psi)(y) + \varepsilon. \tag{3.9}$$

Then we define $\bar{p}_x(dz) \coloneqq \int_y \tilde{p}_y(dz)\,p_x(dy)$, so that $\operatorname{bary}(\bar{p}_x) = x$. Integrating (3.9) with respect to $p_x(dy)$ yields

$$\int \psi\,d\bar{p}_x \leqslant \int \operatorname{conv}\psi\,dp_x + \varepsilon. \tag{3.10}$$

Since $\psi, \operatorname{conv}\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ and $p_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ we conclude

$$\ell + \tfrac{1}{2}\int |y|^2\, \bar{p}_x(dy) \leqslant a + k\int |y|^2\, p_x(dy) + \varepsilon < +\infty,$$

so that $\bar{p}_x \in \mathscr{P}_2^x(\mathbb{R}^d)$.

In order to show the inequality (3.8), we first observe that $p_x \preceq_{\mathrm{c}} \bar{p}_x$ by Jensen's inequality. Together with the Kantorovich duality (see, e.g. [Vil09, Theorem 5.10]), we conclude that[1]

$$\begin{aligned}\operatorname{MCov}(p_x, \gamma) &= \inf_{\substack{f\colon \mathbb{R}^d \to \mathbb{R}\\ \text{convex}}} \Big(\int f\, dp_x + \int f^*\, d\gamma \Big) \\ &\leqslant \inf_{\substack{f\colon \mathbb{R}^d \to \mathbb{R}\\ \text{convex}}} \Big(\int f\, d\bar{p}_x + \int f^*\, d\gamma \Big) = \operatorname{MCov}(\bar{p}_x, \gamma).\end{aligned} \tag{3.11}$$

On the other hand, from $\operatorname{conv}\psi \leqslant \psi$ and (3.10) we have the inequality

$$\int \operatorname{conv}\psi\, d(\nu - p_x) \leqslant \int \psi\, d(\nu - \bar{p}_x) + \varepsilon. \tag{3.12}$$

Finally, summing (3.11) and (3.12), we obtain the inequality (3.8). □

Remark 3.4. On the right-hand side of (3.7) we can further require ψ to be smooth. Indeed, if $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ is convex, then $\psi^\varepsilon := \psi * \gamma^\varepsilon$ is a smooth convex function in $C_{\mathrm{q}}(\mathbb{R}^d)$ and

$$\liminf_{\varepsilon\to 0} \Big(\int \psi^\varepsilon\, d\nu - \int \varphi^{\psi^\varepsilon}\, d\mu \Big) \leqslant \int \psi\, d\nu - \int \varphi^\psi\, d\mu,$$

as follows by dominated convergence and the inequality $\psi^\varepsilon \geqslant \psi$. ◇

For $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$, we recall the definition (3.6) of the dual function $\mathcal{D}(\,\cdot\,)$, which we rewrite as

$$\mathcal{D}(\psi) = \int \Big(\int \psi(y)\, \pi_x(dy) - \varphi^\psi(x) \Big)\, \mu(dx), \tag{3.13}$$

where $\pi(dx, dy) = \pi_x(dy)\, \mu(dx)$ is an arbitrary fixed element of $\mathsf{MT}(\mu, \nu)$. The "relaxed" representation (3.13) of (3.6) allows us to extend the definition of the dual function $\mathcal{D}(\,\cdot\,)$ to convex functions $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ which are not confined to be in $C_{\mathrm{q}}(\mathbb{R}^d)$, but which are only required to be μ-a.s. finite, i.e., satisfy $\mu(\operatorname{dom}\psi) = 1$. This is summarized in the following lemma.

Lemma 3.5. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and fix an arbitrary martingale transport $\pi \in \mathsf{MT}(\mu, \nu)$. Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a convex function which is μ-a.s. finite. Formula (3.13) then defines $\mathcal{D}(\psi) \in [0, +\infty]$ and, recalling (3.2), we have the inequality*

$$\mathcal{D}(\psi) \geqslant \tilde{D}(\mu, \nu). \tag{3.14}$$

In particular, recalling (3.7), we have

$$\tilde{D}(\mu, \nu) = \inf_{\substack{\psi \in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \mathcal{D}(\psi) = \inf_{\substack{\mu(\operatorname{dom}\psi) = 1,\\ \psi \text{ convex}}} \mathcal{D}(\psi). \tag{3.15}$$

[1] In fact, $p_x \preceq_{\mathrm{c}} \bar{p}_x$ is equivalent to the inequality $\operatorname{MCov}(p_x, q) \leqslant \operatorname{MCov}(\bar{p}_x, q)$ being valid for all probability measures $q \in \mathscr{P}_2(\mathbb{R}^d)$, see. e.g. [AP22, Theorem 1] or [WZ23, Corollary 1.2].

Proof. Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a convex function which is μ-a.s. finite. First, note that by Jensen's inequality we have

$$\int \psi(y)\, \pi_x(dy) \geqslant \psi\Big(\int y\, \pi_x(dy) \Big) = \psi(x),$$

and by taking $p = \delta_x$ in (3.3) we obtain $\varphi^\psi(x) \leqslant \psi(x)$. Hence

$$\int \psi(y)\, \pi_x(dy) - \varphi^\psi(x) \geqslant 0, \tag{3.16}$$

for μ-a.e. $x \in \mathbb{R}^d$, so that $\mathcal{D}(\psi) \in [0, +\infty]$. In order to prove the inequality (3.14), we distinguish two cases. In the case $\int \psi\, d\pi_x = +\infty$, for $x \in \mathbb{R}^d$ in a set of positive μ-measure, we conclude from (3.16) that $\mathcal{D}(\psi) = +\infty$, so that (3.14) is trivially satisfied. Now suppose that $\int \psi\, d\pi_x < +\infty$, for μ-a.e. $x \in \mathbb{R}^d$. Recall from Theorem 3.1 that there exists an optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu, \nu)$ of (3.1). Taking $p = \pi_x^{\mathrm{SBM}}$ in (3.3) we get

$$\varphi^\psi(x) \leqslant \int \psi\, d\pi_x^{\mathrm{SBM}} - \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)$$

and therefore

$$\mathcal{D}(\psi) \geqslant \int \Big(\int \psi\, d\pi_x + \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) - \int \psi\, d\pi_x^{\mathrm{SBM}} \Big)\, \mu(dx) = \tilde{P}(\mu, \nu),$$

as π^{SBM} is a primal optimizer and both π^{SBM} and π have ν as second marginal. Since $\tilde{P}(\mu, \nu) = \tilde{D}(\mu, \nu)$ by Theorem 3.1, we again see that (3.14) is satisfied.

Finally, from (3.14) and recalling (3.7), we conclude

$$\tilde{D}(\mu, \nu) \leqslant \inf_{\substack{\mu(\mathrm{dom}\,\psi)=1,\\ \psi \text{ convex}}} \mathcal{D}(\psi) \leqslant \inf_{\substack{\psi \in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \mathcal{D}(\psi) = \tilde{D}(\mu, \nu),$$

which shows (3.15). □

4. Proof of the first part of Theorem 1.7

In Chapter 3 we already noted (by citing Theorem 2.2 of [BBHK20]) that the value $P(\mu,\nu)$ of the continuous-time optimization problem (1.4) equals $\tilde{P}(\mu,\nu)$, the value of the discrete-time formulation (3.1). We also know from Theorem 3.1 that there is no duality gap between the primal problem (3.1) and the dual problem (3.2), i.e., $\tilde{P}(\mu,\nu) = \tilde{D}(\mu,\nu)$. Therefore we can formulate the first part of Theorem 1.7 equivalently as follows.

Proposition 4.1. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then $\tilde{D}(\mu,\nu)$ is equal to*

$$D(\mu,\nu) = \inf_{\substack{\psi\in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big). \tag{4.1}$$

In Section 7.3 we will prove the second part of Theorem 1.7 and in particular discuss the existence of dual optimizers of (4.1). Since in general we cannot expect a dual optimizer to be integrable with respect to the probability measure ν, we need the following "relaxed" formulation of Proposition 4.1.

Proposition 4.2. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then $\tilde{D}(\mu,\nu)$ is equal to*

$$D_{\mathrm{rel}}(\mu,\nu) \coloneqq \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \mathcal{E}(\psi), \tag{4.2}$$

where

$$\mathcal{E}(\psi) \coloneqq \int \Big(\int \psi(y)\, \pi_x(dy) - (\psi^* * \gamma)^*(x) \Big)\, \mu(dx), \tag{4.3}$$

with π being an arbitrary fixed element of $\mathsf{MT}(\mu,\nu)$.

The main idea behind the proofs of Propositions 4.1, 4.2 (which we will present at the end of this chapter) is to apply Proposition 3.3 and then to show that $\varphi^\psi = (\psi^* * \gamma)^*$, for every convex function $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. This motivates our next goal, namely to solve the minimization problem (3.3), which we rewrite as a maximization problem

$$-\varphi^\psi(x) = \sup_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi \, dp \Big). \tag{4.4}$$

As a preliminary step, we consider the simpler problem

$$\varrho^\psi \coloneqq \sup_{p\in\mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,\gamma) - \int \psi \, dp \Big), \tag{4.5}$$

where we do not prescribe the barycenter x of $p \in \mathscr{P}_2(\mathbb{R}^d)$. In Lemma 4.4 below we will show for an arbitrary proper convex function $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$, that the value

ϱ^ψ equals $\int \psi^* \, d\gamma$. In fact, this information is essentially enough to prove Propositions 4.1, 4.2.

Solving the maximization problem (4.5) leads to an interesting connection with Brenier maps in Lemma 4.3 below. By *Brenier's theorem* (see, e.g., [Vil03, Theorem 2.12]), the optimal transport for quadratic cost between γ and $p \in \mathscr{P}_2(\mathbb{R}^d)$ is induced by the γ-a.e. defined gradient ∇v of some convex function $v \colon \mathbb{R}^d \to \mathbb{R}$ via $(\nabla v)(\gamma) = p$.

Lemma 4.3 ("reverse Brenier"). *Let $v \colon \mathbb{R}^d \to \mathbb{R}$ be a finite-valued convex function and $\psi := v^*$ its convex conjugate. Assume that the probability measure $\hat{p} := (\nabla v)(\gamma)$ has finite second moment. Then $\hat{p}$ is the unique maximizer of the optimization problem (4.5) and*

$$\varrho^\psi = \int v \, d\gamma = \int \psi^* \, d\gamma < +\infty.$$

Proof. We denote by T_γ^p the Brenier map from γ to $p \in \mathscr{P}_2(\mathbb{R}^d)$ and note that $T_\gamma^{\hat{p}} = \nabla v$. Since the convex function v is finite-valued, its gradient exists γ-a.e. and we have that

$$v(z) = \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) = \langle \nabla v(z), z \rangle - v^*(\nabla v(z)),$$

for γ-a.e. $z \in \mathbb{R}^d$. Using these observations, for $p \in \mathscr{P}_2(\mathbb{R}^d)$ we get

$$\begin{aligned}
\mathrm{MCov}(p, \gamma) - \int \psi \, dp &= \int \Big(\langle T_\gamma^p(z), z \rangle - v^*\big(T_\gamma^p(z)\big) \Big) \, \gamma(dz) \\
&\leqslant \int \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) \, \gamma(dz) = \int v(z) \, \gamma(dz) \\
&= \int \Big(\langle \nabla v(z), z \rangle - v^*\big(\nabla v(z)\big) \Big) \, \gamma(dz) \\
&= \mathrm{MCov}(\hat{p}, \gamma) - \int \psi \, d\hat{p},
\end{aligned}$$

with equality if and only if $T_\gamma^p(z) = T_\gamma^{\hat{p}}(z)$, for γ-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p = \hat{p}$. Finally, from the convexity of v and the Cauchy–Schwarz inequality we obtain

$$\int |v| \, d\gamma \leqslant |v(0)| + \sqrt{\int |\nabla v|^2 \, d\gamma} \cdot \sqrt{d} < +\infty, \tag{4.6}$$

which proves that $\varrho^\psi < +\infty$. □

Note that in Lemma 4.3 we started with a *finite-valued* convex function $v \colon \mathbb{R}^d \to \mathbb{R}$ and then *defined* $\psi := v^*$. If we additionally *know* that the probability measure $\hat{p} = (\nabla v)(\gamma)$ has finite second moment, then $\hat{p}$ is the unique maximizer of the optimization problem (4.5). These are rather strong assumptions. However, if we are given just a proper convex function $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$, we can still compute the value of the supremum in (4.5), without explicitly constructing a maximizer of this optimization problem.

Lemma 4.4. *Let $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper convex function. Then*

$$\varrho^\psi = \int \psi^* \, d\gamma. \tag{4.7}$$

Proof. Using probabilistic notation, we rewrite the supremum in (4.5) as

$$\varrho^{\psi} = \sup \mathbb{E}\big[\langle Y, Z\rangle - \psi(Y)\big], \tag{4.8}$$

where the supremum in (4.8) is taken over all probability spaces such that $Z \sim \gamma$ and Y is an $\mathbb{R}^d$-valued random variable with finite second moment. Replacing Y by $\mathbb{E}[Y \mid Z]$, we observe that the maximization in (4.8) can be restricted to random variables Y which are measurable functions of Z, and we obtain

$$\varrho^{\psi} = \sup_{Y \in L^2(\gamma;\mathbb{R}^d)} \int \Big(\langle Y(z), z\rangle - \psi(Y(z))\Big)\, \gamma(dz), \tag{4.9}$$

where $L^2(\gamma;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are square-integrable under γ. Clearly, for any $Y \in L^2(\gamma;\mathbb{R}^d)$ we have

$$\int \Big(\langle Y(z), z\rangle - \psi(Y(z))\Big)\, \gamma(dz) \leqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz) = \int \psi^*(z)\, \gamma(dz),$$

which shows the inequality $\varrho^{\psi} \leqslant \int \psi^*\, d\gamma$. In order to see the reverse inequality, we define the auxiliary problem

$$\varrho^{\psi}_{\infty} \coloneqq \sup_{Y \in L^\infty(\gamma;\mathbb{R}^d)} \int \Big(\langle Y(z), z\rangle - \psi(Y(z))\Big)\, \gamma(dz), \tag{4.10}$$

where $L^\infty(\gamma;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are bounded γ-a.e. Comparing (4.9) with (4.10), we obviously have $\varrho^{\psi} \geqslant \varrho^{\psi}_{\infty}$. Now we claim that

$$\varrho^{\psi}_{\infty} \geqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz) = \int \psi^*(z)\, \gamma(dz), \tag{4.11}$$

which will finish the proof of (4.7). To see this, we first write

$$\varrho^{\psi}_{\infty} = \lim_{N\to\infty} \sup_{\substack{Y \in L^\infty(\gamma;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\langle Y(z), z\rangle - \psi(Y(z))\Big)\, \gamma(dz).$$

Using a measurable selection argument, we obtain

$$\sup_{\substack{Y \in L^\infty(\gamma;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\langle Y(z), z\rangle - \psi(Y(z))\Big)\, \gamma(dz) \geqslant \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz).$$

Since ψ is proper we can choose $y_0 \in \operatorname{dom}\psi \neq \varnothing$. Then for N large enough we have

$$\sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y, z\rangle - \psi(y)\big) \geqslant \langle y_0, z\rangle - \psi(y_0),$$

with the right-hand side being integrable with respect to $\gamma(dz)$. Hence we can apply the monotone convergence theorem and deduce that

$$\lim_{N\to\infty} \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz) = \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, \gamma(dz),$$

which completes the proof of (4.11). □

As already announced, with the help of (4.7), we are now able to prove Propositions 4.1, 4.2. We first show a simpler variant, where we optimize over the class of convex functions ψ in $C_{\mathrm{q}}(\mathbb{R}^d)$ (i.e., which have quadratic growth), as in (4.12) of Proposition 4.5 below. Proposition 4.2, where we optimize over *all* μ-a.s. finite-valued convex functions $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ as in (4.2), is then a straightforward consequence. Proposition 4.1, where we optimize over all ν-integrable convex functions ψ as in (4.1), follows from a "sandwich argument".

Proposition 4.5. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then $\tilde{D}(\mu, \nu)$ is equal to*

$$D_{\mathrm{q}}(\mu, \nu) \coloneqq \inf_{\substack{\psi\in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \Big(\int \psi\, d\nu - \int (\psi^* * \gamma)^*\, d\mu \Big). \tag{4.12}$$

Proof. By Proposition 3.3 we have

$$\tilde{D}(\mu, \nu) = \inf_{\substack{\psi\in C_{\mathrm{q}}(\mathbb{R}^d),\\ \psi \text{ convex}}} \Big(\int \psi\, d\nu - \int \varphi^\psi\, d\mu \Big).$$

Hence it remains to show that $\varphi^\psi = (\psi^* * \gamma)^*$, for every convex function $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. In order to do this, we will first prove that $(\varphi^\psi)^* = \psi^* * \gamma$. By definition of the convex conjugate and (4.4), for $\zeta \in \mathbb{R}^d$, we have

$$\begin{aligned}
(\varphi^\psi)^*(\zeta) &= \sup_{x\in\mathbb{R}^d} \big(\langle x, \zeta\rangle - \varphi^\psi(x)\big) && (4.13)\\
&= \sup_{x\in\mathbb{R}^d} \sup_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) - \int \psi_\zeta\, dp\Big) && (4.14)\\
&= \sup_{p\in\mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) - \int \psi_\zeta\, dp\Big) = \varrho^{\psi_\zeta}, && (4.15)
\end{aligned}$$

where the function ψ_ζ is defined by $\psi_\zeta(y) \coloneqq \psi(y) - \langle \zeta, y\rangle$, for $y \in \mathbb{R}^d$. Now applying Lemma 4.4 to the proper convex function ψ_ζ yields

$$(\varphi^\psi)^*(\zeta) = \varrho^{\psi_\zeta} = \int \psi_\zeta^*\, d\gamma = \int \psi^*(\zeta + z)\, d\gamma(z) = (\psi^* * \gamma)(\zeta). \tag{4.16}$$

To complete the proof, we must justify that $(\varphi^\psi)^{**} = \varphi^\psi$, i.e. that the Fenchel–Moreau theorem is applicable. To see this, we first note that the function $x \mapsto \varphi^\psi(x)$ is convex

and we have the upper bound $\varphi^\psi \leqslant \psi < +\infty$. Furthermore, for any $p \in \mathscr{P}_2^x(\mathbb{R}^d)$ we have the inequalities

$$\mathrm{MCov}(p,\gamma) \leqslant \tfrac{1}{2}\int |y|^2\,dp(y) + \tfrac{d}{2}$$

as well as

$$\int \psi\,dp \geqslant \ell + \tfrac{1}{2}\int |y|^2\,dp(y),$$

the latter following from the fact that $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$. As a consequence, we get the lower bound $\varphi^\psi \geqslant \ell - \frac{d}{2} > -\infty$. Altogether, φ^ψ is a convex function which is finite everywhere on $\mathbb{R}^d$, thus it is continuous and we indeed have $(\varphi^\psi)^{**} = \varphi^\psi$. □

Proof of Proposition 4.2. We first show that the function $\mathcal{E}(\,\cdot\,)$ in (4.3) is well-defined for every convex function $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$, which is μ-a.s. finite. To this end, we will prove the inequality $(\psi^* * \gamma)^* \leqslant \psi$. By Jensen's inequality, this implies that the integrand in (4.3) is μ-a.s. nonnegative, and hence $\mathcal{E}(\psi)$ is well-defined and $[0,+\infty]$-valued. Recalling the equation (4.16) above, we have $\psi^* * \gamma = (\varphi^\psi)^*$. Taking the convex conjugate and using that $\varphi^\psi \leqslant \psi$, we obtain $(\psi^* * \gamma)^* = (\varphi^\psi)^{**} \leqslant \varphi^\psi \leqslant \psi$, as required.

Now let us turn to the proof of $\tilde{D}(\mu,\nu) = D_{\mathrm{rel}}(\mu,\nu)$. Recalling (4.2), (4.12), and using Proposition 4.5, we have $D_{\mathrm{rel}}(\mu,\nu) \leqslant D_{\mathrm{q}}(\mu,\nu) = \tilde{D}(\mu,\nu)$, so that we need to show the inequality $\tilde{D}(\mu,\nu) \leqslant D_{\mathrm{rel}}(\mu,\nu)$. Let $\pi \in \mathsf{MT}(\mu,\nu)$ and $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ be convex with $\mu(\operatorname{dom}\psi) = 1$. Since $\tilde{D}(\mu,\nu) = \tilde{P}(\mu,\nu)$ (recall Theorem 3.1 and (3.4)), we have to verify the inequality

$$\int \mathrm{MCov}(\pi_x,\gamma)\,d\mu(x) \leqslant \int \Big(\int \psi(y)\,\pi_x(dy) - (\psi^* * \gamma)^*(x)\Big)\,\mu(dx).$$

It is sufficient to prove, for μ-a.e. $x \in \mathbb{R}^d$, that

$$\mathrm{MCov}(\pi_x,\gamma) \leqslant \int \psi\,d\pi_x - (\psi^* * \gamma)^*(x). \tag{4.17}$$

We express the convex conjugate on the right-hand side of (4.17) as

$$-(\psi^* * \gamma)^*(x) = \inf_{\zeta\in\mathbb{R}^d}\big((\psi^* * \gamma)(\zeta) - \langle\zeta,x\rangle\big) = \inf_{\zeta\in\mathbb{R}^d}\Big(\int \psi_\zeta^*\,d\gamma - \langle\zeta,x\rangle\Big),$$

where the function ψ_ζ is defined by $\psi_\zeta(y) \coloneqq \psi(y) - \langle\zeta,y\rangle$, for $y \in \mathbb{R}^d$. Substituting back into (4.17) yields

$$\mathrm{MCov}(\pi_x,\gamma) \leqslant \inf_{\zeta\in\mathbb{R}^d}\Big(\int \psi_\zeta\,d\pi_x + \int \psi_\zeta^*\,d\gamma\Big). \tag{4.18}$$

Now observe that by the Fenchel–Young inequality we have

$$\mathrm{MCov}(\pi_x,\gamma) \leqslant \int \phi\,d\pi_x + \int \phi^*\,d\gamma,$$

for every proper convex function $\phi\colon \mathbb{R}^d \to (-\infty,+\infty]$. In particular, for every $\zeta \in \mathbb{R}^d$, it holds that

$$\mathrm{MCov}(\pi_x,\gamma) \leqslant \int \psi_\zeta\,d\pi_x + \int \psi_\zeta^*\,d\gamma,$$

which implies (4.18). This completes the proof of Proposition 4.2. □

Proof of Proposition 4.1. The assertion follows immediately from Proposition 4.2 and Proposition 4.5 by a "sandwich argument". Indeed, we have

$$\tilde{D}(\mu,\nu) = D_{\mathrm{rel}}(\mu,\nu) \leqslant D(\mu,\nu) \leqslant D_{\mathrm{q}}(\mu,\nu) = \tilde{D}(\mu,\nu). \tag{4.19}$$

The inequalities in (4.19) are due to the inclusions

$$\{\psi \in C_{\mathrm{q}}(\mathbb{R}^d)\colon \psi \text{ convex}\} \subseteq \{\psi \in L^1(\nu)\colon \psi \text{ convex}\} \subseteq \{\psi \text{ convex}\colon \mu(\operatorname{dom}\psi) = 1\};$$

the equalities on the left-hand side and on the right-hand side of (4.19) are justified by Proposition 4.2 and Proposition 4.5, respectively. □

5. Preparation for the proof of the second part of Theorem 1.7

The goal of this rather technical chapter is to outline and prepare the proof of the second part of Theorem 1.7, for which we will need the results of Chapters 6 and 7. In Chapter 7 we will show that the value $D(\mu,\nu)$ is attained by a convex function ψ_{opt} with $\mu(\mathrm{ri}(\mathrm{dom}\,\psi_{\mathrm{opt}}))=1$ if and only if (μ,ν) is irreducible. In Chapter 6 (see Theorem 6.6) we will prove that there is such a dual optimizer ψ_{opt} if and only if there exists a Bass martingale from μ to ν. At a first reading one might skip the present Chapter 5.

Comparing the primal problem (3.1), with optimizer $\pi^{\mathrm{SBM}}\in\mathsf{MT}(\mu,\nu)$, to the dual problem (3.2), with optimizer ψ_{opt} (supposing that this optimizer exists), we see that the minimization problem $\varphi^{\psi_{\mathrm{opt}}}(x)$ (recall (3.3)) is attained by $\pi_x^{\mathrm{SBM}}\in\mathscr{P}_2^x(\mathbb{R}^d)$, for μ-a.e. $x\in\mathbb{R}^d$. To draw the connection to the Bass martingale, we need to express the optimizer π_x^{SBM} in terms of the (single) convex function ψ_{opt}. More precisely, we will see as a consequence of Proposition 5.3 that $\pi_x^{\mathrm{SBM}}=(\nabla\psi_{\mathrm{opt}}^*)(\gamma_\zeta)$, for some appropriate $\zeta=\zeta(x)\in\mathbb{R}^d$, and where γ_ζ denotes the translated standard Gaussian measure on $\mathbb{R}^d$ with barycenter ζ. For this purpose, we have to return to the optimization problem (3.3) and study its optimizers.

First, we need the following technical result, formulated in Lemma 5.1 below. Let $\psi\colon\mathbb{R}^d\to(-\infty,+\infty]$ be a lower semicontinuous convex function. We recall from convex analysis that ψ is called co-finite if

$$\forall x\in\mathbb{R}^d\setminus\{0\}\colon\qquad\lim_{t\to+\infty}\frac{\psi(tx)}{t}=+\infty.$$

According to [Roc70, Corollary 13.3.1], the convex function ψ is co-finite if and only if its convex conjugate ψ^* is finite everywhere on $\mathbb{R}^d$. We derive now a sufficient condition for the co-finiteness of ψ, in terms of the function φ^ψ defined in (3.3).

Lemma 5.1. *Let $\psi\colon\mathbb{R}^d\to(-\infty,+\infty]$ be a proper, lower semicontinuous convex function. Then φ^ψ is convex on $\mathrm{dom}\,\psi$. If we additionally assume that $\varphi^\psi(x)>-\infty$, for some $x\in\mathrm{int}(\mathrm{dom}\,\psi)$, then $\varphi^\psi>-\infty$ on $\mathrm{int}(\mathrm{dom}\,\psi)$ and ψ is co-finite.*

Proof. To see that φ^ψ is convex on $\mathrm{dom}\,\psi$, we let $x:=cx_1+(1-c)x_2$ for $x_1,x_2\in\mathrm{dom}\,\psi$ and $c\in(0,1)$. If $\varphi^\psi(x_1)=-\infty$, then there are $p_{x_1}^{(n)}\in\mathscr{P}_2^{x_1}(\mathbb{R}^d)$ with

$$\int\psi\,dp_{x_1}^{(n)}-\mathrm{MCov}(p_{x_1}^{(n)},\gamma)\to-\infty.$$

We observe that $p_x^{(n)} := cp_{x_1}^{(n)} + (1-c)\delta_{x_2} \in \mathscr{P}_2^x(\mathbb{R}^d)$, and consequently

$$\begin{aligned}\varphi^\psi(x) &\leqslant c\int \psi\, dp_{x_1}^{(n)} + (1-c)\psi(x_2) - \mathrm{MCov}(p_x^{(n)}, \gamma)\\ &\leqslant c\Big(\int \psi\, dp_{x_1}^{(n)} - \mathrm{MCov}(p_{x_1}^{(n)}, \gamma)\Big) + (1-c)\psi(x_2),\end{aligned}$$

where we have used the convexity of $\mathscr{P}_2(\mathbb{R}^d) \ni p \mapsto -\mathrm{MCov}(p,\gamma)$. We conclude that $\varphi^\psi(x) = -\infty$. The case $\varphi^\psi(x_2) = -\infty$ is treated similarly. If, on the other hand, both $\varphi^\psi(x_1) > -\infty$ and $\varphi^\psi(x_2) > -\infty$, then $\varphi^\psi(x) \leqslant c\varphi^\psi(x_1) + (1-c)\varphi^\psi(x_2)$ follows by standard arguments. All in all we see that φ^ψ is convex on $\operatorname{dom}\psi$.

If $\varphi^\psi(x) > -\infty$ for one $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\varphi^\psi(\tilde{x}) > -\infty$ for all $\tilde{x} \in \operatorname{int}(\operatorname{dom}\psi)$, as can be seen directly by convexity.

Finally, given some $x \in \operatorname{int}(\operatorname{dom}\psi)$ with $\varphi^\psi(x) > -\infty$, we show that ψ is co-finite. Without loss of generality, we assume that $x = 0$ and $\psi(0) = 0$. Using probabilistic notation, we rewrite the supremum in (4.4) as

$$-\varphi^\psi(0) = \sup \mathbb{E}\big[\langle Y, Z\rangle - \psi(Y)\big] < +\infty, \tag{5.1}$$

where the supremum in (5.1) is taken over all probability spaces such that $Z \sim \gamma$ and Y is an $\mathbb{R}^d$-valued random variable with finite second moment and $\mathbb{E}[Y] = 0$. By contradiction, suppose there is $e \in \mathbb{R}^d \setminus \{0\}$ such that

$$\lim_{t\to+\infty} \frac{\psi(te)}{t} < +\infty. \tag{5.2}$$

We define the convex function $\bar{\psi}\colon \mathbb{R} \to (-\infty, +\infty]$ by $\bar{\psi}(t) := \psi(te)$, for $t \in \mathbb{R}$. By assumption we have $0 \in \operatorname{int}(\operatorname{dom}\psi)$, thus also $0 \in \operatorname{int}(\operatorname{dom}\bar{\psi})$. In particular, $\bar{\psi}$ is continuous in a neighbourhood of 0. By (5.2), there is a constant K such that $\bar{\psi}(t) \leqslant Kt$, for all t large enough. As $\bar{\psi}(0) = 0$ and $\bar{\psi}$ is convex, we conclude that $\bar{\psi}(t) \leqslant Kt$, for all $t \geqslant 0$. Next we introduce four parameters, $A < 0$ and $\delta, M, C > 0$, and define the $\mathbb{R}^d$-valued function

$$\mathbb{R}^d \ni z \longmapsto f_{\delta,A}^{M,C}(z) := \begin{cases} Me, & \text{if } \langle e, z\rangle \geqslant C,\\ -\delta e, & \text{if } \langle e, z\rangle \in (A, 0),\\ 0, & \text{else.}\end{cases}$$

Given $Z \sim \gamma$, this induces a random variable $Y_{\delta,A}^{M,C} := f_{\delta,A}^{M,C}(Z)$. We impose the relation

$$M\,\mathbb{P}\big[\langle e, Z\rangle \geqslant C\big] = \delta\,\mathbb{P}\big[\langle e, Z\rangle \in (A, 0)\big] \tag{5.3}$$

between M and δ, so that $\mathbb{E}[Y_{\delta,A}^{M,C}] = 0$. Now we fix $A < 0$ and some $\varepsilon > 0$. Then we choose $\delta > 0$ small enough such that

$$|\bar{\psi}(-\delta)| \vee \delta\,\mathbb{E}\big[|\langle e, Z\rangle|\big] \leqslant \tfrac{\varepsilon}{2},$$

which we achieve by continuity of $\bar{\psi}$ and the fact that $\bar{\psi}(0) = 0$. We leave C as a free parameter, which fixes M via (5.3). We compute

$$\begin{aligned}
&\mathbb{E}\big[\langle Y_{\delta,A}^{M,C}, Z\rangle - \psi(Y_{\delta,A}^{M,C})\big] \\
&= \mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle \geqslant C}\big(M\langle e, Z\rangle - \bar{\psi}(M)\big) + \mathbf{1}_{\langle e,Z\rangle \in (A,0)}\big(-\delta\langle e, Z\rangle - \bar{\psi}(-\delta)\big)\big] \\
&\geqslant M\,\mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle \geqslant C}\big(\langle e, Z\rangle - K\big)\big] - \delta\,\mathbb{E}\big[\mathbf{1}_{\langle e,Z\rangle \in (A,0)}\langle e, Z\rangle\big] - \bar{\psi}(-\delta)\,\mathbb{P}\big[\langle e, Z\rangle \in (A, 0)\big] \\
&\geqslant M(C - K)\,\mathbb{P}\big[\langle e, Z\rangle \geqslant C\big] - \varepsilon \\
&= \delta(C - K)\,\mathbb{P}\big[\langle e, Z\rangle \in (A, 0)\big] - \varepsilon.
\end{aligned}$$

Now taking $C \nearrow +\infty$ we conclude that $-\varphi^{\psi}(0) = +\infty$, which is a contradiction to (5.1). □

Our next result is the analogue of Lemma 4.4. But now, instead of maximizing over all $p \in \mathscr{P}_2(\mathbb{R}^d)$ as in (4.5), we have to maximize over all $p \in \mathscr{P}_2^x(\mathbb{R}^d)$, with a fixed barycenter $x \in \mathbb{R}^d$, as in (4.4).

Proposition 5.2. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function and assume that $\varphi^{\psi}(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$. Then we have the duality formula*

$$\varphi^{\psi}(x) = \sup_{\zeta \in \mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z)\Big), \tag{5.4}$$

and the right-hand side admits a unique maximizer $\zeta(x) \in \mathbb{R}^d$.

Proof. We deploy a similar strategy as in the proof of Proposition 4.5. Recalling equations (4.13) – (4.16), for each $\zeta \in \mathbb{R}^d$, we have

$$(\varphi^{\psi})^*(\zeta) = \sup_{p \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) - \int \psi_\zeta\, dp\Big) = \int \psi^*(\zeta + z)\, d\gamma(z), \tag{5.5}$$

where the function ψ_ζ is defined by $\psi_\zeta(y) \coloneqq \psi(y) - \langle \zeta, y\rangle$, for $y \in \mathbb{R}^d$. Since $\varphi^{\psi}(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$, the convex function $\xi \mapsto \varphi^{\psi}(\xi)$ is finite and thus also continuous in a neighbourhood of x. Hence we can apply the variant [BC17, Proposition 13.44] of the Fenchel–Moreau theorem and obtain

$$\varphi^{\psi}(x) = (\varphi^{\psi})^{**}(x) = \sup_{\zeta \in \mathbb{R}^d} \Big(\langle \zeta, x\rangle - \int \psi^*(\zeta + z)\, d\gamma(z)\Big), \tag{5.6}$$

which proves the duality formula (5.4).

To prove the existence of a maximizer $\zeta(x) \in \mathbb{R}^d$, we define the function

$$f_x(\zeta) \coloneqq \sup_{p \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) - \int \psi\, dp + \int \langle \zeta, y - x\rangle\, dp(y)\Big), \tag{5.7}$$

so that $-\varphi^{\psi}(x) = \inf_{\zeta\in\mathbb{R}^d} f_x(\zeta)$. Indeed, it follows from (5.5), (5.6) that

$$\begin{aligned}-\varphi^{\psi}(x) &= \inf_{\zeta\in\mathbb{R}^d}\Big(\int \psi^*(\zeta+z)\,d\gamma(z) - \langle\zeta, x\rangle\Big)\\ &= \inf_{\zeta\in\mathbb{R}^d}\sup_{p\in\mathscr{P}_2(\mathbb{R}^d)}\Big(\mathrm{MCov}(p,\gamma) - \int \psi_\zeta\,dp - \langle\zeta,x\rangle\Big).\end{aligned}$$

By assumption, the function f_x takes values in $(-\infty,+\infty]$ and is not constant equal to $+\infty$. Equivalently, we can express $f_x(\zeta)$ as

$$f_x(\zeta) = \int \left(\psi^*(\zeta+z) + \psi(x) - \langle\zeta+z, x\rangle\right) d\gamma(z) - \psi(x). \tag{5.8}$$

We will show that $\zeta \mapsto f_x(\zeta)$ is lower semicontinuous and coercive, implying the existence of an optimizer $\zeta(x)$. It is easy to see from the representation (5.8) that f_x is lower semicontinuous. Indeed, this directly follows from the lower semicontinuity of the nonnegative integrand in (5.8) and Fatou's lemma. For the verification of the coercivity, we take a sequence $(\zeta^{(n)})_{n\geqslant 1}$ in $\mathbb{R}^d$ with $|\zeta^{(n)}| \to +\infty$. Then there is a coordinate $k \in \{1,\dots,d\}$ such that $|\zeta_k^{(n)}| \to +\infty$. Since $x \in \operatorname{int}(\operatorname{dom}\psi)$, we can choose $\varepsilon > 0$ small enough such that $x \pm \varepsilon\, e_k \in \operatorname{dom}\psi$, where e_k denotes the k-th standard basis vector of $\mathbb{R}^d$. Defining

$$y^{(n)} := x + \operatorname{sign}(\zeta_k^{(n)})\,\varepsilon\, e_k \in \operatorname{dom}\psi$$

and taking $p = \delta_{y^{(n)}}$, we conclude from (5.7) that

$$f_x(\zeta^{(n)}) \geqslant -\psi(y^{(n)}) + \varepsilon\,|\zeta_k^{(n)}| \to +\infty,$$

which shows $\lim_{|\zeta|\to+\infty} f_x(\zeta) = +\infty$.

We now show that the maximizer $\zeta(x) \in \mathbb{R}^d$ of the right-hand side of (5.4) is unique. As the assumptions of Lemma 5.1 are satisfied, the convex function ψ is co-finite, and thus ψ^* is finite everywhere on $\mathbb{R}^d$. In particular, ψ^* is continuous everywhere and differentiable Lebesgue-a.e. on $\mathbb{R}^d$. In order to establish the uniqueness of $\zeta(x)$, we show that the function $\zeta \mapsto \int \psi^*(\zeta+z)\,d\gamma(z) \in (-\infty,+\infty]$ is strictly convex on its domain. By contradiction, suppose that there are $\zeta_1,\zeta_2 \in \mathbb{R}^d$ with $\zeta_1 \neq \zeta_2$ such that both $\int \psi^*(\zeta_1+z)\,d\gamma(z)$ and $\int \psi^*(\zeta_2+z)\,d\gamma(z)$ are finite and

$$\int \psi^*\big(t\zeta_1 + (1-t)\zeta_2 + z\big)\,d\gamma(z) = t\int \psi^*(\zeta_1+z)\,d\gamma(z) + (1-t)\int \psi^*(\zeta_2+z)\,d\gamma(z),$$

for some $t \in (0,1)$. Then

$$\psi^*\big(t\zeta_1 + (1-t)\zeta_2 + z\big) = t\psi^*(\zeta_1+z) + (1-t)\psi^*(\zeta_2+z),$$

for γ-a.e. $z \in \mathbb{R}^d$. Consequently, the function ψ^* is affine on the line segment from ζ_1+z to ζ_2+z, for Lebesgue-a.e. $z \in \mathbb{R}^d$. By convexity and finiteness of ψ^* on $\mathbb{R}^d$, we thus have that the function

$$\mathbb{R} \in \lambda \longmapsto \psi^*\big(\lambda(\zeta_2-\zeta_1)+z\big)$$

is affine, for every $z \in \mathbb{R}^d$. By [Roc70, Theorem 8.8], there exists some $c \in \mathbb{R}^d$ such that

$$\psi^*(\lambda(\zeta_2 - \zeta_1) + z) = \psi^*(z) + \lambda c.$$

Differentiating at the point $\lambda = 1$ yields $\langle \nabla\psi^*(z), \zeta_2 - \zeta_1 \rangle = c$, for Lebesgue-a.e. $z \in \mathbb{R}^d$. But this is a contradiction to $\varnothing \neq \operatorname{int}(\operatorname{dom}\psi) \subseteq (\partial\psi^*)(\mathbb{R}^d)$, where the symbol ∂, applied to a convex function, denotes its subdifferential. □

The final result of this chapter is the analogue of Lemma 4.3. The duality formula (5.4) established in Proposition 5.2 enables us to identify the structure of the optimizer in the maximization problem (4.4).

Proposition 5.3. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function. For fixed $\zeta \in \mathbb{R}^d$ we define the function ψ_ζ by $\psi_\zeta(y) := \psi(y) - \langle \zeta, y \rangle$, for $y \in \mathbb{R}^d$, so that $\psi_\zeta^*(y) = \psi^*(\zeta + y)$.*

(i) If there are $x \in \mathbb{R}^d$ and $\zeta = \zeta(x) \in \mathbb{R}^d$ such that $(\nabla\psi_\zeta^*)(\gamma) \in \mathscr{P}_2^x(\mathbb{R}^d)$, then this is actually the unique optimizer of the supremum in (4.4). If additionally $\varphi^\psi(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\zeta = \zeta(x)$ is the optimizer of the supremum in (5.4).

(ii) If $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$ and the supremum in (4.4) is attained by some $\hat{p}_x \in \mathscr{P}_2^x(\mathbb{R}^d)$, then $\hat{p}_x = (\nabla\psi_\zeta^*)(\gamma)$ with $\zeta = \zeta(x)$ being the optimizer of the supremum in (5.4).

Proof. (i) We set $\hat{p}_x := (\nabla\psi_\zeta^*)(\gamma)$ and let $p_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ be arbitrary. We denote by $T_\gamma^{p_x}$ the Brenier map from γ to p_x, note that $T_\gamma^{\hat{p}_x} = \nabla\psi_\zeta^*$, and compute

$$\begin{aligned}
\int \psi \, dp_x - \operatorname{MCov}(p_x, \gamma) &= \int \Big(\psi_\zeta(T_\gamma^{p_x}(z)) + \langle \zeta, T_\gamma^{p_x}(z) \rangle - \langle T_\gamma^{p_x}(z), z \rangle \Big) \, d\gamma(z) \\
&\geqslant \langle \zeta, x \rangle + \int \inf_{y \in \mathbb{R}^d} \left(\psi_\zeta(y) - \langle y, z \rangle \right) d\gamma(z) \\
&= \langle \zeta, x \rangle - \int \psi^*(\zeta + z) \, d\gamma(z) \\
&= \int \psi \, d\hat{p}_x - \operatorname{MCov}(\hat{p}_x, \gamma),
\end{aligned}$$

with equality if and only if $T_\gamma^{p_x}(z) = T_\gamma^{\hat{p}_x}(z)$, for γ-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p_x = \hat{p}_x$. We conclude that $\hat{p}_x$ is the unique optimizer of the supremum in (4.4), i.e.

$$\varphi^\psi(x) = \int \psi \, d\hat{p}_x - \operatorname{MCov}(\hat{p}_x, \gamma) = \langle \zeta, x \rangle - \int \psi^*(\zeta + z) \, d\gamma(z). \tag{5.9}$$

On the other hand, if additionally $\varphi^\psi(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}\psi)$, by Proposition 5.2 we have that

$$\varphi^\psi(x) = \sup_{\zeta \in \mathbb{R}^d} \Big(\langle \zeta, x \rangle - \int \psi^*(\zeta + z) \, d\gamma(z) \Big),$$

which in light of (5.9) implies that $\zeta = \zeta(x)$ is the optimizer of the supremum in (5.4).
(ii) By assumption and Proposition 5.2, we have that

$$\begin{aligned}\varphi^{\psi}(x) &= \langle \zeta, x\rangle + \int \left(\psi_{\zeta}(T_{\gamma}^{\hat{p}_x}(z)) - \langle T_{\gamma}^{\hat{p}_x}(z), z\rangle\right) d\gamma(z)\\ &= \langle \zeta, x\rangle - \int \psi_{\zeta}^{*}(z)\, d\gamma(z)\end{aligned}$$

and therefore

$$\int \left(\psi_{\zeta}(T_{\gamma}^{\hat{p}_x}(z)) + \psi_{\zeta}^{*}(z) - \langle T_{\gamma}^{\hat{p}_x}(z), z\rangle\right) d\gamma(z) = 0.$$

As the integrand is pointwise nonnegative, we conclude that

$$\psi(T_{\gamma}^{\hat{p}_x}(z)) + \psi^{*}(\zeta + z) = \langle T_{\gamma}^{\hat{p}_x}(z), \zeta + z\rangle$$

and consequently $T_{\gamma}^{\hat{p}_x}(z) \in \partial\psi^{*}(\zeta + z)$, for γ-a.e. $z \in \mathbb{R}^d$. Hence $T_{\gamma}^{\hat{p}_x}(z) = \nabla\psi^{*}(\zeta + z)$, for γ-a.e. $z \in \mathbb{R}^d$, which implies that $\hat{p}_x = (\nabla\psi_{\zeta}^{*})(\gamma)$. □

6. Existence of dual optimizers and Bass martingales

Throughout this chapter we fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$.

Definition 6.1. A lower semicontinuous convex function $\psi_{\mathrm{opt}} \colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\mathrm{ri}(\mathrm{dom}\, \psi_{\mathrm{opt}})) = 1$ is called an optimizer of the dual problem (3.15) (in short, a dual optimizer) if $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\mathrm{opt}})$, for the dual function $\mathcal{D}(\,\cdot\,)$ as defined in (3.13).

As the next result shows, we could have required in the definition of a dual optimizer ψ_{opt} that the affine dimension of its domain is at least as large as the affine dimension of the support of the probability measure ν.

Lemma 6.2. *If ψ_{opt} is an optimizer according to Definition 6.1, then*

$$\dim(\mathrm{dom}\, \psi_{\mathrm{opt}}) \geqslant \dim(\mathrm{supp}\, \nu).$$

Proof. By contradiction, we suppose that $\dim(\mathrm{dom}\, \psi_{\mathrm{opt}}) < \dim(\mathrm{supp}\, \nu)$. Defining

$$B_\ell \coloneqq \{x \in \mathbb{R}^d \colon \psi_{\mathrm{opt}}(x) \leqslant \ell\}$$

for $\ell \geqslant 1$, we clearly have $\mu(B_\ell) \nearrow 1$ and furthermore

$$\sup_{\ell \geqslant 1} \pi^{\mathrm{SBM}}\big(B_\ell \times (\mathbb{R}^d \setminus \mathrm{dom}\, \psi_{\mathrm{opt}})\big) > 0.$$

Indeed, otherwise ν would be concentrated on $\mathrm{dom}\, \psi_{\mathrm{opt}}$, contradicting the assumption that $\dim(\mathrm{dom}\, \psi_{\mathrm{opt}}) < \dim(\mathrm{supp}\, \nu)$. As a consequence,

$$\int_{B_\ell \times \mathbb{R}^d} \psi_{\mathrm{opt}}(y)\, \pi^{\mathrm{SBM}}(dx, dy) = +\infty$$

for all $\ell \geqslant 1$ large enough. We now show that ψ_{opt} could not have been optimal, by establishing that

$$\mathcal{D}(\psi_{\mathrm{opt}}) = \int \Big(\sup_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) - \int \psi_{\mathrm{opt}}\, dp \Big) + \int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) \Big)\, \mu(dx) = +\infty.$$

To wit, selecting $p_x = \pi_x^{\mathrm{SBM}}$ for $x \in \mathbb{R}^d \setminus B_\ell$ and $p_x = \delta_x$ for $x \in B_\ell$ we find

$$\begin{aligned} \mathcal{D}(\psi_{\mathrm{opt}}) &\geqslant \int_{\mathbb{R}^d \setminus B_\ell} \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, \mu(dx) - \int_{B_\ell} \Big(\psi_{\mathrm{opt}}(x) - \int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) \Big)\, \mu(dx) \\ &\geqslant c - \ell + \int_{B_\ell \times \mathbb{R}^d} \psi_{\mathrm{opt}}(y)\, \pi^{\mathrm{SBM}}(dx, dy), \end{aligned}$$

for a finite constant c. We conclude by taking $\ell \geqslant 1$ large enough. □

We recall Definition 1.1 and list some important properties of Bass martingales. Such martingales were introduced in [BBHK20] under the name of a "standard stretched Brownian motion". In this work, we use — with Richard Bass' permission — the term "Bass martingale" instead.

Remark 6.3. Let $M = (M_t)_{0\leqslant t\leqslant 1}$ be a Bass martingale from μ to ν, with corresponding convex function $v\colon \mathbb{R}^d \to \mathbb{R}$ and initial distribution $\mathscr{P}(\mathbb{R}^d) \ni \alpha \sim B_0$ of the underlying Brownian motion $B = (B_t)_{0\leqslant t\leqslant 1}$.

(i) As shown in [BBHK20] (and as a result of Theorem 6.6 below), the martingale transport

$$\mathrm{Law}(M_0, M_1) \in \mathsf{MT}(\mu, \nu)$$

is equal to the unique optimizer π^{SBM} of (3.1). Furthermore, the knowledge of π^{SBM} already determines the martingale $(M_t)_{0\leqslant t\leqslant 1}$ as well as the function v, which is $(\alpha * \gamma)$-a.e. (equivalently, Lebesgue-a.e.) unique up to an additive constant.

(ii) The convex function v and the probability measure α satisfy the identities

$$(\nabla v * \gamma)(\alpha) = \mu \qquad \text{and} \qquad \nabla v(\alpha * \gamma) = \nu, \tag{6.1}$$

which we summarize in the following graphic:

$$\begin{array}{ccc} \alpha * \gamma & \xrightarrow{\nabla v} & \nu \\ {\scriptstyle *}\uparrow & & \\ \alpha & \xrightarrow{\nabla v * \gamma} & \mu \end{array}$$

(iii) We also remark (see [BBHK20]) that we have

$$M_t = (\nabla v * \gamma^{1-t})(B_t) \quad \text{and} \quad \pi^{\mathrm{SBM}} = \mathrm{Law}\left((\nabla v * \gamma)(B_0), \nabla v(B_1)\right), \tag{6.2}$$

where γ^s denotes the d-dimensional centered Gaussian distribution with covariance matrix sI_d.

(iv) Let E be the affine space generated by the support of ν, and call d_E its dimension and P_E the orthogonal projection from $\mathbb{R}^d$ onto E. Then $v(x) = \tilde{v}(P_E x)$, for a convex function $\tilde{v}\colon E \to \mathbb{R}$. If we denote $\tilde{B}_t \coloneqq P_E B_t$, which is a d_E-dimensional Brownian motion on E with initial distribution $\tilde{\alpha} \coloneqq P_E(\alpha)$, then

$$M_t = (\nabla v * \gamma^{1-t})(B_t) = (\nabla_E \tilde{v} * \tilde{\gamma}^{1-t})(\tilde{B}_t),$$

where $\tilde{\gamma}^{1-t} \coloneqq P_E(\gamma^{1-t})$ is the d_E-dimensional centered Gaussian distribution with covariance matrix $(1-t)I_{d_E}$. This provides a "minimal representation" of the Bass martingale M, and therefore — if necessary — we can always assume that $E = \mathbb{R}^d$ by redefining $d \coloneqq d_E$.

(v) In the same spirit as the previous point, the kernels of every martingale transport π from μ to ν are supported on E. It follows that $\mathrm{MCov}(\pi_x, \gamma) = \mathrm{MCov}(\pi_x, P_E(\gamma))$ and hence the primal problem (3.1) may be redefined as

$$\tilde{P}(\mu, \nu) \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, P_E(\gamma))\, \mu(dx).$$

Thus also the dual problems may be redefined as taking place over functions on the affine space E. ◇

The next result, Theorem 6.6 below, explains how the existence of a dual optimizer ψ_{opt} is related to the existence of a Bass martingale. Along with it we need the following technical lemmas.

Lemma 6.4. *Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a finite convex function such that $\nabla f \in L^2(\gamma_\zeta; \mathbb{R}^d)$, for some $\zeta \in \mathbb{R}^d$. Then*

(i) $\nabla(f * \gamma) = (\nabla f) * \gamma$.

If additionally $\mathrm{int}(\mathrm{dom}\, f^*) \neq \varnothing$, *then*

(ii) $(f * \gamma)^*$ *is differentiable and strictly convex on* $\mathrm{int}(\mathrm{dom}(f * \gamma)^*) \supseteq \mathrm{int}(\mathrm{dom}\, f^*)$,

(iii) $\nabla(f * \gamma)^*\colon \mathrm{int}(\mathrm{dom}(f * \gamma)^*) \to \mathbb{R}^d$ *and* $\nabla(f * \gamma)\colon \mathbb{R}^d \to \mathrm{int}(\mathrm{dom}(f * \gamma)^*)$ *are bijections, and we have*

$$(\nabla f * \gamma)^{-1} = \nabla(f * \gamma)^*.$$

Proof. We first prove the identity (i). We denote $F \coloneqq \nabla f$, which is well-defined Lebesgue-a.e. For $y, \eta \in \mathbb{R}^d$ we have to compute

$$\frac{(f * \gamma)(y + h\eta) - (f * \gamma)(y)}{h} = \int \frac{f(y + h\eta + z) - f(y + z)}{h}\, d\gamma(z),$$

as $h \to 0$. For all y, η and for γ-a.e. $z \in \mathbb{R}^d$, we have

$$\lim_{h \to 0} \frac{f(y + h\eta + z) - f(y + z)}{h} = \langle F(y + z), \eta \rangle,$$

so we only need to justify the exchange of limit and integral. To this end, it suffices to show the uniform integrability, with respect to the reference measure γ, of the family

$$\mathscr{U} \coloneqq \left\{ \frac{f(y + h\eta + \cdot\,) - f(y + \cdot\,)}{h} \colon 0 \leqslant h \leqslant 1 \right\}. \tag{6.3}$$

Using twice the above-tangent characterization of convexity, we obtain

$$\langle F(y + z), \eta \rangle \leqslant \frac{f(y + h\eta + z) - f(y + z)}{h} \leqslant \langle F(y + h\eta + z), \eta \rangle. \tag{6.4}$$

We take $r \in (1,2)$, $p \coloneqq 2/r$, and $q \coloneqq p/(p-1)$ its Hölder conjugate, and compute

$$\begin{aligned}
&\int |F(y+h\eta+z)|^r \, d\gamma(z) = \int |F(\zeta+z)|^r \, d\gamma_{y+h\eta-\zeta}(z) \\
&= \int |F(\zeta+z)|^r \, \exp\big(-|y-\zeta+h\eta|^2/2 + \langle z, y-\zeta+h\eta\rangle\big) \, d\gamma(z) \\
&\leqslant \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta;\mathbb{R}^d)}} \sqrt[q]{\int \exp\big(-q|y-\zeta+h\eta|^2/2 + q\langle z, y-\zeta+h\eta\rangle\big) \, d\gamma(z)} \\
&\leqslant \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta;\mathbb{R}^d)}} \sqrt[q]{\int \exp\big(q\langle z, y-\zeta+h\eta\rangle\big) \, d\gamma(z)} \\
&= \sqrt[p]{\|F\|^2_{L^2(\gamma_\zeta;\mathbb{R}^d)}} \sqrt[q]{\exp\big(q^2|y-\zeta+h\eta|^2/2\big)}. \qquad (6.5)
\end{aligned}$$

We see that the first factor in (6.5) is finite (and independent of h) by assumption, while the second factor is uniformly bounded for $0 \leqslant h \leqslant 1$. Hence both the upper and lower bounds in (6.4) are uniformly integrable for $0 \leqslant h \leqslant 1$, implying the same property for the family $\mathscr{U}$ of (6.3).

We now turn to (ii) and (iii). The arguments given at the end of the proof of Proposition 5.2 show that, under the assumptions made here (but with f instead of ψ^*), the function $f * \gamma$ is strictly convex. We already know that this function is differentiable, and that the interior of its domain is $\mathbb{R}^d$. Hence by [Roc70, Theorem 26.5] we deduce the stated properties. The only thing that merits an explanation is the inclusion $\operatorname{int}(\operatorname{dom}(f*\gamma)^*) \supseteq \operatorname{int}(\operatorname{dom} f^*)$. To this end, observe that

$$\begin{aligned}
(f*\gamma)^*(x) &= \sup_{\zeta\in\mathbb{R}^d} \Big\{\langle \zeta, x\rangle - \int f(\zeta+z)\, d\gamma(z)\Big\} \leqslant \int \sup_{\zeta\in\mathbb{R}^d} \{\langle \zeta, x\rangle - f(\zeta+z)\}\, d\gamma(z) \\
&= \int \sup_{\zeta\in\mathbb{R}^d} \{\langle \zeta - z, x\rangle - f(\zeta)\}\, d\gamma(z) = \int \big(f^*(x) - \langle z, x\rangle\big)\, d\gamma(z) = f^*(x),
\end{aligned}$$

so $\operatorname{dom}(f*\gamma)^* \supseteq \operatorname{dom} f^*$. □

Lemma 6.5. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function and assume that $\varphi^\psi(x) > -\infty$, for some $x \in \operatorname{int}(\operatorname{dom}\psi)$. Furthermore, we suppose that $\nabla\psi^* \in L^2(\gamma_\zeta;\mathbb{R}^d)$, for some $\zeta \in \mathbb{R}^d$. Then, for all $\eta \in \mathbb{R}^d$ and all $t > 0$, we have*

$$\operatorname{int}(\operatorname{dom}\psi) = \operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = (\nabla\psi^* * \gamma_\eta^t)(\mathbb{R}^d), \qquad (6.6)$$

where γ_η^t denotes the d-dimensional Gaussian distribution with barycenter η and covariance matrix tI_d.

Proof. The first equality in (6.6) follows from

$$\operatorname{int}(\operatorname{dom}\psi) \subseteq \operatorname{dom}\partial\psi \subseteq \operatorname{dom}\psi$$

and $\operatorname{dom}\partial\psi = (\partial\psi^*)(\mathbb{R}^d)$. We also note that

$$\operatorname{int}\big((\partial\psi^*)(\mathbb{R}^d)\big) = \operatorname{int}\operatorname{conv}\big((\partial\psi^*)(\mathbb{R}^d)\big), \qquad (6.7)$$

where conv denotes the convex hull. Indeed, as a consequence of a result by Mirsky [Mir61], we have that $\operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)) \subseteq (\partial\psi^*)(\mathbb{R}^d)$, since the subdifferential (multi-valued) mapping of any lower semicontinuous proper convex function is also a maximal monotone mapping by [Roc70, Corollary 31.5.2].

Let us turn to the proof of the second equality in (6.6). Without loss of generality we can assume that $\eta = 0$ and $t = 1$. Since we already verified the first equality and by virtue of (6.7), it suffices to show the inclusions

$$(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)) \tag{6.8}$$

and

$$\operatorname{int}(\operatorname{dom}\psi) \subseteq (\nabla\psi^* * \gamma)(\mathbb{R}^d). \tag{6.9}$$

Note that we have $(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \overline{\operatorname{conv}}((\partial\psi^*)(\mathbb{R}^d))$ and therefore

$$\operatorname{int}(\nabla\psi^* * \gamma)(\mathbb{R}^d) \subseteq \operatorname{int}\overline{\operatorname{conv}}((\partial\psi^*)(\mathbb{R}^d)) = \operatorname{int}\operatorname{conv}((\partial\psi^*)(\mathbb{R}^d)). \tag{6.10}$$

Under the assumptions of Lemma 6.5, the function ψ is co-finite by Lemma 5.1 and therefore we can apply Lemma 6.4 to its convex conjugate $f = \psi^*$. In particular, by (iii), the set $(\nabla\psi^* * \gamma)(\mathbb{R}^d)$ is open and hence the inclusion (6.8) follows from (6.10).

In order to prove (6.9), we first observe that $\operatorname{int}(\operatorname{dom}\psi) \subseteq \operatorname{int}(\operatorname{dom}(\psi^* * \gamma)^*)$ by Lemma 6.4, (ii). Since $\operatorname{int}(\operatorname{dom}(\psi^* * \gamma)^*) = (\nabla\psi^* * \gamma)(\mathbb{R}^d)$ by Lemma 6.4, (iii), the inclusion (6.9) follows. □

Theorem 6.6. *There exists a dual optimizer ψ_{opt} in the sense of Definition 6.1 if and only if there exists a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν. In this case,* $\operatorname{Law}(M_0, M_1)$ *is equal to the optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (3.1), and — after passing to the minimal representation discussed in Remark 6.3, (iv), (v) and redefining the dimension of the problem to be* $\dim(\operatorname{supp}\nu)$ *if necessary — $\psi_{\mathrm{opt}}, v, \alpha, \pi^{\mathrm{SBM}}$ are related via $v^* = \psi_{\mathrm{opt}}$, $\alpha = \zeta(\mu)$ and*

$$\pi^{\mathrm{SBM}}_x = \operatorname{Law}(M_1 \,|\, M_0 = x) = \nabla v(\zeta(x) + \,\cdot\,)(\gamma), \tag{6.11}$$

where the function $\zeta\colon X \to \mathbb{R}^d$ is given by

$$\zeta(x) := \operatorname{argmax}_{\zeta\in\mathbb{R}^d}\Big(\langle\zeta, x\rangle - \int v(\zeta + z)\,d\gamma(z)\Big) \tag{6.12}$$

as in (5.4), and we have

$$\zeta(x) = (\nabla v * \gamma)^{-1}(x) = \nabla(v * \gamma)^*(x), \tag{6.13}$$

for all $x \in X$. Furthermore, the domain X of the function ζ satisfies $X \subseteq \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$ and $\mu(X) = 1$.

Proof. "⇒": By point (v) of Remark 6.3 we can assume without loss of generality that $\dim(\operatorname{supp}\nu) = d$. Suppose that a dual optimizer ψ_{opt} exists. Then $\tilde{D}(\mu,\nu) - \mathcal{D}(\psi_{\mathrm{opt}}) = 0$ and from (3.4), (3.13) we obtain

$$\int \Big(\varphi^{\psi_{\mathrm{opt}}}(x) - \Big(\int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) \Big) \Big)\, \mu(dx) = 0.$$

Hence, for μ-a.e. $x \in \mathbb{R}^d$, the infimum in

$$\varphi^{\psi_{\mathrm{opt}}}(x) = \inf_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int \psi_{\mathrm{opt}}\, dp - \mathrm{MCov}(p,\gamma) \Big)$$

must be attained by $\pi_x^{\mathrm{SBM}} \in \mathscr{P}_2^x(\mathbb{R}^d)$. Furthermore, as $\mathcal{D}(\psi_{\mathrm{opt}})$ is finite, we deduce as in the proof of Lemma 3.5 that $\int \psi_{\mathrm{opt}}\, d\pi_x^{\mathrm{SBM}} < +\infty$, for μ-a.e. $x \in \mathbb{R}^d$. Thus also $\varphi^{\psi_{\mathrm{opt}}}(x)$ is finite, for μ-a.e. $x \in \mathbb{R}^d$. Recalling our assumption that $\dim(\operatorname{supp}\nu) = d$ as well as Lemma 6.2, we have $\mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$ by Definition 6.1. In particular, ψ_{opt} is co-finite by Lemma 5.1. All in all, we can apply part (ii) of Proposition 5.3 to ψ_{opt}, for μ-a.e. $x \in \mathbb{R}^d$. This yields that $\pi_x^{\mathrm{SBM}} = \nabla v(\zeta(x) + \cdot\,)(\gamma)$, with $v := \psi_{\mathrm{opt}}^*$ and $\zeta(x) \in \mathbb{R}^d$ as in (6.12). We now define $\alpha := \zeta(\mu)$. Since π^{SBM} has second marginal equal to ν, we obtain $\nabla v(\alpha * \gamma) = \nu$. On the other hand, as π_x^{SBM} has barycenter x, we have

$$\int \nabla v(\zeta(x) + z)\, d\gamma(z) = (\nabla v * \gamma)(\zeta(x)) = x. \tag{6.14}$$

We denote the set of points $x \in \mathbb{R}^d$ for which the identity (6.14) holds by X and note that $X \subseteq \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$ as well as $\mu(X) = 1$. Since $\alpha = \zeta(\mu)$, we conclude from (6.14) that $(\nabla v * \gamma)(\alpha) = \mu$. By (6.1), this establishes the existence of a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν, which by construction satisfies $\mathrm{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$, and connects $\psi_{\mathrm{opt}}, v, \alpha, \pi^{\mathrm{SBM}}$ as claimed.

It remains to check the identity (6.13). Since $\pi_x^{\mathrm{SBM}} \in \mathscr{P}_2^x(\mathbb{R}^d)$, and as we already established (6.11), we conclude that

$$\int \big|\nabla v(\zeta(x) + z)\big|^2\, d\gamma(z) < +\infty, \tag{6.15}$$

for every $x \in X$. In particular, we can apply part (iii) of Lemma 6.4 to the function $f = v$. Then from (6.14) we derive (6.13).

"⇐": Conversely, suppose that a Bass martingale $M = (M_t)_{0\leqslant t\leqslant 1}$ from μ to ν exists. By points (iv), (v) of Remark 6.3, we can assume without loss of generality that $\dim(\operatorname{supp}\nu) = d$ after passing to the minimal representation of the Bass martingale. We denote by $v\colon \mathbb{R}^d \to \mathbb{R}$ the associated convex function and define the proper, lower semicontinuous convex function $\psi_{\mathrm{opt}} := v^*$. Since $(\partial v)(\mathbb{R}^d) \subseteq \operatorname{dom} v^*$ and

$$\operatorname{int}\operatorname{conv}\big((\partial v)(\mathbb{R}^d)\big) = \operatorname{int}\overline{\operatorname{conv}}\big((\partial v)(\mathbb{R}^d)\big) = \operatorname{int}\widehat{\operatorname{supp}}(\nu) \neq \varnothing,$$

we also have that $\operatorname{int}(\operatorname{dom} v^*) \neq \varnothing$. In the following we show that ψ_{opt} is indeed a dual optimizer in the sense of Definition 6.1.

First, we verify that $\varphi^{\psi_{\mathrm{opt}}}(x) > -\infty$, for all $x \in \mathbb{R}^d$. As a consequence, we will see that the identity (6.13) holds for all $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$. Note that from (6.1) we have the relation $\nabla v(\alpha * \gamma) = \nu$. Since ν has finite second moment, we conclude that

$$\int |\nabla v(\zeta + z)|^2 \, d\gamma(z) < +\infty, \tag{6.16}$$

for α-a.e. $\zeta \in \mathbb{R}^d$. In particular, by analogy with (4.6), we obtain that

$$(v * \gamma)(\zeta) = \int v(\zeta + z) \, d\gamma(z) \in (-\infty, +\infty), \tag{6.17}$$

for α-a.e. $\zeta \in \mathbb{R}^d$. Recalling the second equality in (5.6), we have

$$\varphi^{\psi_{\mathrm{opt}}}(x) \geqslant (\varphi^{\psi_{\mathrm{opt}}})^{**}(x) = \sup_{\zeta \in \mathbb{R}^d} \Big(\langle \zeta, x \rangle - \int \psi^*_{\mathrm{opt}}(\zeta + z) \, d\gamma(z) \Big), \tag{6.18}$$

and by (6.17), the right-hand side of (6.18) is greater than $-\infty$, for every $x \in \mathbb{R}^d$. In particular, we can apply Proposition 5.2 to the function ψ_{opt}. As a result, for every $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$, we obtain an equality in (6.18) and the right-hand side admits a unique maximizer $\zeta(x) \in \mathbb{R}^d$. Differentiating under the integral sign, which is justified by part (i) of Lemma 6.4 applied to the function $f = v = \psi^*_{\mathrm{opt}}$, the first order condition for the optimality of $\zeta(x)$ reads

$$(\nabla \psi^*_{\mathrm{opt}} * \gamma)(\zeta(x)) = x. \tag{6.19}$$

Now we use part (iii) of Lemma 6.4 and obtain the identity (6.13) from (6.19) above, for all $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$.

Next, we show that $\alpha = \zeta(\mu)$ and $\mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$. Recalling from (6.1) that $(\nabla v * \gamma)(\alpha) = \mu$, and using (6.13), we get $\alpha = (\nabla v * \gamma)^{-1}(\mu) = \zeta(\mu)$. For the second claim, we again use (6.13) and obtain

$$1 = \alpha(\mathbb{R}^d) = \mu\big((\nabla v * \gamma)(\mathbb{R}^d)\big) = \mu(\operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})),$$

where the last equality follows from Lemma 6.5.

It remains to check that $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\mathrm{opt}})$ and $\operatorname{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$. For each $x \in \operatorname{int}(\operatorname{dom}\psi_{\mathrm{opt}})$, we define a probability measure $\pi^v_x := \nabla v(\zeta(x) + \,\cdot\,)(\gamma)$. By (6.16) and (6.19), π^v_x is an element of $\mathscr{P}^x_2(\mathbb{R}^d)$, for μ-a.e. $x \in \mathbb{R}^d$. Hence we can apply part (i) of Proposition 5.3, showing that the infimum

$$\varphi^{\psi_{\mathrm{opt}}}(x) = \inf_{p \in \mathscr{P}^x_2(\mathbb{R}^d)} \Big(\int \psi_{\mathrm{opt}} \, dp - \mathrm{MCov}(p, \gamma) \Big)$$

is attained by π^v_x. Therefore, from the representation (3.13) of the dual function $\mathcal{D}(\,\cdot\,)$ and the fact that $\pi^v_x(dy)\,\mu(dx) \in \mathsf{MT}(\mu, \nu)$, we obtain

$$\mathcal{D}(\psi_{\mathrm{opt}}) = \int \Big(\int \psi_{\mathrm{opt}}(y) \, \pi^v_x(dy) - \varphi^{\psi_{\mathrm{opt}}}(x) \Big) \, \mu(dx) = \int \mathrm{MCov}(\pi^v_x, \gamma) \, \mu(dx). \tag{6.20}$$

On the other hand, from (3.1), Theorem 3.1 and (3.14) we have

$$\int \mathrm{MCov}(\pi_x^v, \gamma)\, \mu(dx) \leqslant \tilde{P}(\mu, \nu) = \tilde{D}(\mu, \nu) \leqslant \mathcal{D}(\psi_{\mathrm{opt}}),$$

which in light of (6.20) implies that $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\mathrm{opt}})$. This completes the proof that ψ_{opt} is a dual optimizer. We still have to show that $\mathrm{Law}(M_0, M_1)$ equals the optimizer π^{SBM} of (3.1). Since M is a Bass martingale with associated convex function v, it follows from the identities (6.1) that $\pi_x^v = \mathrm{Law}(M_1 \,|\, M_0 = x)$. But then from (6.20) we conclude that $\pi_x^v = \pi_x^{\mathrm{SBM}}$. In particular, this shows (6.11). □

We now give a one-dimensional example in which a dual optimizer ψ_{opt} in the sense of Definition 6.1 is not integrable with respect to ν. Example 6.7 below illustrates that the form (3.6) of the dual function $\mathcal{D}(\,\cdot\,)$ may fail to make sense when we allow ψ to range more generally than in $C_{\mathrm{q}}(\mathbb{R}^d)$. On the other hand, the dual function written in the form (3.13) makes perfect sense for general convex functions $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$, which are μ-a.s. finite (recall Lemma 3.5). This leads to the satisfactory characterization of dual attainment as given by Theorem 6.6.

Example 6.7. Let $\alpha := \sum_{n=1}^{\infty} 2^{-n} \delta_{z_n} \in \mathscr{P}(\mathbb{R})$, with the sequence $(z_n)_{n \geqslant 1} \subseteq \mathbb{R}$ satisfying $\lim_{n\to\infty} \frac{|z_n|}{2^n} = +\infty$. Consider the convex function $v(z) := z \arctan z - \frac{1}{2}\log(1 + z^2)$, with the derivative $v'(z) = \arctan z$ being a strictly increasing, continuous and bounded function. Define $\mu := (v' * \gamma)(\alpha)$, so that $\mu = \sum_{n=1}^{\infty} 2^{-n} \delta_{x_n}$ for some bounded sequence $(x_n)_{n \geqslant 1} \subseteq \mathbb{R}$, and let $\nu := v'(\alpha * \gamma)$. Note that μ and ν are supported by $(-\frac{\pi}{2}, \frac{\pi}{2})$. Then, according to (6.1), the pair (v, α) defines a Bass martingale from μ to ν. Theorem 6.6 shows that the derivative of the dual optimizer is given by

$$\tfrac{d}{dy}\psi_{\mathrm{opt}}(y) = (v')^{-1}(y) = \tan y.$$

By definition of ν and α, this function is not ν-integrable, and neither is its antiderivative ψ_{opt}. ◊

We finish this chapter by deducing a subtle trajectorial property of Bass martingales from Lemma 6.5 and Theorem 6.6.

Corollary 6.8. *Let $M = (M_t)_{0 \leqslant t \leqslant 1}$ be a Bass martingale with $M_1 \sim \nu$. Define τ as the stopping time*

$$\tau := \inf\big\{t \in [0,1] \colon M_t \notin \mathrm{ri}\,\widehat{\mathrm{supp}}(\nu)\big\} \wedge 1,$$

i.e., the minimum of 1 and the first time that M reaches the boundary of $\widehat{\mathrm{supp}}(\nu)$. Then $\tau = 1$ a.s.

Proof. As in the proof of Theorem 6.6 we assume without loss of generality that $\mathrm{supp}\,\nu$ is d-dimensional, so that $\mathrm{int}(\mathrm{dom}\, v^*) \neq \varnothing$. By (6.2), the Bass martingale is given by $M_t = (\nabla v * \gamma^{1-t})(B_t)$, where $\nabla v(\gamma) = \nu$, for some finite-valued convex function v. We now define the proper, lower semicontinuous convex function $\psi := v^*$. As in the proof of the second implication of Theorem 6.6, we see that $\varphi^{\psi}(x) > -\infty$, for all $x \in \mathbb{R}^d$,

and $\nabla\psi^* \in L^2(\gamma_{B_0}; \mathbb{R}^d)$ a.s. Therefore we can apply Lemma 6.5 and from (6.6), (6.7) we obtain for all $t \in [0,1)$ the a.s. equality

$$\operatorname{int}(\operatorname{dom}\psi) = \operatorname{int}\operatorname{conv}\left((\partial\psi^*)(\mathbb{R}^d)\right) = (\nabla\psi^* * \gamma_{B_0}^{1-t})(\mathbb{R}^d).$$

Since

$$\operatorname{int}\operatorname{conv}\left((\partial\psi^*)(\mathbb{R}^d)\right) = \operatorname{int}\overline{\operatorname{conv}}\left((\partial\psi^*)(\mathbb{R}^d)\right) = \operatorname{int}\widehat{\operatorname{supp}}(\nu)$$

we conclude for all $t \in [0,1)$ that

$$\operatorname{int}\widehat{\operatorname{supp}}(\nu) = (\nabla\psi^* * \gamma_{B_0}^{1-t})(\mathbb{R}^d)$$

a.s. Hence a.s. M_t lies in $\operatorname{int}\widehat{\operatorname{supp}}(\nu)$ as long as $t \in [0,1)$. Thus if M ever reaches the boundary of $\widehat{\operatorname{supp}}(\nu)$, this will happen at time $t = 1$, so $\tau = 1$. Otherwise, i.e., if M never reaches the boundary of $\widehat{\operatorname{supp}}(\nu)$, then by definition $\tau = 1$, too. □

7. Irreducibility and existence of dual optimizers

We fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. The following notation from [DMT19] will be used throughout this chapter.

Definition 7.1. We denote by $C := \widehat{\mathrm{supp}}(\nu)$ the closed convex hull of the support of ν and by $I := \operatorname{ri} C$ its relative interior.

As in the proofs of Theorem 6.6 and Corollary 6.8, we can assume without loss of generality that the support of ν has dimension d, so that I is open in $\mathbb{R}^d$; see also Remark 6.3, (iv) and (v).

Definition 7.2. We say that the pair (μ, ν) is *De March–Touzi irreducible*, if the irreducible convex paving of De March–Touzi [DMT19] consists of the single irreducible component I. By [DMT19, Theorem 2.1], this condition means that

(i) the set $I = \operatorname{ri} C$, with $C = \widehat{\mathrm{supp}}(\nu)$, satisfies $\mu(I) = 1$,

(ii) there exists some martingale transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu, \nu)$ with the property that $C = \widehat{\mathrm{supp}}(\pi_x^{\mathrm{DMT}})$, for μ-a.e. $x \in \mathbb{R}^d$. We refer to π^{DMT} as a *De March–Touzi transport*.

The main assumption of this chapter (with the exception of Lemma 7.9, Lemma 7.19, Lemma 7.20, Proposition 7.21 and Corollary 7.22) is the following.

Assumption 7.3. The pair (μ, ν) is De March–Touzi irreducible.

Assumption 7.3 is equivalent to the irreducibility of the pair (μ, ν) in the sense of Definition 1.2. For a proof of this result and for further equivalent characterizations of irreducibility we recall Theorem 1.4.

In Lemma 7.5 below we give a more quantitative description of the defining property of a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu, \nu)$ between an irreducible pair (μ, ν).

Definition 7.4. Let $K \subseteq I$ be compact. For an element y^* of the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$ we define the slice S_{y^*} of C beyond K by

$$S_{y^*} := \{y \in C \colon \langle y, y^* \rangle > \sup\{\langle \tilde{y}, y^* \rangle \colon \tilde{y} \in K\}\}. \tag{7.1}$$

Lemma 7.5. *Under Assumption 7.3, for μ-a.e. $x \in \mathbb{R}^d$ and for every compact set $K \subseteq I$, there exists a constant $\delta(K, x) > 0$ such that $\pi_x^{\mathrm{DMT}}(S_{y^*}) \geqslant \delta(K, x)$ for every $y^* \in \mathbb{S}^{d-1}$.*

Proof. We fix $x \in \mathbb{R}^d$ with $C = \widehat{\operatorname{supp}}(\pi_x^{\mathrm{DMT}})$ and note that the map

$$r_x \colon \mathbb{S}^{d-1} \longrightarrow [0,1] \colon y^* \longmapsto r_x(y^*) \coloneqq \pi_x^{\mathrm{DMT}}(S_{y^*})$$

is lower semicontinuous. As the set C equals the closed convex hull of the support of π_x^{DMT}, the map r_x is also strictly positive. Indeed, if there was some $y^* \in \mathbb{S}^{d-1}$ with $r_x(y^*) = 0$, then the closed convex set $C \setminus S_{y^*}$ would support π_x^{DMT}, which is in contradiction to $C \setminus S_{y^*} \subsetneq C = \widehat{\operatorname{supp}}(\pi_x^{\mathrm{DMT}})$. Since $\mathbb{S}^{d-1}$ is compact we conclude that r_x attains its minimum $\delta(K,x) \coloneqq \min_{y^* \in \mathbb{S}^{d-1}} r_x(y^*) > 0$. □

The main result of this chapter is the following.

Theorem 7.6. *Under Assumption 7.3, there exists a dual optimizer ψ_{opt} in the sense of Definition 6.1.*

Together with Theorem 6.6, this has the following consequence for the primal optimizer $\pi^{\mathrm{SBM}} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (3.1).

Corollary 7.7. *Under Assumption 7.3, there exists a Bass martingale $(M_t)_{0 \leqslant t \leqslant 1}$ from μ to ν. Moreover, $\operatorname{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$ and, for μ-a.e. $x \in \mathbb{R}^d$, the measure π_x^{SBM} is equivalent to ν. In particular, π^{SBM} is a De March–Touzi transport in the sense of Definition 7.2, (ii).*

Proof. Admitting Theorem 7.6, there exists a dual optimizer ψ_{opt}. Thus, by Theorem 6.6, there is a Bass martingale from μ to ν with the property that $\operatorname{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$. In particular, by (6.11), there is a measurable set A with $\mu(A) = 1$, such that π_x^{SBM} and $\pi_{x'}^{\mathrm{SBM}}$ are image measures of suitably translated Gaussians under the same function ∇v, for all $x, x' \in A$. Hence $\pi_x^{\mathrm{SBM}} \sim \pi_{x'}^{\mathrm{SBM}}$, for all $x, x' \in A$, and this in turn implies the equivalence of $\nu(dy) = \int \pi_x^{\mathrm{SBM}}(dy)\, \mu(dx)$ with $\pi_{x'}^{\mathrm{SBM}}(dy)$, for each $x' \in A$. In particular, $\operatorname{supp}(\pi_x^{\mathrm{SBM}}) = \operatorname{supp}(\nu)$, for μ-a.e. $x \in \mathbb{R}^d$, which implies Definition 7.2, (ii) for π^{SBM}. □

7.1. Proof of Theorem 7.6

The proof is split into two main steps.

7.1.1. Step 1 of the proof of Theorem 7.6

We first construct a *convergent* optimizing sequence of convex functions $(\psi_n)_{n \geqslant 1}$ in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ (see Remark 7.11 below) for the dual problem (3.2).

Proposition 7.8. *Under Assumption 7.3, there is an optimizing sequence $(\psi_n)_{n \geqslant 1}$ of nonnegative convex functions in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ for the dual problem (3.2), which converges compactly on I (i.e., uniformly on compact subsets $K \subseteq I$) to some convex function $\psi_{\lim} \colon I \to [0, +\infty)$.*

Before we turn to the proof of Proposition 7.8, we still need some preparation. The following auxiliary result does not require the irreducibility Assumption 7.3, but solely relies on the finiteness of the value $\tilde{D}(\mu,\nu)$ of the dual problem (3.2).

Lemma 7.9. *Let $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ be an optimizing sequence for the dual problem (3.2) and take any $\pi \in \mathsf{MT}(\mu,\nu)$. Then*

$$\sup_{n\geqslant 1} \int\int (\psi_n(y) - \psi_n(x))\, \pi_x(dy)\, \mu(dx) < +\infty. \tag{7.2}$$

Proof. Recalling (3.3), we consider for $n \geqslant 1$ the functions

$$\mathbb{R}^d \ni x \longmapsto \varphi^{\psi_n}(x) = \inf_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int \psi_n\, dp - \mathrm{MCov}(p,\gamma)\Big). \tag{7.3}$$

By taking $p = \delta_x$ in (7.3) we obtain the trivial estimate $\varphi^{\psi_n}(x) \leqslant \psi_n(x)$ and consequently

$$\int\int (\psi_n(y) - \psi_n(x))\, \pi_x(dy)\, \mu(dx) \leqslant \int\int (\psi_n(y) - \varphi^{\psi_n}(x))\, \pi_x(dy)\, \mu(dx).$$

By definition (3.13) of the dual function $\mathcal{D}(\,\cdot\,)$ we have

$$\mathcal{D}(\psi_n) = \int \Big(\int \psi_n(y)\, \pi_x(dy) - \varphi^{\psi_n}(x)\Big)\, \mu(dx),$$

and as the sequence of real numbers $(\mathcal{D}(\psi_n))_{n\geqslant 1}$ converges to the finite number $\tilde{D}(\mu,\nu)$, we conclude (7.2). □

Definition 7.10. We fix an arbitrary optimizing sequence $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ of *convex* functions for the dual problem (3.2), which is possible thanks to Proposition 3.3. For $x \in \mathbb{R}^d$, we define

$$\psi_n^x(\,\cdot\,) \coloneqq \psi_n(\,\cdot\,) - \psi_n(x) - \langle \partial\psi_n(x), \,\cdot\, - x\rangle, \tag{7.4}$$

so that $\psi_n^x(x) = 0$ and $\psi_n^x(\,\cdot\,)$ takes values in $[0,+\infty)$, for every $n \geqslant 1$. Here — by abuse of notation — $\partial\psi_n(x)$ denotes a subgradient of ψ_n at x, i.e., an arbitrary element of the subdifferential of ψ_n at x.

Remark 7.11. We observe that passing from $\psi \in C_{\mathrm{q}}(\mathbb{R}^d)$ to $\psi + \mathrm{aff}$, where $\mathrm{aff}\colon \mathbb{R}^d \to \mathbb{R}$ is an arbitrary affine function, does not change the value of the dual function (3.13), i.e., $\mathcal{D}(\psi) = \mathcal{D}(\psi + \mathrm{aff})$. In particular, we note that the sequence $(\psi_n^x)_{n\geqslant 1}$ defined in (7.4) is contained in the class of functions

$$C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d) \coloneqq \{\psi(\,\cdot\,) + \mathrm{aff}(\,\cdot\,)\colon\ \psi \in C_{\mathrm{q}}(\mathbb{R}^d),\ \mathrm{aff}\colon \mathbb{R}^d \to \mathbb{R} \text{ affine}\},$$

and hence $\mathcal{D}(\psi_n^x) = \mathcal{D}(\psi_n)$. ◇

Our goal is to show that — under the irreducibility Assumption 7.3 and after passing to Cesàro means of a suitable subsequence — the sequence $(\psi_n^x)_{n\geqslant 1}$ of (7.4) is bounded on compact subsets of I, for μ-a.e. $x \in I$. This is the content of the following lemma.

Lemma 7.12. *Under Assumption 7.3, for any given optimizing sequence $(\psi_n)_{n\geqslant 1}$ of convex functions in $C_{\mathrm{q}}(\mathbb{R}^d)$, there is a sequence of Cesàro means of a subsequence, still denoted by $(\psi_n)_{n\geqslant 1}$, such that, for μ-a.e. $x \in I$ and for every compact set $K \subseteq I$, we have that $\sup_{n\geqslant 1, y\in K} \psi_n^x(y) < +\infty$, for every choice of $(\psi_n^x)_{n\geqslant 1} \subseteq C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ as in (7.4).*

Proof. We rely on [DMT19] and fix a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu,\nu)$ satisfying point (ii) of Definition 7.2, which exists by the irreducibility Assumption 7.3. Let $(\psi_n)_{n\geqslant 1}$ and $(\psi_n^x)_{n\geqslant 1}$, measurable selected for $x \in \mathbb{R}^d$, be as in Definition 7.10 above. By Lemma 7.9, the sequence

$$\Psi_n(x) \coloneqq \int \psi_n^x(y)\,\pi_x^{\mathrm{DMT}}(dy) = \int (\psi_n(y) - \psi_n(x))\,\pi_x^{\mathrm{DMT}}(dy)$$

is bounded in $L^1(\mu)$. Applying Komlós' theorem [Kom67, Theorem 1], we can find a subsequence $(n_k)_{k\geqslant 1} \subseteq \mathbb{N}$, such that the Cesàro means

$$\overline{\Psi}_k \coloneqq \frac{\Psi_{n_1} + \Psi_{n_2} + \ldots + \Psi_{n_k}}{k}$$

converge μ-a.s. to some random variable $\Psi \in L^1(\mu)$. If we define the Cesàro means

$$\bar{\psi}_k^x \coloneqq \frac{\psi_{n_1}^x + \psi_{n_2}^x + \ldots + \psi_{n_k}^x}{k},$$

we have

$$\overline{\Psi}_k(x) = \int \bar{\psi}_k^x(y)\,\pi_x^{\mathrm{DMT}}(dy), \qquad x \in \mathbb{R}^d.$$

Note that passing to a subsequence $(n_k)_{k\geqslant 1} \subseteq \mathbb{N}$ and forming Cesàro means $(\bar{\psi}_k^x)$ preserves the property of being an optimizing sequence of convex functions in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ of the form (7.4). Therefore we can replace the original sequences (Ψ_n) and (ψ_n^x) by $(\overline{\Psi}_k)$ and $(\bar{\psi}_k^x)$, respectively, and may again relabel them as (Ψ_n) and (ψ_n^x), respectively. With that said, as a consequence of the μ-a.s. convergence to a finite limit, we have that

$$m(x) \coloneqq \sup_{n\geqslant 1} \Psi_n(x) = \sup_{n\geqslant 1} \int \psi_n^x(y)\,\pi_x^{\mathrm{DMT}}(dy) < +\infty, \tag{7.5}$$

for μ-a.e. $x \in \mathbb{R}^d$.

Arguing by contradiction to the statement of Lemma 7.12, we assume that there is a compact set $K_0 \subseteq I$ such that the set

$$A_1 \coloneqq \Big\{x \in \mathbb{R}^d \colon \sup_{n\geqslant 1, y\in K_0} \psi_n^x(y) = +\infty\Big\}$$

has positive μ-measure. Furthermore, by Lemma 7.5, the set

$$A_2 \coloneqq \Big\{x \in \mathbb{R}^d \colon\ \exists\, \delta(K_0,x) > 0 \text{ such that } \pi_x^{\mathrm{DMT}}(S_{y^*}) \geqslant \delta(K_0,x) \text{ for every } y^* \in \mathbb{S}^{d-1}\Big\}$$

has full μ-measure. By (7.5) above, also the set

$$A_3 \coloneqq \{x \in \mathbb{R}^d \colon m(x) < +\infty\}$$

has full μ-measure, so that the intersection $A \coloneqq A_1 \cap A_2 \cap A_3$ has positive μ-measure. Pick a point $x_0 \in A$. As $x_0 \in A_1$, for arbitrarily large $M > m(x_0)/\delta(K_0,x_0)$, we can find $n_0 \geqslant 1$ and $y_0 \in K_0$ such that $\psi_{n_0}^{x_0}(y_0) \geqslant M$. The function

$$\ell(y) \coloneqq \langle \partial\psi_{n_0}^{x_0}(y_0), y - y_0\rangle + M, \qquad y \in \mathbb{R}^d,$$

satisfies $\ell \leqslant \psi_{n_0}^{x_0}$ and $\ell(y) \geqslant M$, for all $y \in S_{y^*}$, with $y^* \coloneqq \frac{z}{|z|}$ and $z \coloneqq \partial\psi_{n_0}^{x_0}(y_0)$. We conclude with

$$m(x_0) \geqslant \int \psi_n^x(y)\, \pi_{x_0}^{\mathrm{DMT}}(dy) \geqslant \int \ell^+(y)\, \pi_{x_0}^{\mathrm{DMT}}(dy) \geqslant M\, \pi_{x_0}^{\mathrm{DMT}}(S_{y^*}) \geqslant M\, \delta(K_0, x_0),$$

which is the desired contradiction, as M is arbitrarily large. □

Proof of Proposition 7.8. By Lemma 7.12 we can fix some $x_0 \in I$ such that the sequence of convex functions $(\psi_n^{x_0})_{n\geqslant 1}$ is bounded on all compact subsets $K \subseteq I$. In particular, the sequence $(\psi_n^{x_0})_{n\geqslant 1}$ is pointwise bounded on I. By [Roc70, Theorem 10.9] we can select a subsequence, still denoted by $(\psi_n^{x_0})_{n\geqslant 1}$, which converges uniformly on compact subsets $K \subseteq I$ to some convex function $\psi_{\lim}^{x_0} \colon I \to [0, +\infty)$. Dropping the superscript x_0 to ease notation, we arrive at a sequence $(\psi_n)_{n\geqslant 1}$ with limit $\psi_{\lim}$ as required in the statement of Proposition 7.8. □

7.1.2. Step 2 of the proof of Theorem 7.6

We extend the function $\psi_{\lim} \colon I \to [0, +\infty)$ of Proposition 7.8 to a lower semicontinuous convex function $\psi_{\lim} \colon \mathbb{R}^d \to [0, +\infty]$ which is equal to $+\infty$ on $\mathbb{R}^d \backslash C$. Since $I \subseteq \operatorname{dom} \psi_{\lim}$, I is open and $\mu(I) = 1$, it follows that $\mu(\operatorname{int}(\operatorname{dom} \psi_{\lim})) = 1$.

Proposition 7.13. *Under Assumption 7.3, the convex function $\psi_{\lim} \colon \mathbb{R}^d \to [0, +\infty]$ is a dual optimizer in the sense of Definition 6.1, i.e., satisfies $\tilde{D}(\mu, \nu) = \mathcal{D}(\psi_{\lim})$.*

Definition 7.14. For μ-a.e. $x \in \mathbb{R}^d$, we denote by v_{SBM}^x the Brenier map from γ to π_x^{SBM}, so that $\nabla v_{\mathrm{SBM}}^x(\gamma) = \pi_x^{\mathrm{SBM}}$ and

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \langle \nabla v_{\mathrm{SBM}}^x(z), z \rangle\, \gamma(dz);$$

we write $\psi_{\mathrm{SBM}}^x \coloneqq (v_{\mathrm{SBM}}^x)^*$ for its convex conjugate.

Remark 7.15. As γ has full support, the convex function v_{SBM}^x is finite-valued and continuous everywhere on $\mathbb{R}^d$. In particular, v_{SBM}^x is unique, up to an additive constant. Therefore also its convex conjugate $\psi_{\mathrm{SBM}}^x \colon \mathbb{R}^d \to (-\infty, +\infty]$ is unique, up to an additive constant. Also note that ψ_{SBM}^x is finite-valued on I and takes the value $+\infty$ on $\mathbb{R}^d \setminus C$, with $C = \widehat{\operatorname{supp}}(\nu)$ and $I = \operatorname{int} C$. ◇

Our goal is to show that $\psi_{\lim} \colon \mathbb{R}^d \to [0, +\infty]$ is a dual optimizer. For this purpose, we will prove that, for μ-a.e. $x \in I$, the function $\psi_{\lim}(\,\cdot\,)$ equals $\psi_{\mathrm{SBM}}^x(\,\cdot\,)$, modulo adding an affine function; in short, $\psi_{\lim} \equiv \psi_{\mathrm{SBM}}^x \bmod (\mathrm{aff})$, for μ-a.e. $x \in \mathbb{R}^d$. In Lemma 7.18 below we will show that the set

$$A \coloneqq \{x \in I \colon \psi_{\lim} \not\equiv \psi_{\mathrm{SBM}}^x \bmod (\mathrm{aff})\} \tag{7.6}$$

indeed has μ-measure zero. It will then follow from Lemma 7.19 that $\psi_{\lim}$ is actually a dual optimizer. First, we need some auxiliary results.

Lemma 7.16. *Under Assumption 7.3, for μ-a.e. $x \in A$, there exists a probability measure $\check{\pi}_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ supported by C and there is a constant $\tilde{\beta}(x) > 0$ such that*

$$\mathrm{MCov}(\check{\pi}_x, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \check{\pi}_x) \geqslant \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) + \tilde{\beta}(x). \tag{7.7}$$

Proof. Recalling the representation (3.13) of the dual function $\mathcal{D}(\,\cdot\,)$ and the definition (3.3) of the function $x \mapsto \varphi^{\psi}(x)$, we consider for μ-a.e. $x \in I$ the functions

$$\begin{aligned} \phi^{\psi^x_{\mathrm{SBM}}}(x) &\coloneqq \int \psi^x_{\mathrm{SBM}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi^x_{\mathrm{SBM}}}(x) \\ &= \sup_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) + \int \psi^x_{\mathrm{SBM}}\, d(\pi_x^{\mathrm{SBM}} - p) \Big) \end{aligned} \tag{7.8}$$

and

$$\begin{aligned} \phi^{\psi_{\lim}}(x) &\coloneqq \int \psi_{\lim}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_{\lim}}(x) \\ &= \sup_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - p) \Big). \end{aligned} \tag{7.9}$$

Recalling Definition 7.14, we have $\nabla(\psi^x_{\mathrm{SBM}})^*(\gamma) = \pi_x^{\mathrm{SBM}} \in \mathscr{P}_2^x(\mathbb{R}^d)$. Therefore we can apply part (i) of Proposition 5.3, which yields that the supremum in (7.8) is attained by π_x^{SBM}, so that $\phi^{\psi^x_{\mathrm{SBM}}}(x) = \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)$. By taking $p = \pi_x^{\mathrm{SBM}}$ in (7.9), we obtain the inequality $\phi^{\psi_{\lim}}(x) \geqslant \phi^{\psi^x_{\mathrm{SBM}}}(x)$, for μ-a.e. $x \in I$.

Now we define the sets

$$B \coloneqq \{x \in I \colon \phi^{\psi_{\lim}}(x) > \phi^{\psi^x_{\mathrm{SBM}}}(x)\}, \qquad X \coloneqq \{x \in I \colon \phi^{\psi^x_{\mathrm{SBM}}}(x) < +\infty\}$$

and claim that for the set A defined in (7.6) we have the relation $\tilde{A} \coloneqq A \cap X = B$. In other words, since $\mu(X) = 1$, the sets A and B are equal, up to a set of μ-measure zero.

Indeed, if $x \notin \tilde{A}$, then $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff) or $\phi^{\psi^x_{\mathrm{SBM}}}(x) = +\infty$, so that in both cases $\phi^{\psi_{\lim}}(x) = \phi^{\psi^x_{\mathrm{SBM}}}(x)$. Conversely, if $x \notin B$, then $x \notin X$ or we have the equality $\phi^{\psi_{\lim}}(x) = \phi^{\psi^x_{\mathrm{SBM}}}(x)$ of real numbers and thus also the supremum in the definition (7.9) of $\phi^{\psi_{\lim}}(x)$ is attained by

$$\pi_x^{\mathrm{SBM}} = \nabla(\psi^x_{\mathrm{SBM}})^*(\gamma) \in \mathscr{P}_2^x(\mathbb{R}^d). \tag{7.10}$$

Hence we can apply part (ii) of Proposition 5.3 to the lower semicontinuous convex function $\psi_{\lim} \colon \mathbb{R}^d \to [0, +\infty]$ satisfying $\mu(\mathrm{int}(\mathrm{dom}\, \psi_{\lim})) = 1$, which tells us that the optimizer (7.10) of the supremum in (7.9) is equal to

$$\nabla \psi^*_{\lim}(\gamma_{\zeta(x)}),$$

for some $\zeta(x) \in \mathbb{R}^d$. We conclude the γ-a.s. equality

$$\nabla(\psi^x_{\mathrm{SBM}})^* = \nabla \psi^*_{\lim}(\,\cdot\, + \zeta(x)),$$

which is equivalent to the γ-a.s. equality

$$(\psi^x_{\mathrm{SBM}})^* = \psi^*_{\lim}(\,\cdot\, + \zeta(x)) + c, \tag{7.11}$$

for some constant $c \in \mathbb{R}$. Since the convex function $(\psi^x_{\mathrm{SBM}})^*$ is finite-valued and continuous everywhere on $\mathbb{R}^d$, the equality (7.11) has to hold everywhere on $\mathbb{R}^d$. This implies that $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff) and we deduce that $x \notin A$. Altogether, we have seen that $x \notin B$ implies that $x \notin \tilde{A}$.

As a consequence of $A = B$, up to a set of μ-measure zero, for μ-a.e. $x \in A$ we have $\phi^{\psi_{\lim}}(x) > \phi^{\psi^x_{\mathrm{SBM}}}(x)$, so that we can measurably select some $\check{\pi}_x \in \mathscr{P}^x_2(\mathbb{R}^d)$ with

$$\mathrm{MCov}(\check{\pi}_x, \gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - \check{\pi}_x) > \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma), \tag{7.12}$$

which gives (7.7). As the right-hand side of (7.12) is finite, for μ-a.e. $x \in I$, and since $\psi_{\lim}(y) = +\infty$ for $y \in \mathbb{R}^d \setminus C$, we see that $\check{\pi}_x$ is supported by C. □

Our next step is to modify the measures $\{\check{\pi}_x\}_{x\in A}$ of Lemma 7.16, so that they have compact support and still satisfy (7.7) for some $\beta(x) > 0$ instead of $\tilde{\beta}(x)$. To this end, we choose an increasing sequence $(K_j)_{j\geqslant 1}$ of compact subsets of I such that $\bigcup_{j\geqslant 1} K_j = I$. Denoting by M^x the Bass martingale from δ_x to $\check{\pi}_x$, and by τ^x_j the first exit time of M^x from K_j (similarly as in Corollary 6.8), we define $\check{\pi}^j_x := \mathrm{Law}(M^x_{\tau^x_j \wedge 1})$. By optional sampling, $\check{\pi}^j_x \in \mathscr{P}^x_2(\mathbb{R}^d)$ and, by definition, $\check{\pi}^j_x$ is supported by the compact set K_j.

Lemma 7.17. *Under Assumption 7.3, for μ-a.e. $x \in A$, there exists $\check{\pi}^{j(x)}_x \in \mathscr{P}^x_2(\mathbb{R}^d)$ supported by $K_{j(x)}$ for some $j(x) \in \mathbb{N}$ and there is a constant $\beta(x) > 0$ such that*

$$\mathrm{MCov}(\check{\pi}^{j(x)}_x, \gamma) + \int \psi_{\lim}\, d(\pi^{\mathrm{SBM}}_x - \check{\pi}^{j(x)}_x) \geqslant \mathrm{MCov}(\pi^{\mathrm{SBM}}_x, \gamma) + \beta(x). \tag{7.13}$$

Proof. From Lemma 7.16 we already have the inequality (7.7). In order to derive (7.13), we have to show that

$$\lim_{j\to+\infty} \mathrm{MCov}(\check{\pi}^j_x, \gamma) = \mathrm{MCov}(\check{\pi}_x, \gamma) \tag{7.14}$$

and

$$\limsup_{j\to+\infty} \int \psi_{\lim}(y)\, d\check{\pi}^j_x(y) \leqslant \int \psi_{\lim}(y)\, d\check{\pi}_x(y). \tag{7.15}$$

We begin with the proof of (7.14). First, observe that $M^x_{\tau^x_j\wedge 1} \to M^x_{\tau^x\wedge 1}$ in L^2, where τ^x is the first exit time of M^x from I. Since $\tau^x \wedge 1 = 1$ a.s. by Corollary 6.8, we conclude that $M^x_{\tau^x_j\wedge 1} \to M^x_1$ in L^2. Consequently,

$$\mathcal{W}_2^2(\check{\pi}^j_x, \check{\pi}_x) := \inf_{q\in \mathsf{Cpl}(\check{\pi}^j_x, \check{\pi}_x)} \int |x_1 - x_2|^2\, q(dx_1, dx_2)$$

converges to zero as $j \to +\infty$, and the inequality

$$|\mathrm{MCov}(\check{\pi}^j_x, \gamma) - \mathrm{MCov}(\check{\pi}_x, \gamma)| \leqslant \mathcal{W}_2(\check{\pi}^j_x, \check{\pi}_x)\,\sqrt{d}$$

yields (7.14).

Finally, we show (7.15). Note that by optional sampling, $\check{\pi}_x^j \in \mathscr{P}_2^x(\mathbb{R}^d)$. Since a martingale composed with a convex function is a submartingale, it follows that

$$\forall j \geqslant 1\colon \int \psi_{\lim}\, d\check{\pi}_x^j \leqslant \int \psi_{\lim}\, d\check{\pi}_x$$

and we obtain (7.15). □

Lemma 7.18. *Under Assumption 7.3, the set $A \subseteq \mathbb{R}^d$ defined in (7.6) has μ-measure zero.*

Proof. We assume for contradiction that $\mu(A) > 0$. Thanks to Lemma 7.17, the set A is μ-a.s. equal to the union

$$\bigcup_{\substack{j\in\mathbb{N},\\ \beta\in\mathbb{Q}_+\setminus\{0\}}} \left\{x \in A\colon \mathrm{MCov}(\check{\pi}_x^j, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \check{\pi}_x^j) \geqslant \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) + \beta\right\}.$$

Hence we can find a subset $B \subseteq A$ with $\mu(B) > 0$, such that (7.13) holds with uniform constants $j \in \mathbb{N}$ and $\beta > 0$ for all $x \in B$, i.e.,

$$\exists j\ \exists \beta\ \forall x \in B\colon \quad \mathrm{MCov}(\check{\pi}_x^j, \gamma) + \int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \check{\pi}_x^j) \geqslant \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) + \beta. \tag{7.16}$$

Now we define a measurable collection of probability measures $\{\tilde{\pi}_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ with $\mathrm{bary}(\tilde{\pi}_x) = x$ by

$$\tilde{\pi}_x := \begin{cases} \check{\pi}_x^j, & x \in B, \\ \pi_x^{\mathrm{SBM}}, & x \in \mathbb{R}^d \setminus B. \end{cases} \tag{7.17}$$

From the inequality (7.16) and the definition (7.17) we deduce that

$$\int\int \psi_{\lim}\, d(\pi_x^{\mathrm{SBM}} - \tilde{\pi}_x)\, d\mu(x) + \int \mathrm{MCov}(\tilde{\pi}_x, \gamma)\, d\mu(x) \tag{7.18}$$

$$\geqslant \int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, d\mu(x) + \beta\, \mu(B) \tag{7.19}$$

$$= \tilde{P}(\mu, \nu) + \beta\, \mu(B). \tag{7.20}$$

Recall from Proposition 7.8 that there is an optimizing sequence $(\psi_n)_{n\geqslant 1}$ of convex functions $\psi_n\colon \mathbb{R}^d \to [0, +\infty)$ for the dual problem (3.2), which converges uniformly on compact subsets $K \subseteq I$ to $\psi_{\lim}$. To find the desired contradiction, we want to replace $\psi_{\lim}$ in (7.18) by $\lim_{n\to\infty} \psi_n$ and write the limit outside of the integral. This is clearly no problem if $x \notin B$. For $x \in B$, note that the measure $\tilde{\pi}_x = \check{\pi}_x^j$ is supported by the compact set K_j. As $(\psi_n)_{n\geqslant 1}$ converges to $\psi_{\lim}$ uniformly on K_j, we have

$$\lim_{n\to\infty} \int_B \int \psi_n(y)\, d\tilde{\pi}_x(y)\, d\mu(x) = \int_B \int \psi_{\lim}(y)\, d\tilde{\pi}_x(y)\, d\mu(x). \tag{7.21}$$

On the other hand, Fatou's lemma gives

$$\liminf_{n\to\infty} \int_B \int \psi_n(y)\, d\pi_x^{\mathrm{SBM}}(y)\, d\mu(x) \geqslant \int_B \int \psi_{\lim}(y)\, d\pi_x^{\mathrm{SBM}}(y)\, d\mu(x). \tag{7.22}$$

Now combining (7.18) – (7.20) with (7.21), (7.22) yields the inequality

$$\liminf_{n\to\infty} \int \int \psi_n\, d(\pi_x^{\mathrm{SBM}} - \tilde{\pi}_x)\, d\mu(x) + \int \mathrm{MCov}(\tilde{\pi}_x, \gamma)\, d\mu(x) \geqslant \tilde{P}(\mu,\nu) + \beta\, \mu(B).$$

Recalling the definition (3.3) of the function $x \mapsto \varphi^{\psi}(x)$ and the dual function (3.13), we observe that the left-hand side of this inequality is less than or equal to

$$\liminf_{n\to\infty} \mathcal{D}(\psi_n) = \liminf_{n\to\infty} \int \Big(\int \psi_n(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_n}(x) \Big)\, \mu(dx).$$

Since $(\psi_n)_{n\geqslant 1}$ is an optimizing sequence for the dual problem, it follows that

$$\tilde{D}(\mu,\nu) = \lim_{n\to\infty} \mathcal{D}(\psi_n) = \liminf_{n\to\infty} \mathcal{D}(\psi_n) \geqslant \tilde{P}(\mu,\nu) + \beta\, \mu(B) > \tilde{P}(\mu,\nu).$$

But this is a contradiction to the fact that there is no duality gap by Theorem 3.1. □

Lemma 7.19. *A lower semicontinuous convex function $\psi_{\mathrm{opt}} \colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\mathrm{int}(\mathrm{dom}\, \psi_{\mathrm{opt}})) = 1$ is a dual optimizer in the sense of Definition 6.1 if and only if*

$$\mu\big(\{x \in \mathbb{R}^d \colon \psi_{\mathrm{opt}} \equiv \psi_{\mathrm{SBM}}^x \bmod (\mathrm{aff})\}\big) = 1. \tag{7.23}$$

Proof. "⇐": As we have seen in the proof of Lemma 7.16, for μ-a.e. $x \in \mathbb{R}^d$, the supremum in (7.8) is attained by $p = \pi_x^{\mathrm{SBM}}$. Hence integrating with respect to $\mu(dx)$ yields

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, \mu(dx) = \int \Big(\int \psi_{\mathrm{SBM}}^x(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_{\mathrm{SBM}}^x}(x) \Big)\, \mu(dx).$$

By assumption (7.23) it follows that

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, \mu(dx) = \int \Big(\int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - \varphi^{\psi_{\mathrm{opt}}}(x) \Big)\, \mu(dx). \tag{7.24}$$

Recalling the representation (3.13) of the dual function $\mathcal{D}(\,\cdot\,)$, we see that the expression on the right-hand side of (7.24) equals $\mathcal{D}(\psi_{\mathrm{opt}})$. On the other hand, the left-hand side is equal to $\tilde{P}(\mu,\nu) = \tilde{D}(\mu,\nu)$ by Theorem 3.1, so that ψ_{opt} is a dual optimizer.

"⇒": Let ψ_{opt} be a dual optimizer. As in the proof of the implication "⇒" in Theorem 6.6 we conclude that

$$\pi_x^{\mathrm{SBM}} = \nabla \psi_{\mathrm{opt}}^*(\zeta(x) + \,\cdot\,)(\gamma),$$

for μ-a.e. $x \in \mathbb{R}^d$, with $\zeta(x) \in \mathbb{R}^d$ as in (6.12). On the other hand, by Definition 7.14 we have $\pi_x^{\mathrm{SBM}} = \nabla(\psi_{\mathrm{SBM}}^x)^*(\gamma)$, for μ-a.e. $x \in \mathbb{R}^d$. Arguing as in the proof of Lemma 7.16 we obtain (7.23). □

Let us summarize why the above arguments complete the proof of Proposition 7.13 and thus also of Theorem 7.6.

Proof of Proposition 7.13. According to Lemma 7.18, the function $\psi_{\lim}\colon \mathbb{R}^d \to [0,+\infty]$ equals ψ^x_{SBM} mod (aff), for μ-a.e. $x \in I$. By Lemma 7.19, this implies that $\psi_{\lim}$ is a dual optimizer. □

Proof of Theorem 7.6. The existence of a dual optimizer follows from Proposition 7.8 and Proposition 7.13. □

Let us have one more look at the structure of the proof of Theorem 7.6 and Corollary 7.7 above. We started with an arbitrary optimizing sequence $(\psi_n)_{n\geqslant 1} \subseteq C_{\mathrm{q}}(\mathbb{R}^d)$ of convex functions for the dual problem (3.2), which we normalized to obtain $(\psi^x_n)_{n\geqslant 1}$ as in (7.4). Then we showed in Proposition 7.8 that under the irreducibility Assumption 7.3, which guarantees the existence of a De March–Touzi transport $\pi^{\mathrm{DMT}} \in \mathsf{MT}(\mu,\nu)$, we could find a limiting function $\psi_{\lim}$. However, this was only possible *after passing to a subsequence, forming Cesàro means, and then choosing a further subsequence of* $(\psi^x_n)_{n\geqslant 1}$. In Proposition 7.13 we argued that $\psi_{\lim}$ is a dual optimizer and hence, by Theorem 6.6, there is a Bass martingale from μ to ν.

In the follow-up paper [ST24], the general case of a pair (μ,ν) which is not necessarily irreducible — and thus does not necessarily satisfy Assumption 7.3 — is treated. In this case, a variant of this line of reasoning is needed. Suppose we *already know* that an optimizing sequence $(\psi_n)_{n\geqslant 1}$ is pointwise bounded on a relatively open convex set $I \subseteq \mathbb{R}^d$ with $\mu(I) = 1$ and such that $\widehat{\mathrm{supp}}(\nu)$ is contained in the closure of I. Under this assumption — *but without imposing irreducibility on the pair* (μ,ν) — we will prove in Proposition 7.21 below that there is no need of passing to a subsequence or of forming convex combinations of the optimizing sequence $(\psi_n)_{n\geqslant 1}$. Rather already the original sequence $(\psi_n)_{n\geqslant 1}$ — modulo adding affine functions — converges.

Lemma 7.20. *For every sequence $(\psi_n)_{n\geqslant 1}$ of lower semicontinuous convex functions $\psi_n\colon \mathbb{R}^d \to [0,+\infty)$ there is a closed convex set $C_{\lim} \subseteq \mathbb{R}^d$ with relative interior $I_{\lim}$ and a subsequence $(\psi_{n_k})_{k\geqslant 1}$ such that the limits*

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) < +\infty, \qquad y \in I_{\lim} \tag{7.25}$$

$$\psi_{\lim}(y) := \lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C_{\lim} \tag{7.26}$$

exist. Moreover, the convergence in (7.25) is uniform on compact subsets of $I_{\lim}$.

Proof. Step 1. We denote by $\mathcal{N}$ the collection of increasing subsequences $\mathcal{N} = (n_k)_{k\geqslant 1}$ of $\mathbb{N}$. For two subsequences $\mathcal{N}, \mathcal{N}' \in \mathcal{N}$ we call $\mathcal{N}'$ *finer* than $\mathcal{N}$, denoted by $\mathcal{N}' \succcurlyeq \mathcal{N}$, if $\mathcal{N}' \subseteq \mathcal{N}$, up to finitely many elements.

We write $\mathcal{I}$ for the collection of relatively open convex sets $I \subseteq \mathbb{R}^d$. For two pairs $(I,\mathcal{N}), (I',\mathcal{N}') \in \mathcal{I}\times\mathcal{N}$ we say that $(I',\mathcal{N}')$ is *finer* than $(I,\mathcal{N})$, again denoted by $(I',\mathcal{N}') \succcurlyeq (I,\mathcal{N})$, if $I' \supsetneq I$ and $\mathcal{N}' \succcurlyeq \mathcal{N}$.

We call a pair $(I, \mathscr{N}) \in \mathcal{I} \times \mathcal{N}$ *admissible* if it satisfies

$$\forall y \in I\colon \quad \sup_{n_k \in \mathscr{N}} \psi_{n_k}(y) < +\infty. \tag{7.27}$$

Step 2. Let $(I, \mathscr{N})$ be admissible. Write $\mathscr{N} = (n_k)_{k\geqslant 1} \in \mathcal{N}$ and denote by C the closure of I. Then one of the following two alternatives hold: either

$$\lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C, \tag{7.28}$$

in which case we call $(I, \mathscr{N})$ *maximal*; or there is some $y_0 \in \mathbb{R}^d \setminus C$ and a subsequence $\mathscr{N}' = (n'_k)_{k\geqslant 1} \in \mathcal{N}$ finer than $\mathscr{N}$ such that

$$\sup_{n'_k \in \mathscr{N}'} \psi_{n'_k}(y_0) < +\infty.$$

Now we define the relatively open convex set $I' \coloneqq \operatorname{ri}\operatorname{conv}(y_0, I)$ and note that $I' \supsetneq I$. Then the pair $(I', \mathscr{N}')$ is admissible and finer than $(I, \mathscr{N})$. In this second case we say that $(I, \mathscr{N})$ is *refinable*.

Step 3. Assume that $(I, \mathscr{N})$ is maximal, so that both (7.27) and (7.28) are satisfied. In particular, the sequence $(\psi_{n_k})_{k\geqslant 1}$, with $\mathscr{N} = (n_k)_{k\geqslant 1}$, is pointwise bounded on I. By [Roc70, Theorem 10.9] we can select a further subsequence, still denoted by $(\psi_{n_k})_{k\geqslant 1}$, such that (7.25) holds for all $y \in I_{\lim} \coloneqq I$ and the convergence is uniform on compact subsets of $I_{\lim}$. Furthermore, condition (7.26) is still satisfied for this further subsequence thanks to (7.28).

Step 4. In light of Step 3, in order to prove Lemma 7.20, it suffices to show the existence of a maximal pair $(I, \mathscr{N})$, which we will denote by $(I_{\lim}, \mathscr{N}_{\lim})$. The construction of the maximal pair $(I_{\lim}, \mathscr{N}_{\lim})$ is done via transfinite recursion along the countable ordinal numbers $\alpha < \omega_1$, where ω_1 is the first uncountable ordinal.

We start the construction as follows. If $\sup_{n\in\mathbb{N}} \psi_n(y) = +\infty$ holds for all $y \in \mathbb{R}^d$, the pair $(I_0, \mathscr{N}_0) = (\varnothing, \mathbb{N})$ is admissible. Otherwise, we can find some $y_0 \in \mathbb{R}^d$ such that $\sup_{n\in\mathbb{N}} \psi_n(y_0) < +\infty$ and then the pair $(I_0, \mathscr{N}_0) = (\{y_0\}, \mathbb{N})$ is admissible.

Now let $\alpha < \omega_1$ be an ordinal number and suppose that the transfinite family $(I_\beta, \mathscr{N}_\beta)_{\beta<\alpha}$ in $\mathcal{I} \times \mathcal{N}$ is such that

(i) the pair $(I_\beta, \mathscr{N}_\beta)$ is admissible, for all $\beta < \alpha$,

(ii) the pair $(I_{\beta_2}, \mathscr{N}_{\beta_2})$ is finer than $(I_{\beta_1}, \mathscr{N}_{\beta_1})$, for all $\beta_1 < \beta_2 < \alpha$.

We first assume that α is a successor ordinal, i.e., there is an ordinal β such that $\alpha = \beta + 1$. Since $(I_\beta, \mathscr{N}_\beta)$ is admissible, by Step 2 only two cases are possible. If $(I_\beta, \mathscr{N}_\beta)$ is maximal, we set $(I_{\lim}, \mathscr{N}_{\lim}) \coloneqq (I_\beta, \mathscr{N}_\beta)$ and finish the construction. If $(I_\beta, \mathscr{N}_\beta)$ is refinable, we can find a pair $(I_\alpha, \mathscr{N}_\alpha)$ which is finer than $(I_\beta, \mathscr{N}_\beta)$.

Suppose that α is a limit ordinal, i.e., it can be written as $\alpha = \{\beta\colon \beta < \alpha\}$. Since $\alpha < \omega_1$, the ordinal α is countable and can be enumerated as $\alpha = \{\alpha_k\colon k < \omega\}$, where ω

is the first infinite ordinal. Letting $\beta_0 \coloneqq 0$ and defining inductively $\beta_{m+1} \coloneqq \alpha_\ell$, where $\ell \coloneqq \min\{k\colon \beta_m < \alpha_k\}$, we obtain a strictly increasing cofinal sequence $(\beta_m)_{m<\omega}$ in α. Now we define $I_\alpha \coloneqq \bigcup_{m<\omega} I_{\beta_m}$. By (ii), $(I_{\beta_m})_{m<\omega}$ is a strictly increasing sequence of relatively open convex sets, therefore also $I_\alpha \in \mathcal{I}$. Again by (ii), the sequence $(\mathscr{N}_{\beta_m})_{m<\omega}$ satisfies $\mathscr{N}_{\beta_{m_1}} \succcurlyeq \mathscr{N}_{\beta_{m_2}}$, for all $m_1 < m_2 < \omega$. Thus we can find a diagonal subsequence $\mathscr{N}_\alpha \in \mathcal{N}$ of $(\mathscr{N}_{\beta_m})_{m<\omega}$, meaning that $\mathscr{N}_\alpha \succcurlyeq \mathscr{N}_{\beta_m}$, for all $m < \omega$. Since by (i) the pair $(I_{\beta_m}, \mathscr{N}_{\beta_m})$ is admissible, for all $m < \omega$, by construction also $(I_\alpha, \mathscr{N}_\alpha)$ is admissible. Furthermore, $(I_\alpha, \mathscr{N}_\alpha)$ satisfies $(I_\alpha, \mathscr{N}_\alpha) \succcurlyeq (I_{\beta_m}, \mathscr{N}_{\beta_m})$, for all $m < \omega$. Since $(\beta_m)_{m<\omega}$ is cofinal, it follows that $(I_\alpha, \mathscr{N}_\alpha) \succcurlyeq (I_\beta, \mathscr{N}_\beta)$, for all $\beta < \alpha$.

After countably many steps, this transfinite recursion must stop, which happens when $\alpha < \omega_1$ is such that $(I_\alpha, \mathscr{N}_\alpha)$ is maximal and we can then set $(I_{\lim}, \mathscr{N}_{\lim}) \coloneqq (I_\alpha, \mathscr{N}_\alpha)$. Indeed, otherwise we would have a transfinite sequence $(I_\beta)_{\beta<\omega_1}$ of strictly increasing relatively open convex sets, indexed by the countable ordinals, which is impossible. □

Proposition 7.21. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Let $I \subseteq \mathbb{R}^d$ be a relatively open convex set with $\mu(I) = 1$. Denote by C the closure of I and assume that $\widehat{\operatorname{supp}}(\nu) \subseteq C$. Let $(\psi_n)_{n\geqslant 1}$ be an optimizing sequence of nonnegative convex functions in $C_{\mathrm{q}}^{\mathrm{aff}}(\mathbb{R}^d)$ for the dual problem (3.2) such that*

$$\forall y \in I\colon \quad \sup_{n\geqslant 1} \psi_n(y) < +\infty. \tag{7.29}$$

Then there is a lower semicontinuous convex function $\psi_{\lim}\colon \mathbb{R}^d \to [0, +\infty]$ and a sequence $(\tilde{\psi}_n)_{n\geqslant 1}$ such that $\psi_n \equiv \tilde{\psi}_n$ mod (aff), for each $n \geqslant 1$, and

$$\forall y \in I\colon \quad \psi_{\lim}(y) = \lim_{n\to\infty} \tilde{\psi}_n(y) < +\infty, \tag{7.30}$$

$$\forall y \in \mathbb{R}^d \setminus C\colon \quad \psi_{\lim}(y) = \lim_{n\to\infty} \tilde{\psi}_n(y) = +\infty. \tag{7.31}$$

The convergence in (7.30) is uniform on compact subsets of I. Moreover, we have that $C = \widehat{\operatorname{supp}}(\nu)$ and $\psi_{\lim}$ is a dual optimizer, which is unique modulo adding affine functions.

Proof. Step 1. By Lemma 7.20, there is a closed convex set $C_{\lim} \subseteq \mathbb{R}^d$ with relative interior $I_{\lim}$ and a subsequence $(\psi_{n_k})_{k\geqslant 1}$ such that the limits

$$\psi_{\lim}(y) \coloneqq \lim_{k\to\infty} \psi_{n_k}(y) < +\infty, \qquad y \in I_{\lim} \tag{7.32}$$

$$\psi_{\lim}(y) \coloneqq \lim_{k\to\infty} \psi_{n_k}(y) = +\infty, \qquad y \in \mathbb{R}^d \setminus C_{\lim} \tag{7.33}$$

exist, where the convergence in (7.32) is uniform on compact subsets of $I_{\lim}$. Via (7.32) and (7.33) we define a lower semicontinuous convex function $\psi_{\lim}\colon \mathbb{R}^d \to [0, +\infty]$ on all of $\mathbb{R}^d$.

Step 2. By the hypothesis (7.29) and the construction of $I_{\lim}$, we clearly have $I \subseteq I_{\lim}$. Our goal is to show that $I_{\lim} = I$. To this end, we define the set $A \subseteq \mathbb{R}^d$ as in (7.6), now with I replaced by $I_{\lim}$, i.e.,

$$A = \{x \in I_{\lim}\colon \psi_{\lim} \not\equiv \psi^x_{\mathrm{SBM}} \text{ mod (aff)}\}.$$

Again, we will show that $\mu(A) = 0$. Repeating the reasoning of Lemma 7.16 we obtain, for μ-a.e. $x \in A$, measures $\check{\pi}_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ satisfying (7.7), which now are supported by $C_{\lim}$. Then we choose an increasing sequence $(K_j)_{j\geqslant 1}$ of relatively compact subsets of $I_{\lim}$ such that $\bigcup_{j\geqslant 1} K_j = I_{\lim}$. As in Lemma 7.17, for μ-a.e. $x \in A$, we find measures $\check{\pi}_x^{j(x)} \in \mathscr{P}_2^x(\mathbb{R}^d)$ satisfying (7.13) and supported by $K_{j(x)}$, for some $j(x) \in \mathbb{N}$. Arguing as in Lemma 7.18 we conclude that $\mu(A) = 0$, i.e., $\psi_{\lim} \equiv \psi_{\mathrm{SBM}}^x \bmod (\mathrm{aff})$, for μ-a.e. $x \in I_{\lim}$. In particular, since $I_{\lim} = \operatorname{ri}(\operatorname{dom}\psi_{\lim})$ and $\operatorname{ri}(\operatorname{dom}\psi_{\mathrm{SBM}}^x) = \operatorname{ri}\widehat{\operatorname{supp}}(\nu)$, we conclude from the assumption $\widehat{\operatorname{supp}}(\nu) \subseteq C$ that $I_{\lim} = I$ and consequently $C = \widehat{\operatorname{supp}}(\nu)$. From Lemma 7.19, we conclude that the limiting function $\psi_{\lim}$ is a dual optimizer. Together with Step 1, this proves the statement of Proposition 7.21 for the subsequence $(\psi_{n_k})_{k\geqslant 1}$.

Step 3. Applying Step 1 and Step 2 to an arbitrary subsequence $(\psi_{m_\ell})_{\ell\geqslant 1}$ instead of the original sequence $(\psi_n)_{n\geqslant 1}$, we obtain a further subsequence $(\psi_{m_{\ell_j}})_{j\geqslant 1}$ and a lower semicontinuous convex function $\tilde{\psi}_{\lim}\colon \mathbb{R}^d \to [0, +\infty]$ such that

$$\begin{aligned} \forall y \in I\colon\ & \tilde{\psi}_{\lim}(y) = \lim_{j\to\infty} \psi_{m_{\ell_j}}(y) < +\infty, \\ \forall y \in \mathbb{R}^d \setminus C\colon\ & \tilde{\psi}_{\lim}(y) = \lim_{j\to\infty} \psi_{m_{\ell_j}}(y) = +\infty, \end{aligned}$$

and $\tilde{\psi}_{\lim} \equiv \psi_{\mathrm{SBM}}^x \bmod (\mathrm{aff})$, for μ-a.e. $x \in I$. We conclude that there is an affine function aff depending on the subsequence $(m_{\ell_j})_{j\geqslant 1}$ such that $\psi_{\lim} = \tilde{\psi}_{\lim} + \mathrm{aff}$. In particular,

$$\forall y \in I \cup (\mathbb{R}^d \setminus C)\colon\ \lim_{j\to\infty} \bigl(\psi_{m_{\ell_j}}(y) + \mathrm{aff}(y)\bigr) = \psi_{\lim}(y). \tag{7.34}$$

Step 4. Let $x_0, \dots, x_d$ be affinely independent points in $I \cup (\mathbb{R}^d \setminus C)$. For each $n \geqslant 1$ there is a unique affine function aff_n such that

$$\forall i \in \{0, \dots, d\}\colon\ \psi_n(x_i) + \mathrm{aff}_n(x_i) = \psi_{\lim}(x_i). \tag{7.35}$$

We claim that

$$\forall y \in I \cup (\mathbb{R}^d \setminus C)\colon\ \lim_{n\to\infty} \bigl(\psi_n(y) + \mathrm{aff}_n(y)\bigr) = \psi_{\lim}(y). \tag{7.36}$$

Indeed, if this were not the case, there would be some $y_0 \in I \cup (\mathbb{R}^d \setminus C)$ and a subsequence $(m_\ell)_{\ell\geqslant 1}$ such that

$$\lim_{\ell\to\infty} \bigl(\psi_{m_\ell}(y_0) + \mathrm{aff}_{m_\ell}(y_0)\bigr) \neq \psi_{\lim}(y_0). \tag{7.37}$$

By Step 3 there is a further subsequence $(m_{\ell_j})_{j\geqslant 1}$ and an affine function aff such that (7.34) holds. In particular,

$$\forall i \in \{0, \dots, d\}\colon\ \lim_{j\to\infty} \bigl(\psi_{m_{\ell_j}}(x_i) + \mathrm{aff}(x_i)\bigr) = \psi_{\lim}(x_i), \tag{7.38}$$

$$\lim_{j\to\infty} \bigl(\psi_{m_{\ell_j}}(y_0) + \mathrm{aff}(y_0)\bigr) = \psi_{\lim}(y_0). \tag{7.39}$$

From (7.35) and (7.38) we obtain that

$$\forall i \in \{0, \dots, d\}\colon \quad \lim_{j\to\infty} \operatorname{aff}_{m_{\ell_j}}(x_i) = \operatorname{aff}(x_i). \tag{7.40}$$

Since the points $x_0, \dots, x_d$ are affinely independent, the convergence in (7.40) holds for every $y \in I \cup (\mathbb{R}^d \setminus C)$. Then (7.39) leads to a contradiction to (7.37), which proves claim (7.36). Defining $\tilde{\psi}_n := \psi_n + \operatorname{aff}_n$, for $n \geqslant 1$, we derive from (7.36) the limiting assertions (7.30) and (7.31). Finally, by [Roc70, Theorem 10.8] the convergence in (7.30) is uniform on compact subsets of I and we have already seen in Step 2 that $\psi_{\lim} \equiv \psi^x_{\mathrm{SBM}}$ mod (aff), for μ-a.e. $x \in I$. □

Corollary 7.22. *Under the assumptions of Proposition 7.21, there exists a Bass martingale $(M_t)_{0\leqslant t\leqslant 1}$ from μ to ν. Moreover, $\operatorname{Law}(M_0, M_1) = \pi^{\mathrm{SBM}}$ and, for μ-a.e. $x \in I$, the measure π^{SBM}_x is equivalent to ν. In particular, π^{SBM} is a De March–Touzi transport and the pair (μ, ν) is irreducible.*

Proof. By Proposition 7.21, the limiting function $\psi_{\lim}$ is a dual optimizer. Therefore, by analogy with the deduction of Corollary 7.7 from Theorem 7.6, we deduce Corollary 7.22 from Proposition 7.21. To see that the pair (μ, ν) is irreducible we refer to Theorem 1.4. □

7.2. Proof of Theorem 1.3

We are now in the position to prove our first main result of the introduction.

Proof of Theorem 1.3. The implication "(2) ⇒ (1)" is Theorem 1.10 of [BBHK20]. For the proof of "(1) ⇒ (2)" we apply Corollary 7.7 and obtain the existence of a Bass martingale from μ to ν. By the uniqueness results of Theorem 2.2 in [BBHK20], this Bass martingale has to agree with the given stretched Brownian motion. □

7.3. Proof of the second part of Theorem 1.7

Finally, we complete the proof of Theorem 1.7. Recalling the results of Chapter 4, we can formulate the second part of Theorem 1.7 equivalently as follows.

Proposition 7.23. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. The value*

$$D_{\mathrm{rel}}(\mu, \nu) = \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \mathcal{E}(\psi) \tag{7.41}$$

is attained by a lower semicontinuous convex function $\psi_{\mathrm{opt}}\colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\operatorname{ri}(\operatorname{dom}\psi_{\mathrm{opt}})) = 1$ if and only if (μ, ν) is irreducible. In this case the (unique) optimizer to (MBB) is given by the Bass martingale

$$M_t := \mathbb{E}[\nabla v(B_1) \,|\, \sigma(B_s\colon s \leqslant t)] = \mathbb{E}[\nabla v(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1,$$

where $v = \psi^*_{\mathrm{opt}}$ and $B_0 \sim \nabla(\psi^*_{\mathrm{opt}} * \gamma)^*(\mu)$.

Proof. We call a function ψ_{opt} as in the statement of the proposition a dual optimizer of (7.41) and first show that this notion is equivalent to a dual optimizer in the sense of Definition 6.1. Indeed, if ψ_{opt} is a dual optimizer of (7.41), it follows from the inequality $(\psi^* * \gamma)^* = (\varphi^\psi)^{**} \leqslant \varphi^\psi$ that ψ_{opt} is also a dual optimizer according to Definition 6.1. Conversely, suppose that ψ_{opt} is an optimizer in the latter sense. Then it follows from Lemma 7.19 that (7.23) is satisfied. Note that by the classical Kantorovich duality we have

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \psi^x_{\mathrm{SBM}}\, d\pi_x^{\mathrm{SBM}} + \int (\psi^x_{\mathrm{SBM}})^*\, d\gamma,$$

for μ-a.e. $x \in \mathbb{R}^d$. Thus, by (7.23), for μ-a.e. $x \in \mathbb{R}^d$ there exists $\zeta(x) \in \mathbb{R}^d$ such that

$$\mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma) = \int \big(\psi_{\mathrm{opt}}(\,\cdot\,) - \langle \zeta(x), \cdot\,\rangle\big)\, d\pi_x^{\mathrm{SBM}} + \int \big(\psi_{\mathrm{opt}}(\,\cdot\,) - \langle \zeta(x), \cdot\,\rangle\big)^*\, d\gamma.$$

Reading the proof of Proposition 4.2 backwards we conclude that

$$\int \mathrm{MCov}(\pi_x^{\mathrm{SBM}}, \gamma)\, d\mu(x) = \int \Big(\int \psi_{\mathrm{opt}}(y)\, \pi_x^{\mathrm{SBM}}(dy) - (\psi^*_{\mathrm{opt}} * \gamma)^*(x) \Big)\, \mu(dx),$$

i.e., $D_{\mathrm{rel}}(\mu, \nu) = \mathcal{E}(\psi_{\mathrm{opt}})$. This shows that ψ_{opt} is also a dual optimizer of (7.41).

Now that we know that both definitions of a dual optimizer are equivalent, we can conclude the assertions of Proposition 7.23 from the results we have already established. Indeed, by Theorem 6.6, the existence of a dual optimizer is equivalent to the existence of a Bass martingale $(M_t)_{0 \leqslant t \leqslant 1}$ in the given form. Finally, by Theorem 1.3 and Theorem 1.4, the existence of a Bass martingale is equivalent to the irreducibility of (μ, ν). □

Part VII.

The Bass functional of martingale transport

Joint work with Julio Backhoff-Veraguas and Walter Schachermayer

Abstract

An interesting question in the field of martingale optimal transport is to determine the martingale with prescribed initial and terminal marginals which is most correlated to Brownian motion. Under a necessary and sufficient irreducibility condition, the answer to this question is given by a *Bass martingale*. At an intuitive level, the latter can be imagined as an order-preserving and martingale-preserving space transformation of an underlying Brownian motion starting with an initial law α which is tuned to ensure the marginal constraints.

In this work we study how to determine the above-mentioned initial condition α. This is done by a careful study of what we dub the *Bass functional*. In our main result we show the equivalence between the existence of minimizers of the Bass functional and the existence of a Bass martingale with prescribed marginals. This complements the convex duality approach presented in Part VI with a purely variational perspective. We also establish an infinitesimal version of this result, and furthermore prove the displacement convexity of the Bass functional along certain generalized geodesics in the quadratic Wasserstein space.

1. Introduction

1.1. Martingale optimization problem

Let μ, ν be elements of $\mathscr{P}_2(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$ with finite second moments. Assume that μ, ν are in convex order, denoted by $\mu \preceq_{\mathrm{c}} \nu$, and meaning that $\int \phi \, d\mu \leqslant \int \phi \, d\nu$ holds for all convex functions $\phi \colon \mathbb{R}^d \to \mathbb{R}$. As in [BBHK20, BBST23] we consider the martingale optimization problem

$$MT(\mu,\nu) \coloneqq \inf_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s \, dB_s}} \mathbb{E}\Big[\int_0^1 |\sigma_t - I_d|^2_{\mathrm{HS}} \, dt\Big], \tag{MBB}$$

where B is Brownian motion on $\mathbb{R}^d$ and $|\cdot|_{\mathrm{HS}}$ denotes the Hilbert–Schmidt norm. The abbreviation "MBB" stands for "Martingale Benamou–Brenier" and this designation is motivated from the fact that (MBB) can be seen as a martingale counterpart of the classical formulation in optimal transport by Benamou–Brenier [BB99], see [BBHK20, BBST23]. The problem (MBB) is equivalent to maximizing the covariance between M and Brownian motion subject to the marginal conditions $M_0 \sim \mu$ and $M_1 \sim \nu$, to wit

$$P(\mu,\nu) \coloneqq \sup_{\substack{M_0\sim\mu,\, M_1\sim\nu,\\ M_t = M_0 + \int_0^t \sigma_s \, dB_s}} \mathbb{E}\Big[\int_0^1 \mathrm{tr}(\sigma_t) \, dt\Big]. \tag{1.1}$$

Both problems have the same optimizer and the values are related via

$$MT(\mu,\nu) = d + \int |y|^2 \, d\nu(y) - \int |x|^2 \, d\mu(x) - 2P(\mu,\nu).$$

As shown in [BBHK20], the problem (MBB) admits a strong Markov martingale $\hat{M}$ as a unique optimizer, which is called the *stretched Brownian motion* from μ to ν in [BBHK20].

1.2. Bass martingales and structure of stretched Brownian motion

Owing to the work [BBST23] it is known that the optimality property of stretched Brownian motion is related to a structural / geometric description. For its formulation we start with the following definition.

Definition 1.1. For probability measures μ, ν we say that the pair (μ, ν) is *irreducible* if for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A) > 0$ and $\nu(B) > 0$ there is a martingale $X = (X_t)_{0 \leqslant t \leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

We remark that in the classical theory of optimal transport one can always find couplings (X_0, X_1) of (μ, ν) such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$, for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A), \nu(B) > 0$; e.g., by letting (X_0, X_1) be independent. In martingale optimal transport this property may fail.

Next we recall the following concept from [Bas83, BBHK20, BBST23].

Definition 1.2. Let $B = (B_t)_{0 \leqslant t \leqslant 1}$ be Brownian motion on $\mathbb{R}^d$ with $B_0 \sim \hat{\alpha}$, where $\hat{\alpha}$ is an arbitrary element of $\mathscr{P}(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$. Let $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ be convex such that $\nabla \hat{v}(B_1)$ is square-integrable. We call

$$\hat{M}_t \coloneqq \mathbb{E}[\nabla \hat{v}(B_1) \,|\, \sigma(B_s \colon s \leqslant t)] = \mathbb{E}[\nabla \hat{v}(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1 \tag{1.2}$$

a *Bass martingale* with *Bass measure* $\hat{\alpha}$ joining $\mu = \mathrm{Law}(\hat{M}_0)$ with $\nu = \mathrm{Law}(\hat{M}_1)$.

The reason behind this terminology is that Bass [Bas83] used this construction (with $d = 1$ and $\hat{\alpha}$ a Dirac measure) in order to derive a solution of the Skorokhod embedding problem.

In [BBST23, Theorem 1.3] it is shown that under the irreducibility assumption on the pair (μ, ν) there is a unique Bass martingale $\hat{M}$ from μ to ν, i.e., satisfying $\hat{M}_0 \sim \mu$ and $\hat{M}_1 \sim \nu$:

Theorem 1.3. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and assume that (μ, ν) is irreducible. Then the following are equivalent for a martingale $\hat{M} = (\hat{M}_t)_{0 \leqslant t \leqslant 1}$ with $\hat{M}_0 \sim \mu$ and $\hat{M}_1 \sim \nu$:*

(1) $\hat{M}$ is stretched Brownian motion, i.e., the optimizer of (MBB).

(2) $\hat{M}$ is a Bass martingale.

Since, for probability measures $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$, stretched Brownian motion always exists by [BBHK20, Theorem 1.5], the above theorem states that the existence of a Bass martingale follows from — and is in fact equivalent to — the irreducibility assumption on the pair (μ, ν).

Denoting by $*$ the convolution operator and by γ the standard Gaussian measure on $\mathbb{R}^d$, we remark that the convex function $\hat{v}$ and the Bass measure $\hat{\alpha}$ from Definition 1.2 satisfy the identities

$$(\nabla \hat{v} * \gamma)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * \gamma) = \nu. \tag{1.3}$$

We formalize the fundamental relations (1.3) and their correspondence with Bass martingales in Lemma 2.3 below.

Throughout we write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and set $\hat{v}_t \coloneqq \hat{v} * \gamma^{1-t} \colon \mathbb{R}^d \to \mathbb{R}$, for $0 \leqslant t \leqslant 1$. In these terms, (1.2) amounts to

$$\hat{M}_t = \nabla \hat{v}_t(B_t), \qquad 0 \leqslant t \leqslant 1.$$

1.3. Main results

In the following we denote by MCov the *maximal covariance* between two probability measures $p_1, p_2 \in \mathscr{P}_2(\mathbb{R}^d)$, defined as

$$\mathrm{MCov}(p_1, p_2) := \sup_{q \in \mathsf{Cpl}(p_1, p_2)} \int \langle x_1, x_2 \rangle \, q(dx_1, dx_2), \tag{1.4}$$

where $\mathsf{Cpl}(\mu, \nu)$ is the set of all couplings $\pi \in \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ between μ and ν, i.e., probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal μ and second marginal ν. As is well known, maximizing the covariance between p_1 and p_2 is equivalent to minimizing their expected squared distance; see also (2.3) below.

Definition 1.4. We introduce the *Bass functional*

$$\mathscr{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha) := \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu). \tag{1.5}$$

In our first main result we derive a novel formulation of problem (1.1), which characterizes the Bass measure $\hat{\alpha}$ in (1.3) as the optimizer of the Bass functional (1.5).

Theorem 1.5. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Then*

$$P(\mu, \nu) = \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{1.6}$$

The right-hand side of (1.6) is attained by $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ if and only if there is a Bass martingale from μ to ν with Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$.

The proof of Theorem 1.5 is given in Chapter 3. In Chapter 4 we will show the following infinitesimal version of Theorem 1.5, which constitutes our second main result:

Theorem 1.6. *Let $(M_t)_{0 \leqslant t \leqslant 1}$ be an $\mathbb{R}^d$-valued martingale bounded in L^2, which is given by the stochastic integral*

$$M_t = M_0 + \int_0^t \sigma_s \, dB_s, \qquad 0 \leqslant t \leqslant 1,$$

where $(\sigma_t)_{0 \leqslant t \leqslant 1}$ is a progressively measurable process. Denote by μ_t the law of M_t. For Lebesgue-a.e. $0 \leqslant t \leqslant 1$ we have, for each $\alpha \in \mathscr{P}_2(\mathbb{R}^d)$, the inequality

$$\mathbb{E}[\mathrm{tr}(\sigma_t)] \leqslant \liminf_{h \to 0} \tfrac{1}{h} \Big(\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h}) - \mathrm{MCov}(\alpha, \mu_t) \Big). \tag{1.7}$$

We note that, for a Bass martingale $(\hat{M}_t)_{0 \leqslant t \leqslant 1}$ of the form

$$d\hat{M}_t = \hat{\sigma}_t(\hat{M}_t) \, dB_t,$$

with associated Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ and diffusion function $\hat{\sigma}_t \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$, we have, for Lebesgue-a.e. $0 \leqslant t \leqslant 1$, the equality

$$\mathbb{E}[\mathrm{tr}(\hat{\sigma}_t(\hat{M}_t))] = \frac{d}{dt} \mathrm{MCov}(\hat{\alpha} * \gamma^t, \hat{\mu}_t),$$

where $\hat{\mu}_t = \mathrm{Law}(\hat{M}_t)$. This exhibits the sharpness of (1.7) and shows that Theorem 1.6 is an infinitesimal analogue of Theorem 1.5.

In our final main result we discuss convexity properties of the Bass functional

$$\mathscr{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha)$$

defined in (1.5).

Theorem 1.7. *We have the following results:*

(1) If $d = 1$, then $\mathcal{V}$ is displacement convex, i.e., convex along the geodesics given by McCann interpolations [McC97].

(2) If $d \geqslant 1$, then $\mathcal{V}$ is displacement convex along generalized geodesics with base μ.

The proof of this result, together with a discussion on the various forms of convexity stated therein (see e.g. [McC97, Vil03, AGS08]), and a treatment of the strict convexity of $\mathcal{V}$, are deferred to Chapter 5. We merely stress here that the Bass functional fails to be convex, and can even be concave, if we consider convex combinations of measures in the usual linear sense.

1.4. Related literature

Optimal transport (see the monographs [Vil03, Vil09, AG13, San15] for introductions to the field) has its origins in the works of Monge [Mon81] and Kantorovich [Kan42]. The modern theory is based on the seminal results of Benamou, Brenier, and McCann [Bre87, Bre91, McC94, McC95, BB99].

In martingale optimal transport problems (see e.g. [HN12, BHLP13, TT13, DS14, GHLT14, CLM17]), the transport plan satisfies an additional martingale constraint. This additional requirement arises naturally in finance (e.g. [BHLP13]), but has also remarkable consequences for the study of martingale inequalities (e.g. [BN15, OST15, HLOST16]) and the Skorokhod embedding problem (e.g. [BCH17, KTT17, BNS22]). The analysis of irreducibility of a pair of marginals (μ, ν) was initiated by Beiglböck and Juillet [BJ16] in dimension one and extended in the works [OS17, DMT19, GKL19] to multiple dimensions.

Martingale optimal transport problems in continuous time have attracted a lot of attention in recent years, see e.g. [DS14, BHLT17, COT19, GKP19, GLW19, GKL20, CKPS21, GL21]. In this work we concern ourselves with the specific structure given by the martingale Benamou–Brenier problem, introduced in [BBHK20] in probabilistic language and in [HT19] in PDE language, and subsequently studied through the point of view of duality theory in [BBST23]. The same kind of problem appeared independently in the context of market impact in financial markets in an article by Loeper [Loe18].

It was also shown in [BBHK20] that the optimizer $\hat{M}$ of the problem (MBB) is the process whose evolution follows the movement of Brownian motion as closely as possible with respect to an *adapted Wasserstein distance* (see e.g. [BBBE20, Föl22]) subject to the given marginal constraints.

2. Preliminaries

In this short chapter we give a more detailed review of some of the main results in [BBST23], which will be useful for the coming discussions and proofs.

2.1. Dual viewpoint

As established in [BBST23, Theorem 1.4], the problem (1.1) admits a dual formulation with a particularly appealing structure:

Theorem 2.1. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$. The value $P(\mu,\nu)$ of the problem (1.1) is equal to*

$$D(\mu,\nu) := \inf_{\substack{\psi \in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big). \tag{2.1}$$

The infimum is uniquely attained (modulo adding affine functions) by a lower semi-continuous convex function $\hat{\psi}\colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\mathrm{ri}(\mathrm{dom}\,\hat{\psi})) = 1$ if and only if (μ,ν) is irreducible. In this case, the (unique) optimizer to (MBB), (1.1) is given by the Bass martingale with associated convex function $\hat{v} = \hat{\psi}^*$ and Bass measure $\hat{\alpha} = \nabla(\hat{\psi}^* * \gamma)^*(\mu)$.

Note that the symbol $*$ used as a superscript denotes the convex conjugate of a function. We also remark that attainment of $D(\mu,\nu)$ has to be understood in a "relaxed" sense, since the optimizer $\hat{\psi}$ is not necessarily ν-integrable; see [BBST23, Proposition 4.2].

2.2. Static martingale optimal transport

We fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_c \nu$ and consider a static / discrete-time version of the continuous-time martingale optimization problem (1.1), to wit

$$\tilde{P}(\mu,\nu) := \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx). \tag{2.2}$$

The collection of martingale transports $\mathsf{MT}(\mu,\nu)$ consists of those couplings $\pi \in \mathsf{Cpl}(\mu,\nu)$ that satisfy $\mathrm{bary}(\pi_x) := \int y\, \pi_x(dy) = x$, for μ-a.e. $x \in \mathbb{R}^d$. Here, the family of probability measures $\{\pi_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e., $\pi(dx,dy) = \pi_x(dy)\,\mu(dx)$.

By [BBHK20, Theorem 2.2] the value $\tilde{P}(\mu,\nu)$ of (2.2) is finite and equal to $P(\mu,\nu)$, as defined in (1.1). Furthermore, there exists a unique optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ of (2.2) and

if $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ is the stretched Brownian motion from μ to ν, then the law of $(\hat{M}_0, \hat{M}_1)$ is equal to $\hat{\pi}$.

As already alluded to, maximizing the maximal covariance is equivalent to minimizing the squared quadratic Wasserstein distance, modulo adding constants. More precisely, in the present setting we have the relation

$$\inf_{\pi\in\mathsf{MT}(\mu,\nu)} \int \mathcal{W}_2^2(\pi_x,\gamma)\,\mu(dx) = d + \int |y|^2\,d\nu(y) - 2\tilde{P}(\mu,\nu),$$

where the quadratic Wasserstein distance $\mathcal{W}_2(\,\cdot\,,\,\cdot\,)$ between two probability measures $p_1, p_2 \in \mathscr{P}_2(\mathbb{R}^d)$ is defined as

$$\mathcal{W}_2(p_1,p_2) \coloneqq \sqrt{\inf_{q\in\mathsf{Cpl}(p_1,p_2)} \int |x_1-x_2|^2\,q(dx_1,dx_2)}. \tag{2.3}$$

In these terms, the value of (MBB) can be expressed as

$$MT(\mu,\nu) = \inf_{\pi\in\mathsf{MT}(\mu,\nu)} \int \mathcal{W}_2^2(\pi_x,\gamma)\,\mu(dx) - \int |x|^2\,d\mu(x).$$

2.3. Structure of optimizers

From [BBST23, Theorem 6.6] we recall the following characterization of the dual optimizer $\hat{\psi}$ of (2.1) and of the primal optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ of (2.2).

Lemma 2.2. *Let $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. Suppose that a Bass martingale $(\hat{M}_t)_{0\leqslant t\leqslant 1}$ from μ to ν with Bass measure $\hat{\alpha} \in \mathscr{P}(\mathbb{R}^d)$ and associated convex function $\hat{v}$ exists. Then the Legendre transform $\hat{v}^*$ is equal to the dual optimizer $\hat{\psi}$ of (2.1) and $\mathrm{Law}(\hat{M}_0,\hat{M}_1)$ is equal to the primal optimizer $\hat{\pi}$ of (2.2). Furthermore, we have $\hat{\alpha} = \nabla\hat{\varphi}(\mu)$, where*

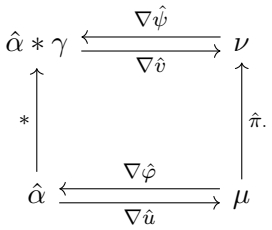
$$\nabla\hat{\varphi}(x) = (\nabla\hat{v} * \gamma)^{-1}(x) = \nabla(\hat{v}*\gamma)^*(x), \tag{2.4}$$

for μ-a.e. $x \in \mathbb{R}^d$, and

$$\hat{\pi}_x = \mathrm{Law}(\hat{M}_1 \,|\, \hat{M}_0 = x) = \nabla\hat{v}(\gamma_{\nabla\hat{\varphi}(x)}), \tag{2.5}$$

where $\gamma_{\nabla\hat{\varphi}(x)}$ denotes the d-dimensional Gaussian distribution with barycenter $\nabla\hat{\varphi}(x)$ and covariance matrix I_d.

We set $\hat{u} \coloneqq \hat{v} * \gamma$, so that $\nabla\hat{u}(\hat{\alpha}) = \mu$. Recalling (1.3), we summarize the relationships between the optimizers in the following diagram:



Finally, we prove the equivalence between the identities (1.3) and the existence of a Bass martingale from μ to ν.

Lemma 2.3. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. There is a Bass martingale $\hat{M}$ with Bass measure $\hat{\alpha} \in \mathscr{P}(\mathbb{R}^d)$ from $\mu = \mathrm{Law}(\hat{M}_0)$ to $\nu = \mathrm{Law}(\hat{M}_1)$ if and only if there is a convex function $\hat{v}\colon \mathbb{R}^d \to \mathbb{R}$ satisfying the identities*

$$(\nabla \hat{v} * \gamma)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * \gamma) = \nu. \tag{2.6}$$

Moreover, the Bass martingale $\hat{M}$ can be expressed as

$$\hat{M}_t = \nabla \hat{v}_t(B_t), \qquad 0 \leqslant t \leqslant 1. \tag{2.7}$$

Proof. Let $\hat{M}$ be a Bass martingale in the sense of Definition 1.2. We first prove (2.7). Let $A \subseteq \mathbb{R}^d$ be a Borel set. We have to show that

$$\mathbb{E}\big[\nabla \hat{v}(B_1)\, \mathbf{1}_{\{B_t \in A\}}\big] = \mathbb{E}\big[(\nabla \hat{v} * \gamma^{1-t})(B_t)\, \mathbf{1}_{\{B_t \in A\}}\big]. \tag{2.8}$$

Denote by $\varphi_t(x, y)$ the Gaussian kernel, for $t \in (0, 1]$ and $x, y \in \mathbb{R}^d$. Then the left-hand side of (2.8) can be expressed as

$$\int \hat{\alpha}(dx_0) \int_A \varphi_t(x_0, dx_t) \int \nabla \hat{v}(x_1)\, \varphi_{1-t}(x_t, dx_1),$$

while the right-hand side is equal to

$$\int \hat{\alpha}(dx_0) \int_A (\nabla \hat{v} * \gamma^{1-t})(x_t)\, \varphi_t(x_0, dx_t).$$

Now we see that (2.8) follows from

$$\int \nabla \hat{v}(x_1)\, \varphi_{1-t}(x_t, dx_1) = \int \nabla \hat{v}(x_1)\, \gamma_{x_t}^{1-t}(dx_1) = (\nabla \hat{v} * \gamma^{1-t})(x_t),$$

where $\gamma_{x_t}^{1-t}$ denotes the d-dimensional Gaussian distribution with barycenter x_t and covariance matrix $(1-t)I_d$. This completes the proof of (2.7).

In particular, at times $t = 0$ and $t = 1$ we obtain from (2.7) that $\hat{M}_0 = (\nabla \hat{v} * \gamma)(B_0)$ and $\hat{M}_1 = \nabla \hat{v}(B_1)$, respectively. If $\hat{M}$ is a Bass martingale from $\mu = \mathrm{Law}(\hat{M}_0)$ to $\nu = \mathrm{Law}(\hat{M}_1)$, this readily gives (2.6).

Conversely, suppose that $\mu, \nu, \hat{\alpha}, \hat{v}$ satisfy the identities (2.6). Let $(B_t)_{0 \leqslant t \leqslant 1}$ be Brownian motion on $\mathbb{R}^d$ with $\mathrm{Law}(B_0) = \hat{\alpha}$. We then define a process $(\hat{M}_t)_{0 \leqslant t \leqslant 1}$ by (1.2). In light of the previous argument, $\hat{M}$ is characterized by (2.7). Since by assumption the identities (2.6) are satisfied, we see that $\mathrm{Law}(\hat{M}_0) = \mu$ and $\mathrm{Law}(\hat{M}_1) = \nu$. Thus $\hat{M}$ is indeed a Bass martingale from μ to ν. □

3. A variational characterization of Bass measures

Throughout this chapter we fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$ and provide the proof of Theorem 1.5. This is done in several steps.

Lemma 3.1. *We have the weak duality*

$$P(\mu, \nu) \leqslant \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{3.1}$$

Proof. Let $\alpha \in \mathscr{P}_2(\mathbb{R}^d)$ be arbitrary. By Brenier's theorem [Vil03, Theorem 2.12] there is a convex function v such that $\nabla v(\alpha * \gamma) = \nu$. Hence from the Kantorovich duality [Vil09, Theorem 5.10] it follows that

$$\mathrm{MCov}(\alpha * \gamma, \nu) = \int v \, d(\alpha * \gamma) + \int v^* \, d\nu = \int (v * \gamma) \, d\alpha + \int v^* \, d\nu.$$

Since $v * \gamma$ is convex, applying once more the Kantorovich duality yields

$$\begin{aligned} \mathrm{MCov}(\alpha * \gamma, \nu) &= \int v^* \, d\nu - \int (v * \gamma)^* \, d\mu + \int (v * \gamma) \, d\alpha + \int (v * \gamma)^* \, d\mu \\ &\geqslant \int v^* \, d\nu - \int (v * \gamma)^* \, d\mu + \mathrm{MCov}(\alpha, \mu). \end{aligned}$$

Finally, from Theorem 2.1 we deduce that

$$\begin{aligned} \mathrm{MCov}(\alpha * \gamma, \nu) &\geqslant \inf_{\psi \text{ convex}} \Big(\int \psi \, d\nu - \int (\psi^* * \gamma)^* \, d\mu \Big) + \mathrm{MCov}(\alpha, \mu) \\ &= P(\mu, \nu) + \mathrm{MCov}(\alpha, \mu), \end{aligned}$$

which gives the inequality (3.1). We remark that it is immaterial whether in (2.1) we optimize over convex functions ψ which are elements of $L^1(\nu)$ or which are just μ-a.s. finite, see [BBST23, Section 4]. □

Lemma 3.2. *Suppose that there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$. Then the right-hand side of (3.1) is attained by $\hat{\alpha}$ and is equal to*

$$\mathcal{V}(\hat{\alpha}) = \int \mathrm{MCov}(\hat{\pi}_x, \gamma) \, \mu(dx), \tag{3.2}$$

where $\hat{\pi} \in \mathsf{MT}(\mu, \nu)$ is the optimizer of (2.2).

Proof. By assumption there exists a Bass martingale from μ to ν, with Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ and associated convex function $\hat{v}$ satisfying (recall Lemma 2.3) the identities (2.6). According to Lemma 2.2, we have that $\hat{\alpha} = \nabla\hat{\varphi}(\mu)$ and

$$\hat{\pi}_x = \nabla\hat{v}(\gamma_{\nabla\hat{\varphi}(x)}),$$

for μ-a.e. $x \in \mathbb{R}^d$. Applying Brenier's theorem, we deduce that

$$\begin{aligned}
\int \mathrm{MCov}(\hat{\pi}_x, \gamma)\, \mu(dx) &= \int\int \langle \nabla\hat{v}(\nabla\hat{\varphi}(x) + z), z\rangle\, \gamma(dz)\, \mu(dx) \\
&= \int\int \Big(\langle \nabla\hat{v}(\nabla\hat{\varphi}(x) + z), \nabla\hat{\varphi}(x) + z\rangle - \langle \nabla\hat{v}(\nabla\hat{\varphi}(x) + z), \nabla\hat{\varphi}(x)\rangle \Big)\, \gamma(dz)\, \mu(dx) \\
&= \int\int \left(\langle \nabla\hat{v}(a + z), a + z\rangle - \langle \nabla\hat{v}(a + z), a\rangle \right) \gamma(dz)\, \hat{\alpha}(da) \\
&= \int \langle \nabla\hat{v}, \mathrm{Id}\rangle\, d(\hat{\alpha} * \gamma) - \int \langle (\nabla\hat{v} * \gamma), \mathrm{Id}\rangle\, d\hat{\alpha} \\
&= \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha}, \mu) = \mathcal{V}(\hat{\alpha}),
\end{aligned}$$

which shows (3.2). Together with the weak duality (3.1) of Lemma 3.1 above, and recalling from Section 2.2 that the right-hand side of (3.2) is equal to $\tilde{P}(\mu, \nu) = P(\mu, \nu)$, we conclude the assertion of Lemma 3.2. □

Lemma 3.3. *We have the duality result*

$$P(\mu, \nu) = \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha). \tag{3.3}$$

Proof. For $\varepsilon > 0$ we define $\mu^\varepsilon := \mu * \gamma^\varepsilon$ and $\nu^\varepsilon := \nu * \gamma^{2\varepsilon}$. Then $\mu^\varepsilon \preceq_{\mathrm{c}} \nu^\varepsilon$ and the pair $(\mu^\varepsilon, \nu^\varepsilon)$ is irreducible. Hence by Theorem 1.3 there is a Bass martingale from μ^ε to ν^ε, so that by Lemma 3.2 we have

$$\sup_{\pi \in \mathsf{MT}(\mu^\varepsilon, \nu^\varepsilon)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu^\varepsilon(dx) = \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(\alpha * \gamma, \nu^\varepsilon) - \mathrm{MCov}(\alpha, \mu^\varepsilon) \Big). \tag{3.4}$$

By weak optimal transport arguments (see [BJMP23, Theorem 2.3]) we know

$$\limsup_{\varepsilon \to 0} \sup_{\pi \in \mathsf{MT}(\mu^\varepsilon, \nu^\varepsilon)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu^\varepsilon(dx) \leqslant \sup_{\pi \in \mathsf{MT}(\mu, \nu)} \int \mathrm{MCov}(\pi_x, \gamma)\, \mu(dx).$$

Therefore, if we can show that the right-hand side of (3.4) converges to the right-hand side of (3.3), we will obtain the inequality

$$P(\mu, \nu) \geqslant \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \mathcal{V}(\alpha) = \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) \Big),$$

which, together with the weak duality of Lemma 3.1, establishes (3.3). But this follows easily from

$$|\mathrm{MCov}(\alpha, \mu^\varepsilon) - \mathrm{MCov}(\alpha, \mu)| \leqslant c_1\varepsilon + \tfrac{1}{2}|\mathcal{W}_2^2(\alpha, \mu^\varepsilon) - \mathcal{W}_2^2(\alpha, \mu)| \leqslant c_2(\varepsilon + \varepsilon^2)$$

and a similar estimate for $|\mathrm{MCov}(\alpha * \gamma, \nu^\varepsilon) - \mathrm{MCov}(\alpha * \gamma, \nu)|$. □

Lemma 3.4. *Suppose that the right-hand side of (3.3) is attained by $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$. Then there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha}$.*

Proof. By Brenier's theorem there is a convex function $\hat{v}$ such that $\nabla\hat{v}(\hat{\alpha} * \gamma) = \nu$. According to Lemma 2.3, for the existence of a Bass martingale from μ to ν, it remains to show the first equality in (2.6), i.e.,

$$(\nabla\hat{v} * \gamma)(\hat{\alpha}) = \mu. \tag{3.5}$$

Let $\hat{Z}$ and X be random variables with laws $\hat{\alpha}$ and μ, respectively, such that

$$\mathrm{MCov}(\hat{\alpha}, \mu) = \mathbb{E}\big[\langle \hat{Z}, X\rangle\big]. \tag{3.6}$$

Denote by $\hat{q}(dz, dx)$ the law of the coupling $(\hat{Z}, X)$. Let $\boldsymbol{w}\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a smooth function with compact support and define probability measures $(\alpha_u)_{u\in\mathbb{R}} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ by

$$\int f\, d\alpha_u \coloneqq \int\int f(z + u\boldsymbol{w}(z,x))\, q(dz, dx), \qquad f \in C_b(\mathbb{R}^d). \tag{3.7}$$

We claim that

$$\liminf_{u\to 0} \tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u, \mu) - \mathrm{MCov}(\hat{\alpha}, \mu)\Big) \geqslant \mathbb{E}\big[\langle \boldsymbol{w}(\hat{Z}, X), X\rangle\big] \tag{3.8}$$

and

$$\lim_{u\to 0} \tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) = \mathbb{E}\big[\langle \boldsymbol{w}(\hat{Z}, X), (\nabla\hat{v} * \gamma)(\hat{Z})\rangle\big]. \tag{3.9}$$

Using the optimality of $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ for the right-hand side of (3.3) and admitting the two claims (3.8), (3.9), we deduce that

$$\begin{aligned} 0 &\leqslant \liminf_{u\to 0} \tfrac{1}{u}\Big(\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) - \Big(\mathrm{MCov}(\alpha_u, \mu) - \mathrm{MCov}(\hat{\alpha}, \mu)\Big)\Big) \\ &\leqslant \mathbb{E}\big[\langle \boldsymbol{w}(\hat{Z}, X), (\nabla\hat{v} * \gamma)(\hat{Z}) - X\rangle\big]. \end{aligned}$$

Since $\boldsymbol{w}$ was arbitrary, it follows that the random variable $(\nabla\hat{v} * \gamma)(\hat{Z})$ has the same law as X, which readily gives (3.5).

We now turn to the proof of the claim (3.8). By the definition of α_u in (3.7), the random variable $Z_u \coloneqq \hat{Z} + u\boldsymbol{w}(\hat{Z}, X)$ has law α_u. Consequently,

$$\mathrm{MCov}(\alpha_u, \mu) \geqslant \mathbb{E}\big[\langle Z_u, X\rangle\big]. \tag{3.10}$$

Combining (3.6) and (3.10) yields (3.8).

It remains to show the claim (3.9). By analogy with the proof of (3.8), we obtain the inequality "$\geqslant$" in (3.9). For the reverse inequality, we note that by the Kantorovich duality we have

$$\begin{aligned} \mathrm{MCov}(\alpha_u * \gamma, \nu) &= \inf_{v \text{ convex}} \Big(\int v\, d(\alpha_u * \gamma) - \int v^*\, d\nu\Big) \\ &\leqslant \int \hat{v}\, d(\alpha_u * \gamma) - \int \hat{v}^*\, d\nu \\ &= \int (\hat{v} * \gamma)\, d\alpha_u - \int \hat{v}^*\, d\nu \end{aligned}$$

and

$$\mathrm{MCov}(\hat{\alpha} * \gamma, \nu) = \int \hat{v}\, d(\alpha * \gamma) + \int \hat{v}^*\, d\nu = \int (\hat{v} * \gamma)\, d\hat{\alpha} + \int \hat{v}^*\, d\nu.$$

Therefore

$$\begin{aligned}\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) &\leqslant \int (\hat{v} * \gamma)\, d\alpha_u - \int (\hat{v} * \gamma)\, d\hat{\alpha} \\ &= \mathbb{E}[(\hat{v} * \gamma)(\hat{Z} + u\boldsymbol{w}(\hat{Z}, X)) - (\hat{v} * \gamma)(\hat{Z})]\end{aligned}$$

Using the convexity of the function $\hat{v} * \gamma$, we deduce that

$$\tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) \leqslant \mathbb{E}[\langle \boldsymbol{w}(\hat{Z}, X), (\nabla \hat{v} * \gamma)(\hat{Z} + u\boldsymbol{w}(\hat{Z}, X))\rangle].$$

Now observe that the expectation on the right-hand side of the above inequality is equal to the expectation of the random variable

$$Y_u := \Big\langle \boldsymbol{w}(\hat{Z}, X)\,, \nabla \hat{v}(\hat{Z} + \Gamma) \exp\left(u\langle \Gamma, \boldsymbol{w}(\hat{Z}, X)\rangle - \tfrac{u^2}{2}|\boldsymbol{w}(\hat{Z}, X)|^2\right)\Big\rangle,$$

where Γ is a standard Gaussian random vector on $\mathbb{R}^d$, independent of $\hat{Z}$ as well as of X. Clearly by continuity

$$\lim_{u \to 0} Y_u = \langle \boldsymbol{w}(\hat{Z}, X), \nabla \hat{v}(\hat{Z} + \Gamma)\rangle, \qquad \mathbb{P}\text{-a.s.}$$

As $\boldsymbol{w}$ is smooth with compact support, for $\delta > 0$ we can find constants c_1, c_2 such that

$$\forall u \in [-\delta, \delta]\colon \quad |Y_u| \leqslant c_1\, |\nabla \hat{v}(\hat{Z} + \Gamma)|\, \mathrm{e}^{c_2|\Gamma|}.$$

By the Cauchy–Schwarz inequality and since $\nabla \hat{v}(\hat{\alpha} * \gamma) = \nu \in \mathscr{P}_2(\mathbb{R}^d)$, we have the bound

$$\mathbb{E}[|\nabla \hat{v}(\hat{Z} + \Gamma)|\, \mathrm{e}^{|\Gamma|}] \leqslant \sqrt{\int |y|^2\, d\nu(y)}\, \sqrt{\mathbb{E}\big[\mathrm{e}^{2|\Gamma|}\big]} < +\infty.$$

Therefore we can apply the dominated convergence theorem and conclude that

$$\limsup_{u \to 0} \tfrac{1}{u}\Big(\mathrm{MCov}(\alpha_u * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu)\Big) \leqslant \mathbb{E}[\langle \boldsymbol{w}(\hat{Z}, X), (\nabla \hat{v} * \gamma)(\hat{Z})\rangle],$$

which completes the proof of the claim (3.9). □

Proof of Theorem 1.5. The assertion of the theorem follows from Lemmas 3.2 – 3.4. □

The reader has certainly noticed that the proof of Lemma 3.2 was given in an analytic style while the proof of Lemma 3.4 was given in a more probabilistic language. In the remainder of this chapter we give an alternative probabilistic proof of Lemma 3.2 and sketch how to translate the proof of Lemma 3.4 into a more analytic language.

The following probabilistic proof of Lemma 3.2 does not require the duality results developed in [BBST23], but only relies on the definition of Bass martingales.

Probabilistic proof of Lemma 3.2. By assumption there exists a Bass martingale from μ to ν, with Bass measure $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$ and associated convex function $\hat{v}$ satisfying (recall Lemma 2.3) the identities (2.6). Let $\alpha \in \mathscr{P}_2(\mathbb{R}^d)$ be arbitrary. We have to show that

$$\begin{aligned} \mathcal{V}(\hat{\alpha}) &= \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha}, \mu) \leqslant \\ &\leqslant \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) = \mathcal{V}(\alpha). \end{aligned} \tag{3.11}$$

Take a random variable $\hat{Z}$ with law $\hat{\alpha}$ and define

$$X := (\nabla \hat{v} * \gamma)(\hat{Z}). \tag{3.12}$$

By Brenier's theorem the coupling $(\hat{Z}, X)$ is optimal and according to (2.6) the random variable X has law μ. Now choose a random variable Z with law α such that the coupling (Z, X) is optimal with respect to the maximal covariance (equivalently, with respect to the quadratic Wasserstein distance). Clearly

$$\mathrm{MCov}(\alpha, \mu) - \mathrm{MCov}(\hat{\alpha}, \mu) = \mathbb{E}\big[\langle Z - \hat{Z}, X \rangle\big]. \tag{3.13}$$

Take a standard Gaussian random vector Γ on $\mathbb{R}^d$, independent of Z as well as of $\hat{Z}$. The random variables $\hat{Z} + \Gamma$ and

$$Y := \nabla \hat{v}(\hat{Z} + \Gamma) \tag{3.14}$$

have laws $\hat{\alpha} * \gamma$ and ν, respectively. As by Brenier's theorem the coupling $(\hat{Z} + \Gamma, Y)$ is optimal, we have

$$\mathrm{MCov}(\hat{\alpha} * \gamma, \nu) = \mathbb{E}\big[\langle \hat{Z} + \Gamma, Y \rangle\big]. \tag{3.15}$$

Since the random variable $Z + \Gamma$ has law $\alpha * \gamma$ we conclude that $(Z + \Gamma, Y)$ is some coupling between $\alpha * \gamma$ and ν, i.e.,

$$\mathrm{MCov}(\alpha * \gamma, \nu) \geqslant \mathbb{E}\big[\langle Z + \Gamma, Y \rangle\big]. \tag{3.16}$$

From (3.13) – (3.16) we obtain the inequality

$$\mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\hat{\alpha} * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) + \mathrm{MCov}(\hat{\alpha}, \mu) \geqslant \mathbb{E}\big[\langle Z - \hat{Z}, Y - X \rangle\big].$$

Therefore, in order to establish the inequality (3.11), it remains to show that

$$\mathbb{E}\big[\langle Z - \hat{Z}, Y - X \rangle\big] = 0. \tag{3.17}$$

For that purpose, we condition $Y - X$ on the random variables Z as well as $\hat{Z}$, so that by (3.12) and (3.14) we obtain

$$\mathbb{E}[Y - X \,|\, Z, \hat{Z}] = 0,$$

which implies (3.17). □

We finally give an alternative heuristic argument for Lemma 3.4, which is based on differentiating the maximal covariance along a continuity equation.

Alternative heuristic proof of Lemma 3.4. Suppose that the right-hand side of (3.3) is attained by $\hat{\alpha} \in \mathscr{P}_2(\mathbb{R}^d)$. We want to show that there exists a Bass martingale from μ to ν with Bass measure $\hat{\alpha}$. The idea is to perturb $\hat{\alpha}$ along a continuity equation

$$\partial_t \alpha_t + \operatorname{div}(\boldsymbol{v}_t \alpha_t) = 0, \qquad t \in (-h, h),$$

with $h > 0$, $\alpha_0 \coloneqq \hat{\alpha}$, and where $\boldsymbol{v}_t$ is a velocity field. Observe that

$$\begin{aligned} \partial_t|_{t=0} \mathcal{V}(\alpha_t) &= \partial_t|_{t=0} \Big(\mathrm{MCov}(\alpha_t * \gamma, \nu) - \mathrm{MCov}(\alpha_t, \mu) \Big) \\ &= \partial_t|_{t=0} \int \hat{v}\, d(\alpha_t * \gamma) - \partial_t|_{t=0} \int \hat{u}\, d\alpha_t, \end{aligned}$$

where $\nabla \hat{v}(\hat{\alpha} * \gamma) = \nu$ is optimal and likewise $\nabla \hat{u}(\hat{\alpha}) = \mu$ is optimal. By the continuity equation we obtain

$$\partial_t|_{t=0} \int \hat{u}\, d\alpha_t = \int \langle \nabla \hat{u}, \boldsymbol{v}_0 \rangle \, d\hat{\alpha}.$$

With similar computations we have

$$\partial_t|_{t=0} \int \hat{v}\, d(\alpha_t * \gamma) = \int \langle \nabla \hat{v} * \gamma, \boldsymbol{v}_0 \rangle \, d\hat{\alpha}.$$

As $\boldsymbol{v}_0$ was arbitrary and $\hat{\alpha}$ was optimal, we conclude that

$$0 = \int \langle \nabla \hat{v} * \gamma - \nabla \hat{u}, \boldsymbol{v}_0 \rangle \, d\hat{\alpha},$$

so that $\nabla \hat{u}$, the optimal map from $\hat{\alpha}$ to μ, is $\hat{\alpha}$-a.s. equal to $\nabla \hat{v} * \gamma$, where $\nabla \hat{v}$ is the optimal map from $\hat{\alpha} * \gamma$ to ν. Recalling (2.6) and Lemma 2.3, this is precisely the structure of the Bass martingale. □

4. An infinitesimal version of Theorem 1.5

We provide the proof of Theorem 1.6, an infinitesimal version of Theorem 1.5.

Proof of Theorem 1.6. For a partition $\Pi = \{t_0, t_1, \dots, t_n\}$ of the interval $[0,1]$ with

$$0 = t_0 < t_1 < \ldots < t_n = 1$$

we denote by Σ^Π the collection of all progressively measurable and L^2-bounded processes $(\sigma_t^\Pi)_{0\leqslant t\leqslant 1}$ such that the stochastic integral

$$M_t^\Pi := M_0 + \int_0^t \sigma_s^\Pi \, dB_s, \qquad 0 \leqslant t \leqslant 1$$

defines an L^2-bounded martingale with $\mathrm{Law}(M_{t_k}^\Pi) = \mu_{t_k}$, for $k = 0, \dots, n$. We define

$$m^\Pi([t_{k-1}, t_k]) := \sup_{\sigma^\Pi \in \Sigma^\Pi} \mathbb{E}\Big[\int_{t_{k-1}}^{t_k} \mathrm{tr}(\sigma_s^\Pi)\, ds\Big]. \tag{4.1}$$

By [BBHK20], we know that the optimizer of

$$m^\Pi([0,1]) = \sup_{\sigma^\Pi \in \Sigma^\Pi} \mathbb{E}\Big[\int_0^1 \mathrm{tr}(\sigma_s^\Pi)\, ds\Big]$$

is given, on each interval $[t_{k-1}, t_k]$, by the stretched Brownian motion from $\mu_{t_{k-1}}$ to μ_{t_k}. By Theorem 1.5 we have

$$m^\Pi([t_{k-1}, t_k]) = \inf_{\alpha \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(\alpha * \gamma^{t_k - t_{k-1}}, \mu_{t_k}) - \mathrm{MCov}(\alpha, \mu_{t_{k-1}})\Big). \tag{4.2}$$

For $t_k \in \Pi$ and a refinement Π_1 of Π we have

$$m^\Pi([0, t_k]) \geqslant m^{\Pi_1}([0, t_k]),$$

as the process $(\sigma_t^{\Pi_1})_{0\leqslant t\leqslant 1}$ has to satisfy more requirements than the process $(\sigma_t^\Pi)_{0\leqslant t\leqslant 1}$. We therefore may pass to a limit $m := \lim m^\Pi$ along the net of finite partitions Π of the interval $[0,1]$, which extends to a finite measure on $[0,1]$, still denoted by m. Clearly the measure m is absolutely continuous with respect to Lebesgue measure on $[0,1]$ and we denote the corresponding density by $g(t)$, for $0 \leqslant t \leqslant 1$. We claim that, for $0 \leqslant r \leqslant u \leqslant 1$, we have

$$\mathbb{E}\Big[\int_r^u \mathrm{tr}(\sigma_s)\, ds\Big] \leqslant m([r, u]). \tag{4.3}$$

Indeed, otherwise we could find a partition Π with $r, u \in \Pi$, such that

$$\mathbb{E}\Big[\int_r^u \operatorname{tr}(\sigma_s^\Pi)\, ds\Big] > m^\Pi([r,u]),$$

which yields a contradiction to the definition of $m^\Pi(\,\cdot\,)$ in (4.1). Since (4.3) holds for all intervals $[r,u] \subseteq [0,1]$, we deduce that

$$\mathbb{E}\big[\operatorname{tr}(\sigma_t)\big] \leqslant g(t), \tag{4.4}$$

for Lebesgue-a.e. $0 \leqslant t \leqslant 1$. From (4.2) we conclude, for Lebesgue-a.e. $0 \leqslant t \leqslant 1$ and for each $\alpha \in \mathscr{P}_2(\mathbb{R}^d)$, the inequality

$$g(t) \leqslant \liminf_{h\to 0} \tfrac{1}{h}\Big(\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h}) - \mathrm{MCov}(\alpha, \mu_t)\Big).$$

Together with (4.4), this finishes the proof of (1.7). □

Again we provide a more analytic argument for Theorem 1.6, at least on a formal level.

Alternative heuristic proof of Theorem 1.6. We will use the Kantorovich duality and the Fokker–Planck equations to get a hold of $\frac{d}{dh}\mathrm{MCov}(\alpha * \gamma^h, \mu_{t+h})$. By a change of variables we then get an equivalent expression which, when minimized, gives the left-hand side of (1.7). We suppose here that M is a strong solution of the stochastic differential equation $dM_u = \sigma_u(M_u)\, dB_u$, with σ as benevolent as needed, so that in particular μ_u admits a density for each u.

We set $\rho_h := \alpha * \gamma^h$, $\Sigma := \sigma\sigma'$, and notice that for fixed t we have

$$\begin{aligned} \partial_h \rho_h(x) &= \tfrac{1}{2}\Delta\rho_h(x), \qquad \rho_0 = \alpha; \\ \partial_h \mu_{t+h}(x) &= \tfrac{1}{2}\sum_{i,k} \partial_{ik}^2(\Sigma_{ik}\mu_{t+h}(x)). \end{aligned}$$

By the Kantorovich duality we have

$$\begin{aligned} \mathrm{MCov}(\rho_h, \mu_{t+h}) &= \inf_{\phi \text{ convex}} \int \phi\, d\rho_h + \int \phi^*\, d\mu_{t+h} \\ &= \int \phi_{\rho_h}^{\mu_{t+h}}\, d\rho_h + \int \phi_{\mu_{t+h}}^{\rho_h}\, d\mu_{t+h}, \end{aligned}$$

where we denote by $\phi_p^q(\,\cdot\,)$ the convex function, which is unique up to a constant, such that $\nabla\phi_p^q(p) = q$. Using this, or more directly [Vil09, Theorem 23.9], we have

$$\begin{aligned} \frac{d}{dh}\,\mathrm{MCov}(\rho_h, \mu_{t+h}) &= \int \phi_{\rho_h}^{\mu_{t+h}}\, \partial_h\rho_h\, d\lambda + \int \phi_{\mu_{t+h}}^{\rho_h}\, \partial_h\mu_{t+h}\, d\lambda \\ &= \int \phi_{\rho_h}^{\mu_{t+h}}\, \tfrac{1}{2}\Delta\rho_h\, d\lambda + \int \phi_{\mu_{t+h}}^{\rho_h}\, \tfrac{1}{2}\sum_{i,k} \partial_{ik}^2(\Sigma_{ik}\mu_{t+h})\, d\lambda \\ &= \tfrac{1}{2}\int \sum_{i,k} \partial_{i,k}^2(\phi_{\rho_h}^{\mu_{t+h}})\, I_{ik}\, d\rho_h + \tfrac{1}{2}\int \sum_{i,k} \partial_{i,k}^2(\phi_{\mu_{t+h}}^{\rho_h})\, \Sigma_{ik}\, d\mu_{t+h} \\ &= \tfrac{1}{2}\int \operatorname{tr}\big(D^2(\phi_{\rho_h}^{\mu_{t+h}})\big)\, \rho_h\, d\lambda + \tfrac{1}{2}\int \operatorname{tr}\big(D^2(\phi_{\mu_{t+h}}^{\rho_h})\Sigma\big)\, \mu_{t+h}\, d\lambda, \end{aligned}$$

where we denote by D and D^2 the Jacobian and Hessian matrix, respectively. During this proof we will use the convention that if $x \mapsto a(x) \in \mathbb{R}^d$ is an invertible vector-valued function, then $a^{-1}(x)$ denotes the inverse function, whereas if $x \mapsto A(x) \in \mathbb{R}^{d\times d}$ is a matrix-valued function, then $[A(x)]^{-1}$ denotes the matrix inverse of $A(x)$. Now observe that

$$D^2(\phi_{\rho_h}^{\mu_{t+h}})(x) = D(\nabla\phi_{\rho_h}^{\mu_{t+h}})(x) = D\big((\nabla\phi_{\mu_{t+h}}^{\rho_h})^{-1}\big)(x) = [D\nabla\phi_{\mu_{t+h}}^{\rho_h} \circ \nabla\phi_{\rho_h}^{\mu_{t+h}}(x)]^{-1},$$

so that

$$\begin{aligned}\int \operatorname{tr}\big(D^2(\phi_{\rho_h}^{\mu_{t+h}})(x)\big)\,\rho_h\,d\lambda &= \int \operatorname{tr}\big([D\nabla\phi_{\mu_{t+h}}^{\rho_h} \circ \nabla\phi_{\rho_h}^{\mu_{t+h}}(x)]^{-1}\big)\,\rho_h\,d\lambda \\ &= \int \operatorname{tr}\big([D^2\phi_{\mu_{t+h}}^{\rho_h}(y)]^{-1}\big)\,\mu_{t+h}\,d\lambda.\end{aligned}$$

Altogether we have

$$\frac{d}{dh}\,\mathrm{MCov}(\rho_h,\mu_{t+h}) = \tfrac{1}{2}\int \Big(\operatorname{tr}\big([D^2\phi_{\mu_{t+h}}^{\rho_h}]^{-1}\big) + \operatorname{tr}\big(D^2(\phi_{\mu_{t+h}}^{\rho_h})\Sigma\big)\Big)\,\mu_{t+h}\,d\lambda.$$

Define now the functional on invertible, positive-semidefinite symmetric matrices

$$A \mapsto J(A) \coloneqq \operatorname{tr}(A^{-1}) + \operatorname{tr}(A\Sigma).$$

We remark that $J(A) \geqslant 2\operatorname{tr}(\Sigma^{1/2})$, since this is equivalent to the trivial statement

$$|A^{-1/2} - A^{1/2}\Sigma^{1/2}|_{\mathrm{HS}} \geqslant 0.$$

Hence in fact the minimum of $J(\,\cdot\,)$ is attained at $A = \Sigma^{-1/2}$. We conclude that

$$\frac{d}{dh}\Big|_{h=0}\,\mathrm{MCov}(\rho_h,\mu_{t+h}) \geqslant \int \operatorname{tr}(\sigma_t(y))\,\mu_t(dy) = \mathbb{E}\big[\operatorname{tr}(\sigma_t(M_t))\big],$$

which completes the proof of Theorem 1.6. □

5. Displacement convexity of the Bass functional

We observe that the Bass functional $\alpha \mapsto \mathcal{V}(\alpha)$ provides a novel example of a convex functional with respect to the almost-Riemannian structure of the quadratic Wasserstein space $\mathscr{P}_2$. As mentioned in [Vil03, Open Problem 5.17], there are only few known examples of so-called displacement convex functionals (see [McC97], [Vil03, Definition 5.10], [AGS08, Definition 9.1.1]), and it is desirable to find new ones.

We shall state two versions of this result. The first one, Proposition 5.1, pertains to the case $d = 1$, while the second one, Proposition 5.2, holds for general $d \in \mathbb{N}$. We also note that, contrary to the rest of this work, we do not assume that $\mu \preceq_{\mathrm{c}} \nu$.

Proposition 5.1. *Suppose $d = 1$. Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R})$. The Bass functional*

$$\mathscr{P}_2(\mathbb{R}) \ni \alpha \longmapsto \mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) \tag{5.1}$$

is displacement convex. Moreover, if a geodesic $(\alpha_u)_{0 \leqslant u \leqslant 1}$ in $\mathscr{P}_2(\mathbb{R})$ is such that α_1 is not a translate of α_0 and if ν is not a Dirac measure, the function $u \mapsto \mathcal{V}(\alpha_u)$ is strictly convex.

Proof. We start by noting that the Bass functional $\mathcal{V}(\,\cdot\,)$ of (5.1) can equivalently be defined in terms of the quadratic Wasserstein distance $\mathcal{W}_2(\,\cdot\,,\,\cdot\,)$ of (2.3) rather than in terms of the maximal covariance $\mathrm{MCov}(\,\cdot\,,\,\cdot\,)$ of (1.4). Indeed, we have the identity

$$\mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu) = \tfrac{1}{2}\mathcal{W}_2^2(\alpha, \mu) - \tfrac{1}{2}\mathcal{W}_2^2(\alpha * \gamma, \nu) + \mathrm{const},$$

where the constant

$$\mathrm{const} = \tfrac{d}{2} + \tfrac{1}{2}\int |y|^2 \, d\nu(y) - \tfrac{1}{2}\int |x|^2 \, d\mu(x)$$

does not depend on α. Therefore showing the (strict) displacement convexity of the Bass functional $\mathcal{V}(\,\cdot\,)$ is equivalent to showing the (strict) displacement convexity of the functional

$$\mathcal{U}(\alpha) \coloneqq \mathcal{W}_2^2(\alpha, \mu) - \mathcal{W}_2^2(\alpha * \gamma, \nu). \tag{5.2}$$

Fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R})$ and let $(\alpha_u)_{0 \leqslant u \leqslant 1}$ be a geodesic in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R})$. Using the hypothesis $d = 1$ we can choose mutually comonotone random variables Z_0, Z_1 and X with laws α_0, α_1 and μ, respectively. As $(\alpha_u)_{0 \leqslant u \leqslant 1}$ is a geodesic, the random variable $Z_u \coloneqq (1-u)Z_0 + uZ_1$ has law α_u, for $0 \leqslant u \leqslant 1$. Also note that each

Z_u is comonotone with X. Let $u_0, u \in [0,1]$. As regards the first Wasserstein distance in (5.2), a straightforward calculation yields

$$\begin{aligned}
&\mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_{u_0}, \mu) = && (5.3)\\
&\qquad = \mathbb{E}[|Z_u - X|^2] - \mathbb{E}[|Z_{u_0} - X|^2] && (5.4)\\
&\qquad = \mathbb{E}[|Z_u - Z_{u_0}|^2] - 2\mathbb{E}[\langle Z_u - Z_{u_0}, X - Z_{u_0}\rangle] && (5.5)\\
&\qquad = (u-u_0)^2\mathbb{E}[|Z_1 - Z_0|^2] - 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, X - Z_{u_0}\rangle]. && (5.6)
\end{aligned}$$

Passing to the second Wasserstein distance in (5.2), we take a standard Gaussian random variable Γ on $\mathbb{R}$, independent of Z_0 as well as of Z_1. Next we choose a random variable Y_{u_0} such that $(Z_{u_0} + \Gamma, Y_{u_0})$ is an optimal coupling of $(\alpha_{u_0} * \gamma, \nu)$. As $(Z_u * \Gamma, Y_{u_0})$ is a (typically sub-optimal) coupling of $(\alpha_u * \gamma, \nu)$, we obtain the inequality

$$\begin{aligned}
&\mathcal{W}_2^2(\alpha_u * \gamma, \nu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu) \leqslant && (5.7)\\
&\qquad \leqslant \mathbb{E}[|Z_u + \Gamma - Y_{u_0}|^2] - \mathbb{E}[|Z_{u_0} + \Gamma - Y_{u_0}|^2] && (5.8)\\
&\qquad = \mathbb{E}[|Z_u - Z_{u_0}|^2] - 2\mathbb{E}[\langle Z_u - Z_{u_0}, Y_{u_0} - (Z_{u_0} + \Gamma)\rangle] && (5.9)\\
&\qquad = (u-u_0)^2\mathbb{E}[|Z_1 - Z_0|^2] - 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - (Z_{u_0} + \Gamma)\rangle]. && (5.10)
\end{aligned}$$

Combining (5.3) – (5.10), we deduce that

$$\begin{aligned}
&\mathcal{U}(\alpha_u) - \mathcal{U}(\alpha_{u_0}) = && (5.11)\\
&\qquad = \Big(\mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu)\Big) - \Big(\mathcal{W}_2^2(\alpha_{u_0}, \mu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu)\Big) \geqslant && (5.12)\\
&\qquad \geqslant 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - X - \Gamma\rangle] && (5.13)\\
&\qquad = 2(u-u_0)\mathbb{E}[\langle Z_1 - Z_0, Y_{u_0} - X\rangle], && (5.14)
\end{aligned}$$

where the last equation follows from conditioning on Z_0, Z_1. The expression in (5.14) defines a linear function in u, which lies below and touches the function

$$\begin{aligned}
u \longmapsto &\mathcal{U}(\alpha_u) - \mathcal{U}(\alpha_{u_0}) = \\
&= \Big(\mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu)\Big) - \Big(\mathcal{W}_2^2(\alpha_{u_0}, \mu) - \mathcal{W}_2^2(\alpha_{u_0} * \gamma, \nu)\Big)
\end{aligned}$$

at the point $u = u_0$. This readily implies the convexity of the function

$$u \longmapsto \mathcal{U}(\alpha_u) = \mathcal{W}_2^2(\alpha_u, \mu) - \mathcal{W}_2^2(\alpha_u * \gamma, \nu).$$

It remains to show the strict convexity assertion of Proposition 5.1. If α_1 is not a translate of α_0, then α_u is not a translate of α_{u_0} either, provided that $u \neq u_0$. As $Z_{u_0} + \Gamma$ is comonotone with Y_{u_0} and Y_{u_0} is assumed to be non-constant, we may find $y_0 \in \mathbb{R}$ and $z_0 \in \mathbb{R}$ such that $\mathbb{P}[Y_{u_0} < y_0] \in (0,1)$ and

$$\{Z_{u_0} + \Gamma < z_0\} = \{Y_{u_0} < y_0\}.$$

If $Z_u + \Gamma$ were also comonotone with Y_{u_0}, we could find $z \in \mathbb{R}$ such that

$$\{Z_{u_0} + \Gamma < z_0\} = \{Y_{u_0} < y_0\} = \{Z_u + \Gamma < z\},$$

where we have used that the law of $Z_u + \Gamma$ is continuous. Conditioning on $\Gamma = \zeta$ this implies that, for Lebesgue-a.e. $\zeta \in \mathbb{R}$,

$$\{Z_{u_0} < z_0 - \zeta\} = \{Z_u < z - \zeta\},$$

so that Z_{u_0} and Z_u are translates. This gives the desired contradiction, showing that there is a strict inequality in (5.7), (5.8) (thus also in (5.12), (5.13)), which implies the strict convexity assertion of Proposition 5.1. □

We now pass to the case of general $d \in \mathbb{N}$. In Proposition 5.2 below we formulate a convexity property of the Bass functional $\mathcal{V}(\,\cdot\,)$ pertaining to the notion of *generalized geodesics* as analyzed in [AGS08, Definition 9.2.2]. Recall that $(\alpha_u)_{0 \leqslant u \leqslant 1}$ is a generalized geodesic with base μ, joining α_0 to α_1, if there are random variables Z_0, Z_1 and X with laws α_0, α_1 and μ, respectively, such that (Z_0, X) and (Z_1, X) are optimal couplings and such that the random variable $Z_u := uZ_1 + (1-u)Z_0$ has law α_u, for $0 \leqslant u \leqslant 1$.

Proposition 5.2. *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$. The Bass functional*

$$\mathscr{P}_2(\mathbb{R}^d) \ni \alpha \longmapsto \mathcal{V}(\alpha) = \mathrm{MCov}(\alpha * \gamma, \nu) - \mathrm{MCov}(\alpha, \mu)$$

is convex along generalized geodesics $(\alpha_u)_{0 \leqslant u \leqslant 1}$ in $\mathscr{P}_2(\mathbb{R}^d)$ with base μ.

Proof. We follow the lines of the proof of Proposition 5.1 and consider again the functional

$$\mathcal{U}(\alpha) = \mathcal{W}_2^2(\alpha, \mu) - \mathcal{W}_2^2(\alpha * \gamma, \nu)$$

as in (5.2). Let $(\alpha_u)_{0 \leqslant u \leqslant 1}$ be a generalized geodesic with base μ, joining α_0 to α_1. Take Z_0, Z_1, Z_u, X as above such that (Z_0, X) and (Z_1, X) are optimal couplings and by definition $Z_u \sim \alpha_u$. Note that (Z_u, X) is an optimal coupling of (α_u, μ) by [AGS08, Lemma 9.2.1], for $0 \leqslant u \leqslant 1$. The equalities (5.3) – (5.6) and the inequality in (5.7) – (5.10) then carry over verbatim and we again arrive at (5.12) – (5.14), which shows the convexity of the function $[0,1] \ni u \mapsto \mathcal{U}(\alpha_u)$. □

Part VIII.

q-Bass martingales

Abstract

An intriguing question in martingale optimal transport is to characterize the martingale with prescribed initial and terminal marginals whose transition kernel is as Gaussian as possible. In this work we address an extension of this question, in which the role of the Gaussian distribution is replaced by an arbitrary reference measure q. Our first main result is a dual formulation of the corresponding martingale optimization problem in terms of convex functions.

In the well-studied case when q is Gaussian, the careful analysis of the solution to the above-mentioned optimization problem is a crucial building block in the construction of Bass martingales, i.e., Brownian martingales induced by gradients of convex functions with possibly non-degenerate starting laws. In our second main result we extend this concept beyond the Gaussian case by introducing the notion of q-Bass martingales in discrete time, and give sufficient conditions for their existence.

1. Introduction

1.1. Martingale optimization problem

Let μ, ν be elements of $\mathscr{P}_2(\mathbb{R}^d)$, the collection of probability measures on $\mathbb{R}^d$ with finite second moments. Assume that μ is dominated by ν in convex order, denoted by $\mu \preceq_{\mathrm{c}} \nu$, and meaning that $\int f \, d\mu \leqslant \int f \, d\nu$ holds for all convex functions $f \colon \mathbb{R}^d \to \mathbb{R}$. In the recent articles [BBHK20, CHL21, BBST23, BST23, AMP23], the martingale optimization problem

$$P^{\gamma}(\mu, \nu) \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, \gamma) \, \mu(dx) \tag{1.1}$$

received particular interest. The standard Gaussian measure γ on $\mathbb{R}^d$ is used as reference measure in (1.1) and the cost function is the maximal covariance defined as

$$\mathrm{MCov}(\rho, \varrho) \coloneqq \sup_{\tilde{\pi} \in \mathsf{Cpl}(\rho,\varrho)} \int \langle y, z \rangle \, \tilde{\pi}(dy, dz), \tag{1.2}$$

for $\rho, \varrho \in \mathscr{P}_2(\mathbb{R}^d)$. The maximization in (1.2) takes place over the set $\mathsf{Cpl}(\rho, \varrho)$ of all couplings $\tilde{\pi}$ between ρ and ϱ, i.e., probability measures $\tilde{\pi}$ on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal ρ and second marginal ϱ. In the martingale optimization problem (1.1), the supremum is taken over the collection $\mathsf{MT}(\mu, \nu)$ of martingale transports, i.e., couplings $\pi \in \mathsf{Cpl}(\mu, \nu)$ satisfying $\mathrm{bary}(\pi_x) \coloneqq \int y \, \pi_x(dy) = x$, for μ-a.e. $x \in \mathbb{R}^d$. Here, the family of probability measures $\{\pi_x\}_{x \in \mathbb{R}^d}$ on $\mathbb{R}^d$ is obtained by disintegrating the coupling π with respect to its first marginal μ, i.e., $\pi(dx, dy) = \pi_x(dy) \, \mu(dx)$.

The optimization problem (1.1) is a weak optimal transport problem in the sense of [GRST17], since the function $\pi \mapsto \int \mathrm{MCov}(\pi_x, \gamma) \, \mu(dx)$ is nonlinear, as opposed to classical optimal transport, where linear problems of the form $\pi \mapsto \int c(x, y) \, \pi(dx, dy)$ are studied. The problem (1.1) can also be viewed as a discrete-time version of a continuous-time martingale optimization problem, see [BBHK20, BBST23, BST23]. Related problems also appear in [Loe18, Loe23] and [HT19].

It is natural to consider more general reference measures than the standard Gaussian γ in (1.1). We thus fix a measure $q \in \mathscr{P}_2(\mathbb{R}^d)$ which does not give mass to small sets, i.e., measurable sets with Hausdorff dimension at most $d - 1$ in $\mathbb{R}^d$. This leads to the following extension of (1.1):

$$P^{q}(\mu, \nu) \coloneqq \sup_{\pi \in \mathsf{MT}(\mu,\nu)} \int \mathrm{MCov}(\pi_x, q) \, \mu(dx). \tag{1.3}$$

1.2. Dual formulation

Our first main result is a dual formulation of the martingale optimization problem (1.3) in terms of convex functions.

Theorem 1.1. *Let $\mu, \nu, q \in \mathscr{P}_2(\mathbb{R}^d)$. Assume that $\mu \preceq_{\mathrm{c}} \nu$ and that q does not give mass to small sets. The primal problem (1.3) is uniquely attained. Its value $P^q(\mu,\nu)$ is finite and equal to*

$$D^q(\mu,\nu) := \inf_{\substack{\psi \in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi \, d\nu - \int (\psi^* \star q)^* \, d\mu \Big). \tag{1.4}$$

A few comments on the notation used in the dual problem (1.4) are in order. For a function $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$, its convex conjugate is denoted by ψ^* and given by

$$\psi^*(x) := \sup_{y \in \mathbb{R}^d} \big(\langle x, y \rangle - \psi(y) \big). \tag{1.5}$$

For a probability measure ϱ on $\mathbb{R}^d$, we denote by $\psi \star \varrho$ the function

$$(\psi \star \varrho)(y) := \int \psi(y+z) \, \varrho(dz) \tag{1.6}$$

provided $\psi(y + \cdot\,) \in L^1(\varrho)$. If ϱ is symmetric (like, e.g., in the Gaussian case $\varrho = \gamma$), the function $\psi \star \varrho$ coincides with the convolution of f and ϱ given by

$$(\psi * \varrho)(y) = \int \psi(y-z) \, \rho(dz).$$

In the Gaussian case $q = \gamma$, the duality result of Theorem 1.1 is contained in the statement of Theorem 1.4 in [BBST23]. It turns out that for the proof of this duality result it is not essential that the reference measure q is Gaussian. In fact, it is sufficient that q is a probability measure on $\mathbb{R}^d$ with finite second moments (so that the maximal covariance with respect to q is well-defined) and which does not give mass to small sets (recall that such sets are measurable subsets of $\mathbb{R}^d$ with Hausdorff dimension at most $d-1$). The latter condition enables us to apply Brenier's theorem (see, e.g., [Vil03, Theorem 2.12]).

1.3. q-Bass martingales

We recall the following special construction of Brownian martingales (see [KS98, p. 182] for the classical definition) induced by gradients of convex functions with possibly non-degenerate starting laws.

Definition 1.2. Let $(B_t)_{0 \leqslant t \leqslant 1}$ be d-dimensional Brownian motion with initial distribution $B_0 \sim \hat{\alpha}$, where $\hat{\alpha}$ is an arbitrary probability measure on $\mathbb{R}^d$. Let $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ be a convex function such that $\nabla \hat{v}(B_1)$ is square-integrable. The martingale

$$\hat{M}_t := \mathbb{E}[\nabla \hat{v}(B_1) \,|\, \sigma(B_s \colon s \leqslant t)] = \mathbb{E}[\nabla \hat{v}(B_1) \,|\, B_t], \qquad 0 \leqslant t \leqslant 1 \tag{1.7}$$

is called *Bass martingale* with *Bass measure* $\hat{\alpha}$, and with initial marginal $\mu \coloneqq \mathrm{Law}(\hat{M}_0)$ and terminal marginal $\nu \coloneqq \mathrm{Law}(\hat{M}_1)$.

Martingales of this form where introduced by Bass [Bas83] (in dimension $d = 1$ and with $\hat{\alpha}$ a Dirac measure) in order to derive a solution of the Skorokhod embedding problem. Definition 1.2 generalizes this concept to multiple dimensions and non-degenerate starting laws; see [BBHK20, BBST23].

We write γ^t for the d-dimensional centered Gaussian distribution with covariance matrix tI_d and set $\hat{v}_t \coloneqq \hat{v} * \gamma^{1-t} \colon \mathbb{R}^d \to \mathbb{R}$, for $0 \leqslant t \leqslant 1$. In these terms, (1.7) amounts to

$$\hat{M}_t = \nabla \hat{v}_t(B_t), \qquad 0 \leqslant t \leqslant 1. \tag{1.8}$$

Furthermore, denoting by $*$ the convolution operator (either between a function and a measure or between two measures), we observe that the convex function $\hat{v}$ and the Bass measure $\hat{\alpha}$ from Definition 1.2 satisfy the identities

$$(\nabla \hat{v} * \gamma)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * \gamma) = \nu. \tag{1.9}$$

Motivated by (1.9), it is natural to extend the definition of Bass martingales as follows.

Definition 1.3. Let $\mu, \nu, q \in \mathscr{P}_2(\mathbb{R}^d)$ and assume that q does not give mass to small sets. A pair $(\hat{v}, \hat{\alpha})$ consisting of a convex function $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ and a probability measure $\hat{\alpha}$ on $\mathbb{R}^d$ is called a *q-Bass martingale* with initial marginal μ and terminal marginal ν if

$$(\nabla \hat{v} \star q)(\hat{\alpha}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha} * q) = \nu. \tag{1.10}$$

Recall that the symbol $\star$ is defined as in (1.6). We summarize the fundamental relations (1.10) in the following graphic.

$$\begin{array}{ccc} \hat{\alpha} * q & \xrightarrow{\nabla \hat{v}} & \nu \\ {\scriptstyle *}\uparrow & & \\ \hat{\alpha} & \xrightarrow{\nabla \hat{v} \star q} & \mu \end{array}$$

We note that q-Bass martingales in the sense of Definition 1.3 are discrete-time stochastic processes. In the Gaussian case $q = \gamma$, the corresponding continuous-time martingales can be represented as in (1.7), (1.8). We leave the definition of general q-Bass martingales in continuous time as an open question.

Next, we define an appropriate class of convex functions $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ which generate q-Bass martingales in a natural way.

Definition 1.4. Let $\mu, q \in \mathscr{P}_2(\mathbb{R}^d)$ and assume that q does not give mass to small sets. Let $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ be a convex function and denote by $\hat{\psi} \coloneqq \hat{v}^*$ its convex conjugate. We say that $\hat{v}$ is *q-Bass martingale generating* with initial marginal μ if

(i) $\mu(\mathrm{int}(\mathrm{dom}\, \hat{\psi})) = 1$,

(ii) the function $\hat{v} \star q$ is finite-valued, strictly convex, and $\nabla(\hat{v} \star q) = (\nabla \hat{v}) \star q$,

(iii) the measure $\nu^{\hat{v}} := \nabla \hat{v}(\hat{\alpha}^{\hat{v}} * q)$ has finite second moment, where $\hat{\alpha}^{\hat{v}} := \nabla(\hat{v} \star q)^*(\mu)$.

In our second main result we show that q-Bass martingale generating functions indeed induce q-Bass martingales in the natural way. Beyond that, we establish the connection to the primal problem (1.3) as well as the dual problem (1.4).

Theorem 1.5. *Let $\mu, q \in \mathscr{P}_2(\mathbb{R}^d)$ and assume that q does not give mass to small sets. Let $\hat{v}$ be q-Bass martingale generating with initial marginal μ and let $\nu^{\hat{v}}, \hat{\alpha}^{\hat{v}}$ be as in Definition 1.4.*

(1) The pair $(\hat{v}, \hat{\alpha}^{\hat{v}})$ is a q-Bass martingale with initial marginal μ and with terminal marginal $\nu^{\hat{v}}$.

(2) The optimizer $\hat{\pi}(dx, dy) = \hat{\pi}_x(dy)\, \mu(dx)$ of the primal problem (1.3) between μ and $\nu = \nu^{\hat{v}}$ is given by

$$\hat{\pi}_x = \nabla \hat{v}\Big(\nabla(\hat{v} \star q)^*(x) + \cdot\Big)(q),$$

for μ-a.e. $x \in \mathbb{R}^d$.

(3) The function $\hat{\psi} = \hat{v}^*$ is an optimizer of the dual problem (1.4) between μ and $\nu = \nu^{\hat{v}}$.

The proof of Theorem 1.5 is presented in Section 3.1 of Chapter 3.

Remark 1.6.

(i) The conditions (i) – (iii) of Definition 1.4 do not necessarily imply that the function $\hat{\psi} = \hat{v}^*$ is integrable with respect to the probability measure $\nu^{\hat{v}}$. Therefore attainment of the dual problem (1.4) as claimed in part (3) of Theorem 1.5 has to be understood in a "relaxed" sense frequently encountered in martingale transport problems [BJ16, BNT17, BNS22]; see Definition 3.4 below.

(ii) Already in the Gaussian case $q = \gamma$, the conditions (i) – (iii) of Definition 1.4 are not only sufficient but also necessary for a result in the form of Theorem 1.5 to be true. More precisely, if $q = \gamma$, it follows from the results of [BBST23] that a function $\hat{v}$ satisfying the properties (1) – (3) of Theorem 1.5 has to be γ-Bass martingale generating in the sense of Definition 1.4. ◇

Finally, we raise the question under which assumptions a q-Bass martingale with given marginals μ and ν exists. For γ-Bass martingales this question was answered in [BBST23, Theorem 1.3]: a γ-Bass martingale with marginals μ and ν exists if and only if the pair (μ, ν) is irreducible in the following sense.

Definition 1.7. For probability measures μ, ν we say that the pair (μ, ν) is *irreducible* if for all measurable sets $A, B \subseteq \mathbb{R}^d$ with $\mu(A) > 0$ and $\nu(B) > 0$ there is a martingale $(X_t)_{0 \leqslant t \leqslant 1}$ with $X_0 \sim \mu$, $X_1 \sim \nu$ such that $\mathbb{P}(X_0 \in A, X_1 \in B) > 0$.

Given an irreducible pair (μ, ν), we conjecture that a q-Bass martingale with marginals μ and ν exists, provided that q is absolutely continuous and satisfies suitable integrability conditions.

1.4. Related literature

The formulation of optimal transport as a field in mathematics goes back to Monge [Mon81] and Kantorovich [Kan42]. The modern theory is based on the seminal contributions of Benamou, Brenier, and McCann [Bre87, Bre91, McC94, McC95, BB99]. We refer to the books [Vil03, Vil09, AG13, San15] for introductions to the field of optimal transport as well as a variety of applications in different areas.

In martingale optimal transport (see e.g. [HN12, BHLP13, TT13, DS14, GHLT14, CLM17] among many others), the transport plan satisfies an additional martingale constraint. This additional requirement leads to notable consequences in mathematical finance [BHLP13], the study of martingale inequalities [BN15, OST15, HLOST16], and the Skorokhod embedding problem [BCH17, KTT17, BNS22]. For structural descriptions of optimal martingale transport plans we refer to [BJ16, OS17, DMT19, GKL19, Cio23a, Cio23b].

Interesting problems in the field of martingale optimal transport are formulated in continuous time, see e.g. [DS14, BHLT17, COT19, GKP19, GLW19, GKL20, CKPS21, GL21]. Most notably, the martingale version of the Benamou–Brenier problem was introduced in [BBHK20] in probabilistic language and in [HT19] in PDE language. It was then analyzed from the point of view of duality theory in [BBST23] and complemented by a variational perspective in [BST23]. In the context of market impact in finance, the same kind of problem appeared independently in [Loe18, Loe23]. For the interpretation of the martingale Benamou–Brenier problem as a local volatility model we refer to the recent articles [BBHK20, CHL21, AMP23].

2. Proof of the duality result

Throughout the rest of this chapter we fix $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$ with $\mu \preceq_{\mathrm{c}} \nu$. We also fix a reference measure $q \in \mathscr{P}_2(\mathbb{R}^d)$ which does not give mass to small sets. The goal of this chapter is to provide the proof of the duality result as stated in Theorem 1.1. The proof is based on [BBST23, Theorem 1.4], which is formulated in the Gaussian setting $q = \gamma$.

2.1. A first dual formulation

To formulate a first version of a dual problem to (1.3), we consider the set of continuous functions with quadratic growth

$$C_2(\mathbb{R}^d) \coloneqq \{\psi \colon \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \exists\, a, k, \ell \in \mathbb{R} \text{ with } \ell + \tfrac{|\cdot|^2}{2} \leqslant \psi(\,\cdot\,) \leqslant a + k|\cdot|^2\}.$$

We define the dual problem

$$\tilde{D}^q(\mu, \nu) \coloneqq \inf_{\psi \in C_2(\mathbb{R}^d)} \Big(\int \psi \, d\nu - \int \varphi^\psi \, d\mu \Big), \tag{2.1}$$

where the function $\mathbb{R}^d \ni x \longmapsto \varphi^\psi(x)$ is given by

$$\varphi^\psi(x) \coloneqq \inf_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int \psi \, dp - \mathrm{MCov}(p, q) \Big). \tag{2.2}$$

Here, the set $\mathscr{P}_2^x(\mathbb{R}^d)$ denotes the elements p of $\mathscr{P}_2(\mathbb{R}^d)$ with barycenter

$$\mathrm{bary}(p) = \int y \, p(dy) = x \in \mathbb{R}^d.$$

Proposition 2.1. *There is no duality gap between the primal problem (1.3) and the dual problem (2.1), i.e., $P^q(\mu, \nu) = \tilde{D}^q(\mu, \nu)$. Moreover, the primal problem is uniquely attained and has a finite value, i.e., there exists a unique $\hat{\pi} \in \mathsf{MT}(\mu, \nu)$ such that*

$$P^q(\mu, \nu) = \int \mathrm{MCov}(\hat{\pi}_x, q) \, \mu(dx) < +\infty. \tag{2.3}$$

Proof. For $x \in \mathbb{R}^d$ and $p \in \mathscr{P}_2(\mathbb{R}^d)$ we define the cost function

$$C(x, p) \coloneqq \begin{cases} -\mathrm{MCov}(p, q) + \frac{1}{2} \int |y|^2 \, dp, & \text{if } \mathrm{bary}(p) = x, \\ +\infty, & \text{if } \mathrm{bary}(p) \neq x, \end{cases}$$

which is bounded from below and convex in the second argument. If we equip $\mathscr{P}_2(\mathbb{R}^d)$ with the topology induced by the quadratic Wasserstein distance, one verifies that the

function $(x,p) \mapsto C(x,p)$ is jointly lower semicontinuous with respect to the product topology on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$. Hence we can apply [BBP19, Theorem 1.2], which guarantees the existence of an optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (1.3). This optimizer is unique, as the value $P^q(\mu,\nu)$ is finite and $(x,p) \mapsto C(x,p)$ is strictly convex in the second argument. The former follows from the fact that μ and ν have finite second moments, the latter is a consequence of Brenier's theorem.

In order to show that there is no duality gap, we introduce the space of continuous functions which are bounded from below and have at most quadratic growth

$$C_{\mathrm{b},2}(\mathbb{R}^d) \coloneqq \{\tilde{\psi}\colon \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \exists\, a,k,\ell \in \mathbb{R} \text{ with } \ell \leqslant \tilde{\psi}(\,\cdot\,) \leqslant a + k|\cdot|^2\}.$$

Then by [BBP19, Theorem 1.3] the value $P^q(\mu,\nu)$ of the primal problem is equal to

$$\tilde{D}^q_{\mathrm{b},2}(\mu,\nu) \coloneqq \inf_{\tilde{\psi} \in C_{\mathrm{b},2}(\mathbb{R}^d)} \Big(\int (\tilde{\psi}(\,\cdot\,) + \tfrac{|\cdot|^2}{2})\, d\nu - \int \bar{\varphi}^{\tilde{\psi}}\, d\mu \Big),$$

where

$$\bar{\varphi}^{\tilde{\psi}}(x) \coloneqq \inf_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int (\tilde{\psi}(\,\cdot\,) + \tfrac{|\cdot|^2}{2})\, dp - \mathrm{MCov}(p,q) \Big).$$

Finally, passing from the functions $\tilde{\psi} \in C_{\mathrm{b},2}(\mathbb{R}^d)$ to $\psi(\,\cdot\,) \coloneqq \tilde{\psi}(\,\cdot\,) + \frac{|\cdot|^2}{2} \in C_2(\mathbb{R}^d)$, we see that $\tilde{D}^q_{\mathrm{b},2}(\mu,\nu) = \tilde{D}^q(\mu,\nu)$. □

We close this section by providing a heuristic argument why (2.1) is indeed a plausible dual formulation of (1.3). Considering the "indicator function"

$$\chi(\pi) \coloneqq \inf_{\psi \in C_2(\mathbb{R}^d)} \Big(\int \psi\, d\nu - \int \psi\, d\pi_x\, d\mu(x) \Big) = \begin{cases} 0, & \text{if } \int \pi_x\, d\mu(x) = \nu, \\ -\infty, & \text{else,} \end{cases}$$

we formally obtain the desired duality relation by interchanging inf and sup, to wit

$$\begin{aligned} P^q(\mu,\nu) &= \sup_{\substack{\pi(dx,dy) = \pi_x(dy)\,\mu(dx),\\ \pi_x \in \mathscr{P}_2^x(\mathbb{R}^d)}} \Big(\int \mathrm{MCov}(\pi_x,q)\, d\mu(x) + \chi(\pi) \Big) \\ &= \inf_{\psi \in C_2(\mathbb{R}^d)} \Big(\int \sup_{\pi_x \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(\pi_x,q) + \int \psi\, d(\nu - \pi_x) \Big)\, d\mu(x) \Big) = \tilde{D}^q(\mu,\nu). \end{aligned}$$

2.2. Convexity of the dual variables

In this section we make the crucial observation that in the dual problem (2.1) it suffices to optimize over the class of functions $\psi \in C_2(\mathbb{R}^d)$ which are convex. We then show that it is also equivalent to optimize over all convex functions $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ which are only μ-a.s. finite, but not necessarily of quadratic growth.

Given some $\psi \in C_2(\mathbb{R}^d)$, we denote by $\operatorname{conv}\psi$ the convex hull of ψ, i.e., the greatest convex function smaller or equal to ψ. It will be convenient to have an explicit representation for the convex hull of a function, as in (2.4) below. This identity is usually stated in the more specific form

$$(\operatorname{conv}\psi)(y) = \inf\Big\{\sum_{i=1}^{d+1}\lambda_i\psi(y_i)\colon \sum_{i=1}^{d+1}\lambda_i y_i = y\Big\},$$

where the infimum is taken over all expressions of y as a convex combination of $d+1$ points, see [Roc70, Corollary 17.1.5].

Lemma 2.2. *For a function $\psi \in C_2(\mathbb{R}^d)$ its convex hull $\operatorname{conv}\psi$ is given by*

$$\mathbb{R}^d \ni y \longmapsto (\operatorname{conv}\psi)(y) = \inf_{p\in\mathscr{P}_2^y(\mathbb{R}^d)}\int \psi\,dp \tag{2.4}$$

and again $\operatorname{conv}\psi \in C_2(\mathbb{R}^d)$.

Recalling the dual problem (2.1), we define the dual function

$$C_2(\mathbb{R}^d) \ni \psi \longmapsto \tilde{\mathcal{D}}(\psi) := \int \psi\,d\nu - \int \varphi^\psi\,d\mu. \tag{2.5}$$

Now we prove that it suffices to optimize the dual function over the class of functions $\psi \in C_2(\mathbb{R}^d)$ which are convex.

Proposition 2.3. *For all $\psi \in C_2(\mathbb{R}^d)$ we have $\tilde{\mathcal{D}}(\operatorname{conv}\psi) \leqslant \tilde{\mathcal{D}}(\psi)$ and consequently*

$$\tilde{D}^q(\mu,\nu) = \inf_{\psi\in C_2(\mathbb{R}^d)}\tilde{\mathcal{D}}(\psi) = \inf_{\substack{\psi\in C_2(\mathbb{R}^d),\\ \psi \text{ convex}}}\tilde{\mathcal{D}}(\psi). \tag{2.6}$$

Proof. Let $\varepsilon > 0$, $\psi \in C_2(\mathbb{R}^d)$ and $\{p_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ be a measurable collection of probability measures with $\operatorname{bary}(p_x) = x$. To show the claim, it is sufficient to construct a measurable family $\{\bar{p}_x\}_{x\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ with $\operatorname{bary}(\bar{p}_x) = x$ such that

$$\operatorname{MCov}(p_x,q) + \int \operatorname{conv}\psi\,d(\nu - p_x) \leqslant \operatorname{MCov}(\bar{p}_x,q) + \int \psi\,d(\nu-\bar{p}_x) + \varepsilon. \tag{2.7}$$

By Lemma 2.2 and a measurable selection argument we can choose a measurable collection of probability measures $\{\tilde{p}_y\}_{y\in\mathbb{R}^d} \subseteq \mathscr{P}_2(\mathbb{R}^d)$ with $\operatorname{bary}(\tilde{p}_y) = y$ such that

$$\int \psi\,d\tilde{p}_y \leqslant (\operatorname{conv}\psi)(y) + \varepsilon. \tag{2.8}$$

Then we define $\bar{p}_x(dz) := \int_y \tilde{p}_y(dz)\,p_x(dy)$, so that $\operatorname{bary}(\bar{p}_x) = x$. Integrating (2.8) with respect to $p_x(dy)$ yields

$$\int \psi\,d\bar{p}_x \leqslant \int \operatorname{conv}\psi\,dp_x + \varepsilon. \tag{2.9}$$

Since $\psi, \operatorname{conv}\psi \in C_2(\mathbb{R}^d)$ and $p_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ we conclude

$$\ell + \tfrac{1}{2}\int |y|^2\, \bar{p}_x(dy) \leqslant a + k\int |y|^2\, p_x(dy) + \varepsilon < +\infty,$$

so that $\bar{p}_x \in \mathscr{P}_2^x(\mathbb{R}^d)$.

In order to show the inequality (2.7), we first observe that $p_x \preceq_{\mathrm{c}} \bar{p}_x$ by Jensen's inequality. Together with the Kantorovich duality (see, e.g. [Vil09, Theorem 5.10]), we conclude that[1]

$$\begin{aligned}\operatorname{MCov}(p_x, q) &= \inf_{\substack{f\colon \mathbb{R}^d\to\mathbb{R}\\ \text{convex}}} \Big(\int f\, dp_x + \int f^*\, dq\Big)\\ &\leqslant \inf_{\substack{f\colon \mathbb{R}^d\to\mathbb{R}\\ \text{convex}}} \Big(\int f\, d\bar{p}_x + \int f^*\, dq\Big) = \operatorname{MCov}(\bar{p}_x, q).\end{aligned} \tag{2.10}$$

On the other hand, from $\operatorname{conv}\psi \leqslant \psi$ and (2.9) we have the inequality

$$\int \operatorname{conv}\psi\, d(\nu - p_x) \leqslant \int \psi\, d(\nu - \bar{p}_x) + \varepsilon. \tag{2.11}$$

Finally, summing (2.10) and (2.11), we obtain the inequality (2.7). □

Next, we extend the definition (2.5) of the dual function $\tilde{\mathcal{D}}(\,\cdot\,)$ to convex functions $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ which are not confined to be in $C_2(\mathbb{R}^d)$, but which are only required to satisfy $\mu(\operatorname{dom}\psi) = 1$, where

$$\operatorname{dom}\psi := \{y \in \mathbb{R}^d\colon \psi(y) < +\infty\}$$

denotes the domain of ψ. In order for the difference of the integrals in (2.5) still to be well-defined, the dual function has to be understood in a "relaxed" sense, as frequently encountered in martingale transport problems; see [BJ16, BNT17, BNS22]. More precisely, we define the "relaxed" representation of (2.5) by

$$\tilde{\mathcal{E}}(\psi) := \int \Big(\int \psi(y)\, \pi_x(dy) - \varphi^{\psi}(x)\Big)\, \mu(dx), \tag{2.12}$$

where $\pi(dx, dy) = \pi_x(dy)\, \mu(dx)$ is an arbitrary fixed element of $\mathsf{MT}(\mu, \nu)$. Then $\tilde{\mathcal{E}}(\psi)$ is well-defined for every convex function ψ satisfying $\mu(\operatorname{dom}\psi) = 1$. Taking the infimum of (2.12) over all such ψ again leads to the same value as in (2.6). This is summarized in the following lemma.

Lemma 2.4. *Fix $\pi \in \mathsf{MT}(\mu, \nu)$ and let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a convex function with $\mu(\operatorname{dom}\psi) = 1$. Then $\tilde{\mathcal{E}}(\psi) \in [0, +\infty]$ and we have the inequality*

$$\tilde{D}^q(\mu, \nu) \leqslant \tilde{\mathcal{E}}(\psi). \tag{2.13}$$

In particular, recalling (2.6), we have

$$\tilde{D}^q(\mu, \nu) = \inf_{\substack{\psi\in C_2(\mathbb{R}^d),\\ \psi \text{ convex}}} \tilde{\mathcal{D}}(\psi) = \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \tilde{\mathcal{E}}(\psi). \tag{2.14}$$

[1] In fact, $p_x \preceq_{\mathrm{c}} \bar{p}_x$ is equivalent to the inequality $\operatorname{MCov}(p_x, q) \leqslant \operatorname{MCov}(\bar{p}_x, q)$ being valid for all probability measures $q \in \mathscr{P}_2(\mathbb{R}^d)$, see. e.g. [AP22, Theorem 1] or [WZ23, Corollary 1.2].

Proof. Fix $\pi \in \mathsf{MT}(\mu,\nu)$ and let $\psi\colon \mathbb{R}^d \to (-\infty,+\infty]$ be a convex function with $\mu(\operatorname{dom}\psi) = 1$. First, note that by Jensen's inequality we have

$$\int \psi(y)\,\pi_x(dy) \geqslant \psi\Big(\int y\,\pi_x(dy)\Big) = \psi(x),$$

and by taking $p = \delta_x$ in (2.2) we obtain $\varphi^\psi(x) \leqslant \psi(x)$. Hence

$$\int \psi(y)\,\pi_x(dy) - \varphi^\psi(x) \geqslant 0, \tag{2.15}$$

for μ-a.e. $x \in \mathbb{R}^d$, so that $\tilde{\mathcal{E}}(\psi) \in [0,+\infty]$. In order to prove the inequality (2.13), we distinguish two cases. In the case $\int \psi\,d\pi_x = +\infty$, for $x \in \mathbb{R}^d$ in a set of positive μ-measure, we conclude from (2.15) that $\tilde{\mathcal{E}}(\psi) = +\infty$, so that (2.13) is trivially satisfied. Now suppose that $\int \psi\,d\pi_x < +\infty$, for μ-a.e. $x \in \mathbb{R}^d$. Recall from Proposition 2.1 that there exists an optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ of the primal problem (1.3). Taking $p = \hat{\pi}_x$ in (2.2) we get

$$\varphi^\psi(x) \leqslant \int \psi\,d\hat{\pi}_x - \mathrm{MCov}(\hat{\pi}_x, q)$$

and therefore

$$\tilde{\mathcal{E}}(\psi) \geqslant \int \Big(\int \psi\,d\pi_x + \mathrm{MCov}(\hat{\pi}_x,q) - \int \psi\,d\hat{\pi}_x\Big)\,\mu(dx) = P^q(\mu,\nu),$$

where we also used that both $\hat{\pi}$ and π have ν as second marginal, and that $\hat{\pi}$ is a primal optimizer, i.e., satisfying (2.3). Since $P^q(\mu,\nu) = \tilde{D}^q(\mu,\nu)$ by Proposition 2.1, we again see that (2.13) does hold.

Finally, from (2.13) and recalling (2.6), we conclude

$$\tilde{D}^q(\mu,\nu) \leqslant \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \tilde{\mathcal{E}}(\psi) \leqslant \inf_{\substack{\psi \in C_2(\mathbb{R}^d),\\ \psi \text{ convex}}} \tilde{\mathcal{D}}(\psi) = \tilde{D}^q(\mu,\nu),$$

which shows (2.14). □

2.3. Proof of Theorem 1.1

According to Proposition 2.1, there is no duality gap between the primal problem (1.3) and the dual problem (2.1), i.e., $P^q(\mu,\nu) = \tilde{D}^q(\mu,\nu)$. Hence for the proof of Theorem 1.1 we have to show that

$$\tilde{D}^q(\mu,\nu) = \inf_{\substack{\psi \in L^1(\nu),\\ \psi \text{ convex}}} \Big(\int \psi\,d\nu - \int (\psi^* \star q)^*\,d\mu\Big). \tag{2.16}$$

Recalling the definitions (2.1), (2.2), and (2.5), it should come as no surprise that the main idea behind the proof of the identity (2.16) is to apply Proposition 2.3 and to show that $\varphi^\psi = (\psi^* \star q)^*$ holds for every convex function $\psi \in C_2(\mathbb{R}^d)$. This motivates

our next goal, namely to solve the minimization problem (2.2), which we rewrite as a maximization problem, to wit

$$-\varphi^{\psi}(x) = \sup_{p \in \mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,q) - \int \psi \, dp \Big). \tag{2.17}$$

As a preliminary step, we consider the simpler problem

$$\phi^{\psi} := \sup_{p \in \mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,q) - \int \psi \, dp \Big), \tag{2.18}$$

where we do not prescribe the barycenter $x \in \mathbb{R}^d$ of $p \in \mathscr{P}_2(\mathbb{R}^d)$. In Lemma 2.6 below we will show for an arbitrary proper convex function $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$, that the value ϕ^{ψ} of (2.18) is equal to $\int \psi^* \, dq$. Recall that ψ is called a proper convex function if ψ is convex and $\operatorname{dom} \psi \neq \varnothing$.

Solving the maximization problem (2.18) leads to an interesting connection with Brenier maps, see Lemma 2.5 below. By Brenier's theorem (see, e.g., [Vil03, Theorem 2.12]), the optimal transport for quadratic cost between $q \in \mathscr{P}_2(\mathbb{R}^d)$ and $p \in \mathscr{P}_2(\mathbb{R}^d)$ is induced by the q-a.e. defined gradient ∇v (the "Brenier map") of some convex function $v \colon \mathbb{R}^d \to \mathbb{R}$ via $(\nabla v)(q) = p$.

Lemma 2.5. *Let $v \colon \mathbb{R}^d \to \mathbb{R}$ be a finite-valued convex function and $\psi := v^*$ its convex conjugate. Assume that the probability measure $\hat{p} := (\nabla v)(q)$ has finite second moment. Then $\hat{p}$ is the unique maximizer of the optimization problem (2.18) and*

$$\phi^{\psi} = \int v \, dq = \int \psi^* \, dq < +\infty.$$

Proof. For $p \in \mathscr{P}_2(\mathbb{R}^d)$, we denote by T_q^p the Brenier map from q to $p \in \mathscr{P}_2(\mathbb{R}^d)$ and note that $T_q^{\hat{p}} = \nabla v$ by Brenier's theorem. Furthermore, we have that

$$v(z) = \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) = \big\langle \nabla v(z), z \big\rangle - v^*\big(\nabla v(z)\big),$$

for q-a.e. $z \in \mathbb{R}^d$. Using these observations, for $p \in \mathscr{P}_2(\mathbb{R}^d)$ we get

$$\begin{aligned}
\mathrm{MCov}(p,q) - \int \psi \, dp &= \int \Big(\big\langle T_q^p(z), z \big\rangle - v^*\big(T_q^p(z)\big) \Big) \, q(dz) \\
&\leqslant \int \sup_{y \in \mathbb{R}^d} \big(\langle y, z \rangle - v^*(y) \big) \, q(dz) = \int v(z) \, q(dz) \\
&= \int \Big(\big\langle \nabla v(z), z \big\rangle - v^*\big(\nabla v(z)\big) \Big) \, q(dz) \\
&= \mathrm{MCov}(\hat{p}, q) - \int \psi \, d\hat{p},
\end{aligned}$$

with equality if and only if $T_q^p(z) = T_q^{\hat{p}}(z)$, for q-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p = \hat{p}$. Finally, from the convexity of v and the Cauchy–Schwarz inequality we obtain

$$\int |v| \, dq \leqslant |v(0)| + \sqrt{\int |\nabla v|^2 \, dq} \, \sqrt{\int |z|^2 \, dq(z)} < +\infty,$$

which proves that $\phi^\psi < +\infty$. □

Next, we relax the rather strong assumptions of Lemma 2.5. If we are given just a proper convex function $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$, we can still compute the value of the supremum in (2.18), without explicitly constructing a maximizer of this optimization problem.

Lemma 2.6. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper convex function. Then*

$$\phi^\psi = \int \psi^* \, dq. \tag{2.19}$$

Proof. Using probabilistic notation, we rewrite the supremum in (2.18) as

$$\phi^\psi = \sup \mathbb{E}\big[\langle Y, Z\rangle - \psi(Y)\big], \tag{2.20}$$

where the supremum in (2.20) is taken over all probability spaces such that $Z \sim q$ and Y is an $\mathbb{R}^d$-valued random variable with finite second moment. Replacing Y by $\mathbb{E}[Y \mid Z]$, we observe that the maximization in (2.20) can be restricted to random variables Y which are measurable functions of Z, and we obtain

$$\phi^\psi = \sup_{Y \in L^2(q;\mathbb{R}^d)} \int \Big(\langle Y(z), z\rangle - \psi\big(Y(z)\big)\Big)\, q(dz), \tag{2.21}$$

where $L^2(q;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are square-integrable under q. Clearly, for any $Y \in L^2(q;\mathbb{R}^d)$ we have

$$\int \Big(\langle Y(z), z\rangle - \psi\big(Y(z)\big)\Big)\, q(dz) \leqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, q(dz) = \int \psi^*(z)\, q(dz),$$

which shows the inequality $\phi^\psi \leqslant \int \psi^* \, dq$. In order to see the reverse inequality, we define the auxiliary problem

$$\phi^\psi_\infty := \sup_{Y \in L^\infty(q;\mathbb{R}^d)} \int \Big(\langle Y(z), z\rangle - \psi\big(Y(z)\big)\Big)\, q(dz), \tag{2.22}$$

where $L^\infty(q;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued Borel measurable functions on $\mathbb{R}^d$, which are bounded q-a.e. Comparing (2.21) with (2.22), we obviously have $\phi^\psi \geqslant \phi^\psi_\infty$. Now we claim that

$$\phi^\psi_\infty \geqslant \int \sup_{y\in\mathbb{R}^d} \big(\langle y, z\rangle - \psi(y)\big)\, q(dz) = \int \psi^*(z)\, q(dz), \tag{2.23}$$

which will finish the proof of (2.19). To see this, we first write

$$\phi^\psi_\infty = \lim_{N\to\infty} \sup_{\substack{Y \in L^\infty(q;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\langle Y(z), z\rangle - \psi\big(Y(z)\big)\Big)\, q(dz).$$

Using a measurable selection argument, we obtain

$$\sup_{\substack{Y\in L^\infty(q;\mathbb{R}^d),\\ |Y|\leqslant N}} \int \Big(\langle Y(z),z\rangle - \psi\big(Y(z)\big)\Big)\, q(dz) \geqslant \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big)\, q(dz).$$

Since ψ is proper we can choose $y_0 \in \operatorname{dom}\psi \neq \varnothing$. Then for N large enough we have

$$\sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big) \geqslant \langle y_0,z\rangle - \psi(y_0),$$

with the right-hand side being integrable with respect to $q(dz)$. Hence we can apply the monotone convergence theorem and deduce that

$$\lim_{N\to\infty} \int \sup_{\substack{y\in\mathbb{R}^d,\\ |y|\leqslant N}} \big(\langle y,z\rangle - \psi(y)\big)\, q(dz) = \int \sup_{y\in\mathbb{R}^d} \big(\langle y,z\rangle - \psi(y)\big)\, q(dz),$$

which completes the proof of (2.23). □

Lemma 2.6 is an important auxiliary result for the proof of Theorem 1.1, more precisely, for the proof of the identity (2.16). We first show a simpler variant of this identity, where we optimize over the class of convex functions ψ in $C_2(\mathbb{R}^d)$ (i.e., which have quadratic growth), as in (2.24) of Proposition 2.7 below.

Proposition 2.7. *The value $\tilde{D}^q(\mu,\nu)$ of the dual problem (2.1) is equal to*

$$D_2^q(\mu,\nu) \coloneqq \inf_{\substack{\psi\in C_2(\mathbb{R}^d),\\ \psi \text{ convex}}} \Big(\int \psi\, d\nu - \int (\psi^* \star q)^*\, d\mu\Big). \tag{2.24}$$

Proof. By Proposition 2.3 we have

$$\tilde{D}^q(\mu,\nu) = \inf_{\substack{\psi\in C_2(\mathbb{R}^d),\\ \psi \text{ convex}}} \Big(\int \psi\, d\nu - \int \varphi^\psi\, d\mu\Big).$$

Hence it remains to show that $\varphi^\psi = (\psi^* \star q)^*$, for every convex function $\psi \in C_2(\mathbb{R}^d)$. In order to do this, we will first prove that $(\varphi^\psi)^* = \psi^* \star q$. By definition of the convex conjugate (1.5) and of the maximization problem (2.17), for $y \in \mathbb{R}^d$, we have

$$\begin{aligned}
(\varphi^\psi)^*(y) &= \sup_{x\in\mathbb{R}^d} \big(\langle x,y\rangle - \varphi^\psi(x)\big) && (2.25)\\
&= \sup_{x\in\mathbb{R}^d} \sup_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,q) - \int \psi_y\, dp\Big) && (2.26)\\
&= \sup_{p\in\mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,q) - \int \psi_y\, dp\Big), && (2.27)
\end{aligned}$$

where the function ψ_y is defined by $\psi_y(z) \coloneqq \psi(z) - \langle y, z\rangle$, for $z \in \mathbb{R}^d$. Now applying Lemma 2.6 to the proper convex function ψ_y yields

$$(\varphi^\psi)^*(y) = \phi^{\psi_y} = \int \psi_y^* \, dq = \int \psi^*(y+z)\, dq(z) = (\psi^* \star q)(y), \tag{2.28}$$

where for the last equality we recall the definition (1.6). To complete the proof, we must justify that $(\varphi^\psi)^{**} = \varphi^\psi$, i.e., that the Fenchel–Moreau theorem is applicable. To see this, we first note that the function $x \mapsto \varphi^\psi(x)$ is convex and we have the upper bound $\varphi^\psi \leqslant \psi < +\infty$. Furthermore, for any $p \in \mathscr{P}_2^x(\mathbb{R}^d)$ we have the inequalities

$$\mathrm{MCov}(p,q) \leqslant \tfrac{1}{2}\int |y|^2 \, dp(y) + \tfrac{1}{2}\int |z|^2 \, dq(z) < +\infty$$

as well as

$$\int \psi \, dp \geqslant \ell + \tfrac{1}{2}\int |y|^2 \, dp(y) > -\infty,$$

the latter following from the fact that $\psi \in C_2(\mathbb{R}^d)$. As a consequence, we get the lower bound

$$\varphi^\psi \geqslant \ell - \tfrac{1}{2}\int |z|^2 \, dq(z) > -\infty.$$

Altogether, φ^ψ is a convex function which is finite everywhere on $\mathbb{R}^d$, thus it is continuous and we indeed have $(\varphi^\psi)^{**} = \varphi^\psi$. □

Before we finally turn to the proof of Theorem 1.1 at the end of this section, we formulate a "relaxed" version of (1.4), (2.24). Similarly as in Lemma 2.4 we now admit convex functions $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ with $\mu(\operatorname{dom}\psi) = 1$ as dual variables, see (2.29) of Proposition 2.8 below.

Proposition 2.8. *The value $\tilde{D}^q(\mu,\nu)$ of the dual problem (2.1) is equal to*

$$D^q_{\mathrm{rel}}(\mu,\nu) \coloneqq \inf_{\substack{\mu(\operatorname{dom}\psi)=1,\\ \psi \text{ convex}}} \mathcal{E}(\psi), \tag{2.29}$$

where the dual function $\psi \mapsto \mathcal{E}(\psi)$ is defined by

$$\mathcal{E}(\psi) \coloneqq \int \Big(\int \psi(y)\, \pi_x(dy) - (\psi^* \star q)^*(x) \Big)\, \mu(dx), \tag{2.30}$$

with π being an arbitrary fixed element of $\mathsf{MT}(\mu,\nu)$.

Proof. We start with showing that the dual function $\mathcal{E}(\,\cdot\,)$ given in (2.30) is well-defined for every convex function $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ which is μ-a.s. finite. To this end, we first prove the inequality $(\psi^* \star q)^* \leqslant \psi$. Since ψ is convex with $\mu(\operatorname{dom}\psi) = 1$, for every $y \in \mathbb{R}^d$, the function

$$\mathbb{R}^d \ni z \longmapsto \psi_y(z) \coloneqq \psi(z) - \langle y, z\rangle \tag{2.31}$$

is a proper convex function, so that as in the proof of Proposition 2.7 above we can apply Lemma 2.6. In particular, recalling the equation (2.28), we again have $\psi^* \star q = (\varphi^\psi)^*$. Taking the convex conjugate and using that $\varphi^\psi \leqslant \psi$, we obtain

$$(\psi^* \star q)^* = (\varphi^\psi)^{**} \leqslant \varphi^\psi \leqslant \psi, \tag{2.32}$$

as required. Applying Jensen's inequality as in the proof of Lemma 2.4, this implies that the integrand in (2.30) is μ-a.s. nonnegative, and hence $\mathcal{E}(\psi)$ is well-defined and takes values in the interval $[0, +\infty]$.

Now let us turn to the proof of $\tilde{D}^q(\mu, \nu) = D^q_{\mathrm{rel}}(\mu, \nu)$. Recalling (2.29), (2.24), and using Proposition 2.7, we have

$$D^q_{\mathrm{rel}}(\mu, \nu) \leqslant D^q_2(\mu, \nu) = \tilde{D}^q(\mu, \nu),$$

so that we need to show the inequality $\tilde{D}^q(\mu, \nu) \leqslant D^q_{\mathrm{rel}}(\mu, \nu)$. Let $\pi \in \mathsf{MT}(\mu, \nu)$ and $\psi \colon \mathbb{R}^d \to (-\infty, +\infty]$ be convex with $\mu(\operatorname{dom} \psi) = 1$. Since $\tilde{D}^q(\mu, \nu) = P^q(\mu, \nu)$ (recall Proposition 2.1 and (2.3)), we have to verify the inequality

$$\int \mathrm{MCov}(\pi_x, q)\, d\mu(x) \leqslant \int \Big(\int \psi(y)\, \pi_x(dy) - (\psi^* \star q)^*(x) \Big)\, \mu(dx).$$

It is sufficient to prove, for μ-a.e. $x \in \mathbb{R}^d$, that

$$\mathrm{MCov}(\pi_x, q) \leqslant \int \psi\, d\pi_x - (\psi^* \star q)^*(x). \tag{2.33}$$

We express the convex conjugate on the right-hand side of (2.33) as

$$-(\psi^* \star q)^*(x) = \inf_{y \in \mathbb{R}^d} \big((\psi^* \star q)(y) - \langle x, y \rangle \big) = \inf_{y \in \mathbb{R}^d} \Big(\int \psi^*_y\, dq - \langle x, y \rangle \Big);$$

for the last equality we again relied on the equation (2.28), and the function ψ_y is defined as in (2.31). Substituting back into (2.33) yields

$$\mathrm{MCov}(\pi_x, q) \leqslant \inf_{y \in \mathbb{R}^d} \Big(\int \psi_y\, d\pi_x + \int \psi^*_y\, dq \Big). \tag{2.34}$$

This is what we need to show to complete the proof of Proposition 2.8. Now observe that by the Fenchel–Young inequality we have

$$\mathrm{MCov}(\pi_x, q) \leqslant \int f\, d\pi_x + \int f^*\, dq,$$

for every proper convex function $f \colon \mathbb{R}^d \to (-\infty, +\infty]$. In particular, for every $y \in \mathbb{R}^d$, it holds that

$$\mathrm{MCov}(\pi_x, q) \leqslant \int \psi_y\, d\pi_x + \int \psi^*_y\, dq,$$

which implies (2.34), as required. □

Finally, we give the proof of Theorem 1.1, where we optimize over all ν-integrable convex functions ψ as in (1.4), (2.16).

Proof of Theorem 1.1. We first show that the integral

$$\int (\psi^* \star q)^* \, d\mu \tag{2.35}$$

appearing on the right-hand side of (1.4), (2.16) is well-defined for every convex function $\psi \in L^1(\nu)$. To see this, recall from (2.32) above that $(\psi^* \star q)^* \leqslant \psi$. On the other hand, since $\mu \preceq_{\mathrm{c}} \nu$ and $\psi \in L^1(\nu)$ is convex, the integral of ψ with respect to μ exists. We conclude that also the integral (2.35) exists.

Now let us turn to the actual proof of Theorem 1.1. Recalling the beginning of Section 2.3, we have to prove the equality (2.16). But this assertion follows immediately from Proposition 2.8 and Proposition 2.7 by a "sandwich argument". Indeed, we have

$$\tilde{D}^q(\mu,\nu) = D^q_{\mathrm{rel}}(\mu,\nu) \leqslant D^q(\mu,\nu) \leqslant D^q_2(\mu,\nu) = \tilde{D}^q(\mu,\nu). \tag{2.36}$$

The inequalities in (2.36) are due to the inclusions

$$\{\psi \in C_2(\mathbb{R}^d)\colon \psi \text{ convex}\} \subseteq \{\psi \in L^1(\nu)\colon \psi \text{ convex}\} \subseteq \{\psi \text{ convex}\colon \mu(\operatorname{dom}\psi) = 1\};$$

the equalities on the left-hand side and on the right-hand side of (2.36) are justified by Proposition 2.8 and Proposition 2.7, respectively.

Finally, we recall that existence and uniqueness of the optimizer $\hat{\pi} \in \mathsf{MT}(\mu,\nu)$ as well as finiteness of $P^q(\mu,\nu)$ have already been shown in Proposition 2.1. □

3. Existence and properties of q-Bass martingales

Throughout this chapter we fix $\mu, q \in \mathscr{P}_2(\mathbb{R}^d)$ and assume that q does not give mass to small sets. Our goal is to prove our second main result, Theorem 1.5. Its proof is based on the work [BBST23], where the Gaussian case $q = \gamma$ is studied in detail. First, we need some auxiliary results.

Lemma 3.1. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a proper, lower semicontinuous convex function.*

(i) The function $x \mapsto \varphi^\psi(x)$ defined in (2.2) is convex on $\operatorname{dom}\psi$.

(ii) If $\varphi^\psi(x) > -\infty$, for some $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\varphi^\psi > -\infty$ on $\operatorname{int}(\operatorname{dom}\psi)$.

Proof. (i) To see that the function φ^ψ is convex on $\operatorname{dom}\psi$, we let $x \coloneqq \lambda x_1 + (1-\lambda)x_2$, for $x_1, x_2 \in \operatorname{dom}\psi$ and $\lambda \in (0,1)$. If $\varphi^\psi(x_1) = -\infty$, then there are $p_{x_1}^{(n)} \in \mathscr{P}_2^{x_1}(\mathbb{R}^d)$ with

$$\int \psi \, dp_{x_1}^{(n)} - \operatorname{MCov}(p_{x_1}^{(n)}, q) \to -\infty.$$

We observe that $p_x^{(n)} \coloneqq \lambda p_{x_1}^{(n)} + (1-\lambda)\delta_{x_2} \in \mathscr{P}_2^x(\mathbb{R}^d)$, and consequently

$$\begin{aligned}\varphi^\psi(x) &\leqslant \lambda \int \psi \, dp_{x_1}^{(n)} + (1-\lambda)\psi(x_2) - \operatorname{MCov}(p_x^{(n)}, q)\\ &\leqslant \lambda\Big(\int \psi \, dp_{x_1}^{(n)} - \operatorname{MCov}(p_{x_1}^{(n)}, q)\Big) + (1-\lambda)\psi(x_2),\end{aligned}$$

where we have used the convexity of $\mathscr{P}_2(\mathbb{R}^d) \ni p \longmapsto -\operatorname{MCov}(p, q)$ in the last inequality. We conclude that $\varphi^\psi(x) = -\infty$. The case $\varphi^\psi(x_2) = -\infty$ is treated similarly. If, on the other hand, both $\varphi^\psi(x_1) > -\infty$ and $\varphi^\psi(x_2) > -\infty$, then

$$\varphi^\psi(x) \leqslant \lambda\varphi^\psi(x_1) + (1-\lambda)\varphi^\psi(x_2)$$

follows by standard arguments. Altogether we see that φ^ψ is convex on $\operatorname{dom}\psi$.

(ii) If $\varphi^\psi(x) > -\infty$ for one $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\varphi^\psi(\tilde{x}) > -\infty$ for all $\tilde{x} \in \operatorname{int}(\operatorname{dom}\psi)$, as can be seen directly by convexity. □

Our next result is the analogue of Lemma 2.6. But now, instead of maximizing over all $p \in \mathscr{P}_2(\mathbb{R}^d)$ as in (2.18), we maximize over all $p \in \mathscr{P}_2^x(\mathbb{R}^d)$, with a fixed barycenter $x \in \mathbb{R}^d$, as in (2.17).

Lemma 3.2. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function and assume that $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$. Then we have the duality formula*

$$\varphi^\psi(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x, y\rangle - \int \psi^*(y+z)\, dq(z) \Big). \tag{3.1}$$

Proof. We first show that $(\varphi^\psi)^* = \psi^* \star q$. To see this, we recall equations (2.25) – (2.28) of Proposition 2.7 and note that these are valid for any proper, lower semicontinuous convex function ψ. Therefore, for each $y \in \mathbb{R}^d$, we have

$$(\varphi^\psi)^*(y) = \sup_{p\in\mathscr{P}_2(\mathbb{R}^d)} \Big(\mathrm{MCov}(p,q) - \int \psi_y\, dp \Big) = \int \psi^*(y+z)\, dq(z), \tag{3.2}$$

where the function ψ_y is defined by $\psi_y(z) \coloneqq \psi(z) - \langle y, z\rangle$, for $z \in \mathbb{R}^d$.

Since $\varphi^\psi(x) > -\infty$ for some $x \in \operatorname{int}(\operatorname{dom}\psi)$ and φ^ψ is convex by Lemma 3.1, the function φ^ψ is finite and thus also continuous in a neighbourhood of x. Hence we can apply the variant [BC17, Proposition 13.44] of the Fenchel–Moreau theorem and deduce from (3.2) that

$$\varphi^\psi(x) = (\varphi^\psi)^{**}(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x, y\rangle - \int \psi^*(y+z)\, dq(z) \Big), \tag{3.3}$$

which proves the duality formula (3.1). □

The duality formula (3.1) established in Lemma 3.2 enables us to generalize Lemma 2.5 and identify the structure of the optimizer in the minimization problem (2.2).

Lemma 3.3. *Let $\psi\colon \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous convex function.*

(i) If there are $x \in \mathbb{R}^d$ and $\hat{y}(x) \in \mathbb{R}^d$ such that

$$\hat{p}_x \coloneqq \nabla\psi^*(\hat{y}(x) + \,\cdot\,)(q) \in \mathscr{P}_2^x(\mathbb{R}^d), \tag{3.4}$$

then this is actually the unique optimizer of the infimum in (2.2).

(ii) If additionally $\varphi^\psi(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom}\psi)$, then $\hat{y}(x)$ is the optimizer of the supremum in (3.1).

Proof. (i) For fixed $y \in \mathbb{R}^d$ we define the function ψ_y by $\psi_y(z) \coloneqq \psi(z) - \langle y, z\rangle$, for $z \in \mathbb{R}^d$, so that $\psi_y^*(z) = \psi^*(y+z)$. With this notation, we can express the probability measure defined in (3.4) as $\hat{p}_x = (\nabla\psi^*_{\hat{y}(x)})(q)$. We denote by $T_q^{p_x}$ the Brenier map from q to $p_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ and note that $T_q^{\hat{p}_x} = \nabla\psi^*_{\hat{y}(x)}$ by Brenier's theorem. For arbitrary $p_x \in \mathscr{P}_2^x(\mathbb{R}^d)$ we compute

$$\begin{aligned}
\int \psi\, dp_x - \mathrm{MCov}(p_x, q) &= \int \Big(\psi_{\hat{y}(x)}(T_q^{p_x}(z)) + \langle T_q^{p_x}(z), \hat{y}(x)\rangle - \langle T_q^{p_x}(z), z\rangle \Big)\, dq(z) \\
&\geqslant \langle x, \hat{y}(x)\rangle + \int \inf_{y\in\mathbb{R}^d} \left(\psi_{\hat{y}(x)}(y) - \langle y, z\rangle\right) dq(z) \\
&= \langle x, \hat{y}(x)\rangle - \int \psi^*(\hat{y}(x)+z)\, dq(z) \\
&= \int \psi\, d\hat{p}_x - \mathrm{MCov}(\hat{p}_x, q),
\end{aligned}$$

with equality if and only if $T_q^{p_x}(z) = T_q^{\hat{p}_x}(z)$, for q-a.e. $z \in \mathbb{R}^d$. This in turn is the case if and only if $p_x = \hat{p}_x$. We conclude that $\hat{p}_x$ is the unique optimizer of the infimum in (2.2), i.e.,

$$\varphi^{\psi}(x) = \int \psi \, d\hat{p}_x - \mathrm{MCov}(\hat{p}_x, q) = \langle x, \hat{y}(x) \rangle - \int \psi^*(\hat{y}(x) + z) \, dq(z). \tag{3.5}$$

(ii) If $\varphi^{\psi}(x) > -\infty$ and $x \in \operatorname{int}(\operatorname{dom} \psi)$, then by Lemma 3.2 we have that

$$\varphi^{\psi}(x) = \sup_{y \in \mathbb{R}^d} \Big(\langle x, y \rangle - \int \psi^*(y + z) \, dq(z) \Big),$$

which in light of (3.5) implies that $\hat{y}(x)$ is the optimizer of the supremum in (3.1). □

3.1. Proof of Theorem 1.5

Definition 3.4. A lower semicontinuous convex function $\hat{\psi} \colon \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $\mu(\operatorname{int}(\operatorname{dom} \hat{\psi})) = 1$ is called an optimizer of the dual problem (1.4) (in short, a dual optimizer) if $D^q(\mu, \nu) = \mathcal{E}(\hat{\psi})$, for the dual function $\mathcal{E}(\,\cdot\,)$ as defined in (2.30).

Proof of Theorem 1.5. (1) Let $\hat{v} \colon \mathbb{R}^d \to \mathbb{R}$ be q-Bass martingale generating with initial marginal μ. Then, according to Definition 1.4, the function $\hat{v}$ is convex, its convex conjugate $\hat{\psi} = \hat{v}^*$ satisfies $\mu(\operatorname{int}(\operatorname{dom} \hat{\psi})) = 1$, and the finite-valued function $\hat{v} \star q$ is strictly convex with gradient $\nabla(\hat{v} \star q) = (\nabla \hat{v}) \star q$. By [Roc70, Theorem 26.5], it follows that the convex conjugate $(\hat{v} \star q)^*$ is differentiable and strictly convex on $\operatorname{int}(\operatorname{dom}(\hat{v} \star q)^*)$. Furthermore, the function

$$\nabla(\hat{v} \star q) \colon \mathbb{R}^d \longrightarrow \operatorname{int}(\operatorname{dom}(\hat{v} \star q)^*)$$

is a bijection with inverse

$$(\nabla \hat{v} \star q)^{-1} = \nabla(\hat{v} \star q)^*. \tag{3.6}$$

Next, we show that $\mu(\operatorname{int}(\operatorname{dom}(\hat{v} \star q)^*)) = 1$. To this end, we will prove the inclusion $\operatorname{int}(\operatorname{dom} \hat{\psi}) \subseteq \operatorname{int}(\operatorname{dom}(\hat{v} \star q)^*)$. Indeed, for $x \in \operatorname{dom} \hat{\psi}$, we have

$$\begin{aligned}
(\hat{v} \star q)^*(x) &= \sup_{y \in \mathbb{R}^d} \Big(\langle x, y \rangle - \int \hat{v}(y + z) \, dq(z) \Big) \leqslant \int \sup_{y \in \mathbb{R}^d} \Big(\langle x, y \rangle - \hat{v}(y + z) \Big) \, dq(z) \\
&= \int \sup_{y \in \mathbb{R}^d} \Big(\langle x, y - z \rangle - \hat{v}(y) \Big) \, dq(z) = \int \Big(\hat{v}^*(x) - \langle x, z \rangle \Big) \, dq(z) \\
&\leqslant \hat{v}^*(x) + |x| \sqrt{\int |z|^2 \, dq(z)} < +\infty,
\end{aligned}$$

so that $\operatorname{dom} \hat{\psi} \subseteq \operatorname{dom}(\hat{v} \star q)^*$.

Now we are ready to show that the pair $(\hat{v}, \hat{\alpha}^{\hat{v}})$ is a q-Bass martingale with initial marginal μ and with terminal marginal $\nu^{\hat{v}}$, i.e.,

$$(\nabla \hat{v} \star q)(\hat{\alpha}^{\hat{v}}) = \mu \qquad \text{and} \qquad \nabla \hat{v}(\hat{\alpha}^{\hat{v}} * q) = \nu^{\hat{v}}. \tag{3.7}$$

According to Definition 1.4, we have $\hat{\alpha}^{\hat{v}} = \nabla(\hat{v} \star q)^*(\mu)$ and $\nu^{\hat{v}} = \nabla\hat{v}(\hat{\alpha}^{\hat{v}} * q)$. Together with (3.6), we conclude the relations (3.7).

(2) Recalling the second equality in (3.3), we have

$$\varphi^{\hat{\psi}}(x) \geqslant (\varphi^{\hat{\psi}})^{**}(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x, y\rangle - \int \hat{v}(y+z)\, dq(z) \Big). \tag{3.8}$$

As the right-hand side of (3.8) is greater than $-\infty$, for every $x \in \mathbb{R}^d$, we can apply Lemma 3.2. Consequently, for every $x \in \operatorname{int}(\operatorname{dom}\hat{\psi})$, we obtain an equality in (3.8), i.e.,

$$\varphi^{\hat{\psi}}(x) = \sup_{y\in\mathbb{R}^d} \Big(\langle x, y\rangle - (\hat{v} \star q)(y) \Big). \tag{3.9}$$

Recall that the finite-valued function $\hat{v} \star q$ is strictly convex and differentiable with gradient $\nabla(\hat{v} \star q) = (\nabla\hat{v}) \star q$. The first order condition for an optimizer $\hat{y}(x) \in \mathbb{R}^d$ of the right-hand side of (3.9) reads

$$(\nabla\hat{v} \star q)(\hat{y}(x)) = x. \tag{3.10}$$

From (3.6) we deduce that the right-hand side of (3.9) admits a unique maximizer $\hat{y}(x)$ given by

$$\hat{y}(x) = (\nabla\hat{v} \star q)^{-1}(x) = \nabla(\hat{v} \star q)^*(x),$$

for all $x \in \operatorname{int}(\operatorname{dom}\hat{\psi})$.

For each $x \in \operatorname{int}(\operatorname{dom}\hat{\psi})$, we define a probability measure

$$\pi_x^{\hat{v}} := \nabla\hat{v}(\hat{y}(x) + \cdot\,)(q).$$

Since $\nu^{\hat{v}}$ has finite second moment, we observe that

$$\int |\nabla\hat{v}(y+z)|^2\, dq(z) < +\infty, \tag{3.11}$$

for $\hat{\alpha}^{\hat{v}}$-a.e. $y \in \mathbb{R}^d$. By (3.10) and (3.11), $\pi_x^{\hat{v}}$ is an element of $\mathscr{P}_2^x(\mathbb{R}^d)$, for μ-a.e. $x \in \mathbb{R}^d$. Hence we can apply part (i) of Lemma 3.3, showing that the infimum

$$\varphi^{\hat{\psi}}(x) = \inf_{p\in\mathscr{P}_2^x(\mathbb{R}^d)} \Big(\int \hat{\psi}\, dp - \mathrm{MCov}(p, q) \Big)$$

is attained by $\pi_x^{\hat{v}}$, for μ-a.e. $x \in \mathbb{R}^d$. Therefore, from the representation (2.12) of the dual function (2.5) and the fact that $\pi_x^{\hat{v}}(dy)\,\mu(dx) \in \mathsf{MT}(\mu, \nu^{\hat{v}})$, we obtain

$$\tilde{\mathcal{E}}(\hat{\psi}) = \int \Big(\int \hat{\psi}(y)\, \pi_x^{\hat{v}}(dy) - \varphi^{\hat{\psi}}(x) \Big)\, \mu(dx) = \int \mathrm{MCov}(\pi_x^{\hat{v}}, q)\, \mu(dx). \tag{3.12}$$

On the other hand, from (1.3), Proposition 2.1 and (2.13) we have

$$\int \mathrm{MCov}(\pi_x^{\hat{v}}, q)\, \mu(dx) \leqslant P^q(\mu, \nu^{\hat{v}}) = \tilde{D}^q(\mu, \nu^{\hat{v}}) \leqslant \tilde{\mathcal{E}}(\hat{\psi}), \tag{3.13}$$

which in light of (3.12) implies that

$$P^q(\mu, \nu^{\hat{v}}) = \int \mathrm{MCov}(\pi_x^{\hat{v}}, q)\, \mu(dx).$$

By uniqueness of the optimizer $\hat{\pi} \in \mathsf{MT}(\mu, \nu^{\hat{v}})$, we conclude that $\pi_x^{\hat{v}} = \hat{\pi}_x$, for μ-a.e. $x \in \mathbb{R}^d$.

(3) According to (3.9), for every $x \in \mathrm{int}(\mathrm{dom}\, \hat{\psi})$, we have that

$$\varphi^{\hat{\psi}}(x) = (\hat{v} \star q)^*(x).$$

Recalling the definition of the dual function (2.30), we observe that

$$\begin{aligned}\mathcal{E}(\hat{\psi}) &= \int \Big(\int \hat{\psi}(y)\, \pi_x(dy) - (\hat{\psi}^* \star q)^*(x) \Big)\, \mu(dx) \\ &= \int \Big(\int \hat{\psi}(y)\, \pi_x^{\hat{v}}(dy) - \varphi^{\hat{\psi}}(x) \Big)\, \mu(dx) = \tilde{\mathcal{E}}(\hat{\psi}).\end{aligned}$$

Finally, from (3.12) and (3.13) we deduce that $\hat{\psi}$ is a dual optimizer in the sense of Definition 3.4. □

Bibliography

[ABP+13] B. Acciaio, M. Beiglböck, F. Penkner, W. Schachermayer, and J. Temme. A trajectorial interpretation of Doob's martingale inequalities. *Ann. Appl. Probab.*, **23**(4):1494–1505, 2013.

[ADPZ13] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. Large deviations and gradient flows. *Philos. Trans. Roy. Soc. A*, **371**(2005), 2013.

[AG13] L. Ambrosio and N. Gigli. A User's Guide to Optimal Transport. In *Modelling and Optimisation of Flows on Networks*, volume **2062** of *Lecture Notes in Math.*, pages 1–155. Springer, Berlin, Heidelberg, 2013.

[AGK04] M. Agueh, N. Ghoussoub, and X. Kang. Geometric inequalities via a general comparison principle for interacting gases. *Geom. Funct. Anal.*, **14**(1):215–244, 2004.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lect. Math. ETH Zürich. Birkhäuser, Basel, second edition, 2008.

[AMP23] B. Acciaio, A. Marini, and G. Pammer. Calibration of the Bass Local Volatility model. arXiv:2311.14567, 2023.

[AMTU01] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations. *Comm. Partial Differential Equations*, **26**(1–2):43–100, 2001.

[AP22] B. Acciaio and G. Pammer. A short proof of the characterisation of convex order using the 2-Wasserstein distance. arXiv:2207.02096, 2022.

[Bac00] L. Bachelier. Théorie de la spéculation. *Ann. Sci. Éc. Norm. Supér.*, Série 3: 17:21–86, 1900.

[Bac06] L. Bachelier. *Louis Bachelier's Theory of Speculation: The Origins of Modern Finance — Translated and with Commentary by Mark Davis & Alison Etheridge*. Princeton University Press, Princeton, NJ, 2006.

[Bas83] R.F. Bass. Skorokhod imbedding via stochastic integrals. In J. Azéma and M. Yor, editors, *Sémin. Probab. XVII 1981/82 — Proceedings*, volume **986** of *Lecture Notes in Math.*, pages 221–224. Springer, Berlin, Heidelberg, 1983.

[BB99] J.-D. Benamou and Y. Brenier. A numerical method for the optimal time-continuous mass transport problem and related problems. In L.A. Caffarelli and M. Milman, editors, *Monge Ampère Equation: Applications to Geometry and Optimization (Deerfield Beach, FL, 1997)*, volume **226** of *Contemp. Math.*, pages 1–11. American Mathematical Society, Providence, RI, 1999.

[BBBE20] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, **24**(3):601–632, 2020.

[BBHK20] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou–Brenier: A probabilistic perspective. *Ann. Probab.*, **48**(5):2258–2289, 2020.

[BBP19] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differential Equations*, **58**(6):Article number: 203, 2019.

[BBST23] J. Backhoff-Veraguas, M. Beiglböck, W. Schachermayer, and B. Tschiderer. The structure of martingale Benamou–Brenier in $\mathbb{R}^d$. arXiv:2306.11019, 2023.

[BC17] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer International Publishing, Switzerland, second edition, 2017.

[BCCP98] D. Benedetto, E. Caglioti, J.A. Carrillo, and M. Pulvirenti. A Non-Maxwellian Steady Distribution for One-Dimensional Granular Media. *J. Stat. Phys.*, **91**(5–6):979–990, 1998.

[BCGL20] J. Backhoff-Veraguas, G. Conforti, I. Gentil, and C. Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probab. Theory Relat. Fields*, **178**(1–2):475–530, 2020.

[BCH17] M. Beiglböck, A.M.G. Cox, and M. Huesmann. Optimal transport and Skorokhod embedding. *Invent. Math.*, **208**(2):327–400, 2017.

[BÉ85] D. Bakry and M. Émery. Diffusions hypercontractives. In J. Azéma and M. Yor, editors, *Sémin. Probab. XIX 1983/84*, volume **1123** of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, Heidelberg, 1985.

[Bel95] D.R. Bell. *Degenerate Stochastic Differential Equations and Hypoellipticity*, volume **79** of *Pitman Monogr. Surv. Pure Appl. Math.* Longman Group Limited, Harlow, 1995.

[Bel06] L.R. Bellet. Ergodic Properties of Markov Processes. In S. Attal, A. Joye, and C.-A. Pillet, editors, *Open Quantum Systems II — The Markovian Approach*, volume **1881** of *Lecture Notes in Math.*, pages 1–39. Springer, Berlin, Heidelberg, 2006.

[BFL24] E. Bayraktar, Q. Feng, and W. Li. Exponential Entropy Dissipation for Weakly Self-Consistent Vlasov–Fokker–Planck Equations. *J. Nonlinear Sci.*, **34**(7), 2024.

[BGG13] F. Bolley, I. Gentil, and A. Guillin. Uniform Convergence to Equilibrium for Granular Media. *Arch. Ration. Mech. Anal.*, **208**(2):429–445, 2013.

[BHLP13] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices — a mass transport approach. *Finance Stoch.*, **17**(3):477–501, 2013.

[BHLT17] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and Skorokhod embedding. *Stochastic Process. Appl.*, **127**(9):3005–3013, 2017.

[BJ16] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, **44**(1):42–106, 2016.

[BJMP23] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Stability of the weak martingale optimal transport problem. *Ann. Appl. Probab.*, **33**(6B):5382–5412, 2023.

[Bla92] A. Blaquière. Controllability of a Fokker–Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version). *Dyn. Control*, **2**(3):235–253, 1992.

[BN15] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, **25**(2):823–859, 2015.

[BNS22] M. Beiglböck, M. Nutz, and F. Stebegg. Fine properties of the optimal Skorokhod embedding problem. *J. Eur. Math. Soc.*, **24**(4):1389–1429, 2022.

[BNT17] M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Probab.*, **45**(5):3038–3074, 2017.

[Bol96] L. Boltzmann. *Vorlesungen über Gastheorie — I. Theil.* Johann Ambrosius Barth Verlag, Leipzig, 1896.

[Bol98a] L. Boltzmann. Ueber die sogenannte H-Curve. *Math. Ann.*, **50**(2–3):325–332, 1898.

[Bol98b] L. Boltzmann. *Vorlesungen über Gastheorie — II. Theil.* Johann Ambrosius Barth Verlag, Leipzig, 1898.

[Bre87] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, **305**(1):805–808, 1987.

[Bre91] Y. Brenier. Polar Factorization and Monotone Rearrangement of Vector-Valued Functions. *Comm. Pure Appl. Math.*, **44**(4):375–417, 1991.

[BRTV98] S. Benachour, B. Roynette, D. Talay, and P. Vallois. Nonlinear self-stabilizing processes — I Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.*, **75**(2):173–201, 1998.

[BS15] M. Beiglböck and P. Siorpaes. Pathwise versions of the Burkholder–Davis–Gundy inequality. *Bernoulli*, **21**(1):360–373, 2015.

[BST23] J. Backhoff-Veraguas, W. Schachermayer, and B. Tschiderer. The Bass functional of martingale transport. arXiv:2309.11181, 2023.

[Car84] E.A. Carlen. Conservative diffusions. *Comm. Math. Phys.*, **94**(3):293–315, 1984.

[Cat04] P. Cattiaux. A Pathwise Approach of Some Classical Inequalities. *Potential Anal.*, **20**(4):361–394, 204.

[CD18] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I — Mean Field FBSDEs, Control, and Games*, volume **83** of *Probab. Theory Stoch. Model.* Springer International Publishing, Switzerland, 2018.

[CE02] D. Cordero-Erausquin. Some Applications of Mass Transport to Gaussian-Type Inequalities. *Arch. Ration. Mech. Anal.*, **161**(3):257–269, 2002.

[CEGH04] D. Cordero-Erausquin, W. Gangbo, and C. Houdré. Inequalities for generalized entropy and optimal transportation. In M.C. Carvalho and J.F. Rodrigues, editors, *Recent Advances in the Theory and Applications of Mass Transport*, volume **353** of *Contemp. Math.*, pages 73–94. American Mathematical Society, Providence, RI, 2004.

[CG03] E.A. Carlen and W. Gangbo. Constrained steepest descent in the 2-Wasserstein metric. *Ann. of Math.*, **157**(3):807–846, 2003.

[CG14] P. Cattiaux and A. Guillin. Semi Log-Concave Markov Diffusions. In C. Donati-Martin, A. Lejay, and A. Rouault, editors, *Sémin. Probab. XLVI*, volume **2123** of *Lecture Notes in Math.*, pages 231–292. Springer, Berlin, Heidelberg, 2014.

[CGM08] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Relat. Fields*, **140**(1–2):19–40, 2008.

[CGP16a] Y. Chen, T.T. Georgiou, and M. Pavon. On the Relation Between Optimal Transport and Schrödinger Bridges: A Stochastic Control Viewpoint. *J. Optim. Theory Appl.*, **169**(2):671–691, 2016.

[CGP16b] Y. Chen, T.T. Georgiou, and M. Pavon. Stochastic control, entropic interpolation and gradient flows on Wasserstein product spaces. In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems*, Minneapolis, MN, July 11–15, 2016.

[CGP21] Y. Chen, T.T. Georgiou, and M. Pavon. Stochastic Control Liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger Bridge. *SIAM Rev.*, **63**(2):249–313, 2021.

[CGPS20] J.A. Carrillo, R.S. Gvalani, G.A. Pavliotis, and A. Schlichting. Long-Time Behaviour and Phase Transitions for the McKean–Vlasov Equation on the Torus. *Arch. Ration. Mech. Anal.*, **235**(1):635–690, 2020.

[CHL21] A. Conze and P. Henry-Labordère. Bass Construction with Multi-Marginals: Lightspeed Computation in a New Local Volatility Model. SSRN:3853085, 2021.

[Chu01] K.L. Chung. *A Course in Probability Theory*. Academic Press, San Diego, third edition, 2001.

[Cio23a] K.J. Ciosmak. General localisation scheme I: theory. arXiv:2312.12281, 2023.

[Cio23b] K.J. Ciosmak. General localisation scheme II: applications. arXiv:2312.13167, 2023.

[CJM+01] J.A. Carrillo, A. Jüngel, P.A. Markowich, G. Toscani, and A. Unterreiter. Entropy Dissipation Methods for Degenerate Parabolic Problems and Generalized Sobolev Inequalities. *Monatsh. Math.*, **133**(1):1–82, 2001.

[CKPS21] P. Cheridito, M. Kiiski, D.J. Prömel, and H.M. Soner. Martingale optimal transport duality. *Math. Ann.*, **379**(3–4):1685–1712, 2021.

[CL98] G. Casella and E.L. Lehmann. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, New York, second edition, 1998.

[CL22] G. Conforti and C. Léonard. Time reversal of Markov processes with jumps under a finite entropy condition. *Stochastic Process. Appl.*, **144**:85–124, 2022.

[CLM17] L. Campi, I. Laachir, and C. Martini. Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.*, **21**(2):471–486, 2017.

[CMV03] J.A. Carrillo, R.J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.*, **19**(3):971–1018, 2003.

[CMV06] J.A. Carrillo, R.J. McCann, and C. Villani. Contractions in the 2-Wasserstein Length Space and Thermalization of Granular Media. *Arch. Ration. Mech. Anal.*, **179**(2):217–263, 2006.

[COT19] A.M.G. Cox, J. Obłój, and N. Touzi. The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reversal approach. *Probab. Theory Relat. Fields*, **173**(1–2):211–259, 2019.

[CS91] E.A. Carlen and A. Soffer. Entropy Production by Block Variable Summation and Central Limit Theorems. *Comm. Math. Phys.*, **140**(2):339–371, 1991.

[CT06] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Ser. Telecommun. Signal Process. John Wiley & Sons, Hoboken, NJ, second edition, 2006.

[DG93] E. De Giorgi. New problems on minimizing movements. In C. Baiocchi and J.-L. Lions, editors, *Boundary Value Problems for Partial Differential Equations and Applications — Dedicated to E. Magenes*, volume **29** of *Research Notes in Applied Mathematics*, pages 81–98. Masson, Paris, 1993.

[DGMT80] E. De Giorgi, A. Marino, and M. Tosques. Problemi di evoluzione in spazi metrici e curve di massima pendenza. *Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8*, **68**(3):180–187, 1980.

[DK94] M.H.A. Davis and I. Karatzas. A Deterministic Approach to Optimal Stopping. In *Probability, Statistics and Optimization: A Tribute to Peter Whittle*, volume **104** of *Wiley Ser. Probab. Stat. — Appl. Probab. Statist. Sect.*, pages 455–466. John Wiley & Sons, Chichester, 1994.

[DMT85] M. Degiovanni, A. Marino, and M. Tosques. Evolution equations with lack of convexity. *Nonlinear Analysis: Theory, Methods & Applications*, **9**(12):1401–1443, 1985.

[DMT19] H. De March and N. Touzi. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *Ann. Probab.*, **47**(3):1726–1774, 2019.

[DP91] P. Dai Pra. A stochastic control approach to reciprocal diffusion processes. *Appl. Math. Optim.*, **23**(1):313–329, 1991.

[DPP89] P. Dai Pra and M. Pavon. Variational path-integral representations for the density of a diffusion process. *Stochastics*, **26**(4):205–226, 1989.

[DPP90] P. Dai Pra and M. Pavon. On the Markov processes of Schrödinger, the Feynman–Kac formula and stochastic control. In M.A. Kaashoek, J.H. van Schuppen, and A.C.M. Ran, editors, *Realization and Modelling in System Theory*, volume **3** of *Prog. Syst. Control Theory*, pages 497–504. Birkhäuser, Boston, 1990.

[DS14] Y. Dolinsky and H.M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, **160**(1–2):391–427, 2014.

[DT16] M.H. Duong and J. Tugaut. Stationary solutions of the Vlasov–Fokker–Planck equation: Existence, characterization and phase-transition. *Appl. Math. Lett.*, **52**:38–45, 2016.

[DT18] M.H. Duong and J. Tugaut. The Vlasov–Fokker–Planck equation in non-convex landscapes: convergence to equilibrium. *Electron. Commun. Probab.*, **23**:1–10, 2018.

[EFLS16] M. Erbar, M. Fathi, V. Laschos, and A. Schlichting. Gradient flow structure for McKean–Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, **36**(12):6799–6833, 2016.

[ELS20] R. Eldan, J. Lehec, and Y. Shenfeld. Stability of the logarithmic Sobolev inequality via the Föllmer process. *Ann. Inst. Henri Poincaré Probab. Stat.*, **56**(3):2253–2269, 2020.

[EM20] R. Eldan and D. Mikulincer. Stability of the Shannon–Stam inequality via the Föllmer process. *Probab. Theory Relat. Fields*, **177**(3–4):891–922, 2020.

[Fat16] M. Fathi. A gradient flow approach to large deviations for diffusion processes. *J. Math. Pures Appl.*, **106**(5):957–993, 2016.

[FJ16] J. Fontbona and B. Jourdain. A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *Ann. Probab.*, **44**(1):131–170, 2016.

[FK06] J. Feng and T.G. Kurtz. *Large Deviations for Stochastic Processes*, volume **131** of *Math. Surveys Monogr.* American Mathematical Society, Providence, RI, 2006.

[FL23a] Q. Feng and W. Li. Entropy Dissipation for Degenerate Stochastic Differential Equations via Sub-Riemannian Density Manifold. *Entropy*, **25**(5):Article number: 786, 2023.

[FL23b] Q. Feng and W. Li. Hypoelliptic entropy dissipation for stochastic differential equations. arXiv:2102.00544, 2023.

[Fle77] W.H. Fleming. Exit probabilities and optimal stochastic control. *Appl. Math. Optim.*, **4**(1):329–346, 1977.

[Fle82] W.H. Fleming. Logarithmic transformations and stochastic control. In W.H. Fleming and L.G. Gorostiza, editors, *Advances in Filtering and Optimal Stochastic Control*, volume **42** of *Lect. Notes Control Inf. Sci.*, pages 131–141. Springer, Berlin, Heidelberg, 1982.

[Föl85] H. Föllmer. An entropy approach to the time reversal of diffusion processes. In M. Métivier and É. Pardoux, editors, *Stochastic Differential Systems — Filtering and Control*, volume **69** of *Lect. Notes Control Inf. Sci.*, pages 156–163. Springer, Berlin, Heidelberg, 1985.

[Föl86] H. Föllmer. Time reversal on Wiener space. In S.A. Albeverio, P. Blanchard, and L. Streit, editors, *Stochastic Processes — Mathematics and Physics*, volume **1158** of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, Heidelberg, 1986.

[Föl88] H. Föllmer. Random fields and diffusion processes. In P.-L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume **1362** of *Lecture Notes in Math.*, pages 101–203. Springer, Berlin, Heidelberg, 1988.

[Föl22] H. Föllmer. Optimal Couplings on Wiener Space and An Extension of Talagrands Transport Inequality. In G. Yin and T. Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization — A Commemorative Volume to Honor Mark H.A. Davis's Contributions*, pages 147–175. Springer, Cham, 2022.

[Fri75] A. Friedman. *Stochastic Differential Equations and Applications — Volume 1*, volume **28** of *Probab. Math. Statist. Ser. Monogr. Textb.* Academic Press, New York, 1975.

[FS85] W.H. Fleming and S.-J. Sheu. Stochastic variational formula for fundamental solutions of parabolic PDE. *Appl. Math. Optim.*, **13**(1):193–204, 1985.

[Gar09] C. Gardiner. *Stochastic Methods — A Handbook for the Natural and Social Sciences*, volume **13** of *Springer Ser. Synergetics*. Springer, Berlin, Heidelberg, fourth edition, 2009.

[GHLT14] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, **24**(1):312–336, 2014.

[GJ20] N. Gozlan and N. Juillet. On a mixture of Brenier and Strassen Theorems. *Proc. Lond. Math. Soc.*, **120**(3):434–463, 2020.

[GKL19] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.*, **47**(1):109–164, 2019.

[GKL20] N. Ghoussoub, Y.-H. Kim, and T. Lim. Optimal Brownian Stopping When the Source and Target Are Radially Symmetric Distributions. *SIAM J. Control Optim.*, **58**(5):2765–2789, 2020.

[GKP19] N. Ghoussoub, Y.-H. Kim, and A.Z. Palmer. PDE methods for optimal Skorokhod embeddings. *Calc. Var. Partial Differential Equations*, **58**(3):Article number: 113, 2019.

[GL10] N. Gozlan and C. Léonard. Transport Inequalities. A Survey. *Markov Process. Relat. Fields*, **16**(4):635–736, 2010.

[GL21] I. Guo and G. Loeper. Path dependent optimal transport and model calibration on exotic derivatives. *Ann. Appl. Probab.*, **31**(3):1232–1263, 2021.

[GLR20] I. Gentil, C. Léonard, and L. Ripani. Dynamical aspects of the generalized Schrödinger problem via Otto calculus — A heuristic point of view. *Rev. Mat. Iberoam.*, **36**(4):1071–1112, 2020.

[GLRT20] I. Gentil, C. Léonard, L. Ripani, and L. Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Process. Appl.*, **130**(2):907–923, 2020.

[GLW19] I. Guo, G. Loeper, and S. Wang. Local Volatility Calibration by Optimal Transport. In J. de Gier, C.E. Praeger, and T. Tao, editors, *2017 MATRIX Annals*, volume **2** of *MATRIX Book Series*, pages 51–64. Springer, Cham, 2019.

[GRST17] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, **273**(11):3327–3405, 2017.

[HLOST16] P. Henry-Labordère, J. Obłój, P. Spoida, and N. Touzi. The maximum maximum of a martingale with given n marginals. *Ann. Appl. Probab.*, **26**(1):1–44, 2016.

[HLT16] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, **20**(3):635–668, 2016.

[HN12] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, **22**(1):31–56, 2012.

[Hol77] C.J. Holland. A new energy characterization of the smallest eigenvalue of the Schrödinger equation. *Comm. Pure Appl. Math.*, **30**(6):755–765, 1977.

[HP86] U.G. Haussmann and É. Pardoux. Time Reversal of Diffusions. *Ann. Probab.*, **14**(4):1188–1205, 1986.

[HRŠS21] 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.*, **57**(4):2043–2065, 2021.

[HT10] S. Herrmann and J. Tugaut. Non-uniqueness of stationary measures for self-stabilizing processes. *Stochastic Process. Appl.*, **120**(7):1215–1246, 2010.

[HT19] M. Huesmann and D. Trevisan. A Benamou–Brenier formulation of martingale optimal transport. *Bernoulli*, **25**(4A):2729–2757, 2019.

[JK96] R. Jordan and D. Kinderlehrer. An extended variational principle. In *Partial Differential Equations and Applications: Collected Papers in Honor of Carlo Pucci*, volume **177** of *Lect. Notes Pure Appl. Math.*, chapter 18, pages 187–200. CRC Press, 1996.

[JK23] B. Jahnel and J. Köppl. Trajectorial Dissipation of Φ-Entropies for Interacting Particle Systems. *J. Stat. Phys.*, **190**(7):119, 2023.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The Variational Formulation of the Fokker–Planck Equation. *SIAM J. Math. Anal.*, **29**(1):1–17, 1998.

[Joh04] O. Johnson. *Information Theory and the Central Limit Theorem.* Imperial College Press, London, 2004.

[Kan42] L.V. Kantorovich. On the translocation of masses. *Dokl. Akad. Nauk. USSR (NS)*, **37**:199–201, 1942.

[KK21] I. Karatzas and C. Kardaras. *Portfolio Theory and Arbitrage: A Course in Mathematical Finance*, volume **214** of *Grad. Stud. Math.* American Mathematical Society, Providence, RI, 2021.

[KMS21] I. Karatzas, J. Maas, and W. Schachermayer. Trajectorial dissipation and gradient flow for the relative entropy in Markov chains. *Commun. Inf. Syst.*, **21**(4):481–536, 2021.

[Kol31] A.N. Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, **104**(1):415–458, 1931.

[Kol37] A.N. Kolmogorov. Zur Umkehrbarkeit der statistischen Naturgesetze. *Math. Ann.*, **113**(1):766–772, 1937.

[Kom67] J. Komlós. A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.*, **18**(1–2):217–229, 1967.

[KS98] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume **113** of *Grad. Texts in Math.* Springer, New York, second edition, 1998.

[KST20] I. Karatzas, W. Schachermayer, and B. Tschiderer. Trajectorial Otto calculus. arXiv:1811.08686, 2020.

[KST22] I. Karatzas, W. Schachermayer, and B. Tschiderer. A trajectorial approach to the gradient flow properties of Langevin–Smoluchowski diffusions. *Theory Probab. Appl.*, **66**(4):668–707, 2022.

[KT22] I. Karatzas and B. Tschiderer. A variational characterization of Langevin–Smoluchowski diffusions. In G. Yin and T. Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization — A Commemorative Volume to Honor Mark H.A. Davis's Contributions*, pages 239–265. Springer, Cham, 2022.

[KTT17] S. Källblad, X. Tan, and N. Touzi. Optimal Skorokhod embedding given full marginals and Azéma-Yor peacocks. *Ann. Appl. Probab.*, **27**(2):686–719, 2017.

[Kun97] H. Kunita. *Stochastic flows and stochastic differential equations*, volume **24** of *Cambridge Stud. Adv. Math.* Cambridge University Press, Cambridge, 1997.

[KY22] D. Kim and L.C. Yeung. A trajectorial approach to entropy dissipation for degenerate parabolic equations. arXiv:2210.16158, 2022.

[Lee18] J.M. Lee. *Introduction to Riemannian Manifolds*, volume **176** of *Grad. Texts in Math.* Springer International Publishing, Switzerland, second edition, 2018.

[Leh13] J. Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.*, **49**(3):885–899, 2013.

[Léo14a] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, **34**(4):1533–1574, 2014.

[Léo14b] C. Léonard. Some Properties of Path Measures. In C. Donati-Martin, A. Lejay, and A. Rouault, editors, *Sémin. Probab. XLVI*, volume **2123** of *Lecture Notes in Math.*, pages 207–230. Springer International Publishing, Switzerland, 2014.

[Léo17] C. Léonard. On the convexity of the entropy along entropic interpolations. In *Measure Theory in Non-Smooth Spaces*, Open Access Partial Differ. Equ. Meas. Theory, pages 194–242. De Gruyter Open, 2017.

[Lio07] P.-L. Lions. Théorie des jeux de champ moyen et applications. *Lectures at the Collège de France.* https://www.college-de-france.fr/fr/agenda/cours/theorie-des-jeux-de-champ-moyen-et-applications-suite, 2007.

[Lis09] S. Lisini. Nonlinear diffusion equations with variable coefficients as gradient flows in wasserstein spaces. *ESAIM Control Optim. Calc. Var.*, **15**(3):712–740, 2009.

[Loe18] G. Loeper. Option pricing with linear market impact and nonlinear Black–Scholes equations. *Ann. Appl. Probab.*, **28**(5):2664–2726, 2018.

[Loe23] G. Loeper. The Measure Preserving Martingale Sinkhorn Algorithm. arXiv:2310.13797, 2023.

[Lot08] J. Lott. Some Geometric Calculations on Wasserstein Space. *Comm. Math. Phys.*, **277**(2):423–437, 2008.

[LS01] R.S. Liptser and A.N. Shiryaev. *Statistics of Random Processes — I. General Theory*, volume **5** of *Stoch. Model. Appl. Probab.* Springer, Berlin, Heidelberg, second edition, 2001.

[LV09] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, **169**(3):903–991, 2009.

[Mal03] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, **13**(2):540–560, 2003.

[McC94] R.J. McCann. A convexity theory for interacting gases and equilibrium crystals. *Princeton University ProQuest Dissertations Publishing*, **9505121**, 1994.

[McC95] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, **80**(2):309–323, 1995.

[McC97] R.J. McCann. A Convexity Principle for Interacting Gases. *Adv. Math.*, **128**(1):153–179, 1997.

[McC01] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, **11**(3):589–608, 2001.

[McK66] H.P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA*, **56**(6):1907–1911, 1966.

[Mey94] P.A. Meyer. Sur une transformation du mouvement brownien due à Jeulin et Yor. In J. Azéma, M. Yor, and P.A. Meyer, editors, *Sémin. Probab. XXVIII*, volume **1583** of *Lecture Notes in Math.*, pages 98–101. Springer, Berlin, Heidelberg, 1994.

[Mir61] L. Mirsky. Majorization of vectors and inequalities for convex functions. *Monatsh. Math.*, **65**(2):159–169, 1961.

[MMN18] 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*, **115**(33):E7665–E7671, 2018.

[MNS89] A. Millet, D. Nualart, and M. Sanz. Integration by Parts and Time Reversal for Diffusion Processes. *Ann. Probab.*, **17**(1):208–238, 1989.

[Mon81] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[MV00] P.A. Markowich and C. Villani. On the trend to equilibrium for the Fokker–Planck equation: an interplay between physics and functional analysis. In *Sixth Workshop on Partial Differential Equations — Part II*, volume **19** of *Mat. Contemp.*, pages 1–29. Sociedade Brasileira de Matemática, Rio de Janeiro, 2000.

[Nag64] M. Nagasawa. Time reversions of Markov processes. *Nagoya Math. J.*, **24**:177–204, 1964.

[Nag93] M. Nagasawa. *Schrödinger Equations and Diffusion Theory*, volume **86** of *Monogr. Math.* Birkhäuser, Basel, 1993.

[Nel01] E. Nelson. *Dynamical Theories of Brownian Motion.* Math. Notes. Princeton University Press, Princeton, NJ, second edition, 2001.

[NM79] M. Nagasawa and T. Maruyama. An Application of Time Reversal of Markov Processes to a Problem of Population Genetics. *Adv. in Appl. Probab.*, **11**(3):457–478, 1979.

[Nua06] D. Nualart. *The Malliavin Calculus and Related Topics.* Probab. Appl. Springer, Berlin, Heidelberg, second edition, 2006.

[Oht09] S. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *Amer. J. Math.*, **131**(2):475–516, 2009.

[OS17] J. Obłój and P. Siorpaes. Structure of martingale transports in finite dimensions. arXiv:1702.08433, 2017.

[OST15] J. Obłój, P. Spoida, and N. Touzi. Martingale Inequalities for the Maximum via Pathwise Arguments. In C. Donati-Martin, A. Lejay, and A. Rouault, editors, *In Memoriam Marc Yor — Sémin. Probab. XLVII*, volume **2137** of *Lecture Notes in Math.*, pages 227–247. Springer, Cham, 2015.

[Ott01] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, **26**(1–2):101–174, 2001.

[OV00] F. Otto and C. Villani. Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *J. Funct. Anal.*, **173**(2):361–400, 2000.

[Par86] É. Pardoux. Grossissement d'une filtration et retournement du temps d'une diffusion. In J. Azéma and M. Yor, editors, *Sémin. Probab. XX 1984/85*, volume **1204** of *Lecture Notes in Math.*, pages 48–55. Springer, Berlin, Heidelberg, 1986.

[Pav89] M. Pavon. Stochastic control and nonequilibrium thermodynamical systems. *Appl. Math. Optim.*, **19**(1):187–202, 1989.

[PW91] M. Pavon and A. Wakolbinger. On Free Energy, Stochastic Control, and Schrödinger Processes. In G.B. Di Masi, A. Gombani, and A.B. Kurzhansky, editors, *Modeling, Estimation and Control of Systems with Uncertainty*, volume **10** of *Prog. Syst. Control Theory*, pages 334–348. Birkhäuser, Boston, MA, 1991.

[Ris96] H. Risken. *The Fokker–Planck Equation — Methods of Solution and Applications*, volume **18** of *Springer Ser. Synergetics*. Springer, Berlin, Heidelberg, second edition, 1996.

[Roc70] R.T. Rockafellar. *Convex analysis*, volume **28** of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1970.

[Rog85] L.C.G. Rogers. Smooth Transition Densities for One-Dimensional Diffusions. *Bull. Lond. Math. Soc.*, **17**(2):157–161, 1985.

[RVW01] F. Russo, P. Vallois, and J. Wolf. A generalized class of Lyons-Zheng processes. *Bernoulli*, **7**(2):363–379, 2001.

[RW00a] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales — Volume 1: Foundations*. Cambridge Math. Libr. Cambridge University Press, second edition, 2000.

[RW00b] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales — Volume 2: Itô Calculus*. Cambridge Math. Libr. Cambridge University Press, second edition, 2000.

[RY99] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume **293** of *Grundlehren Math. Wiss.* Springer, Berlin, Heidelberg, third edition, 1999.

[San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians — Calculus of Variations, PDEs, and Modeling*, volume **87** of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Switzerland, 2015.

[Sch31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitz.-Ber. d. Preuß. Akad. d. Wiss. — Phys.-math. Klasse*, pages 144–153, 1931.

[Sch32] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. Henri Poincaré*, **2**(4):269–310, 1932.

[Sch80] Z. Schuss. Singular Perturbation Methods in Stochastic Differential Equations of Mathematical Physics. *SIAM Rev.*, **22**(2):119–155, 1980.

[ST24] W. Schachermayer and B. Tschiderer. The decomposition of stretched Brownian motion into Bass martingales. arXiv, 2024.

[Sta59] A.J. Stam. Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon. *Inf. Control*, **2**:101–112, 1959.

[Str65] V. Strassen. The Existence of Probability Measures with Given Marginals. *Ann. Math. Statist.*, **36**(2):423–439, 1965.

[Stu06a] K.-T. Sturm. On the geometry of metric measure spaces I. *Acta Math.*, **196**(1):65–131, 2006.

[Stu06b] K.-T. Sturm. On the geometry of metric measure spaces II. *Acta Math.*, **196**(1):133–177, 2006.

[Tsc23] B. Tschiderer. Diffusion processes as Wasserstein gradient flows via stochastic control of the volatility matrix. arXiv:2310.18678, 2023.

[Tsc24] B. Tschiderer. q-Bass martingales. arXiv:2402.05669, 2024.

[TT13] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *Ann. Probab.*, **41**(5):3201–3240, 2013.

[Tug13] J. Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.*, **41**(3A):1427–1460, 2013.

[Tug14a] J. Tugaut. Phase transitions of McKean–Vlasov processes in double-wells landscape. *Stochastics*, **86**(2):257–284, 2014.

[Tug14b] J. Tugaut. Self-stabilizing Processes in Multi-wells Landscape in $\mathbb{R}^d$-Invariant Probabilities. *J. Theoret. Probab.*, **27**(1):57–79, 2014.

[TY23] B. Tschiderer and L.C. Yeung. A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: gradient flows and HWBI inequalities. *Bernoulli*, **29**(1):725–756, 2023.

[Ver06] A.Y. Veretennikov. On Ergodic Measures for McKean–Vlasov Stochastic Equations. In H. Niederreiter and D. Talay, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 471–486. Springer, Berlin, Heidelberg, 2006.

[Vil03] C. Villani. *Topics in Optimal Transportation*, volume **58** of *Grad. Stud. Math.* American Mathematical Society, Providence, RI, 2003.

[Vil06] C. Villani. Mathematics of Granular Materials. *J. Stat. Phys.*, **124**(2–4):781–822, 2006.

[Vil09] C. Villani. *Optimal Transport — Old and New*, volume **338** of *Grundlehren Math. Wiss.* Springer, Berlin, Heidelberg, 2009.

[Wal82] J.B. Walsh. A non reversible semi-martingale. In *Sémin. Probab. XVI*, volume **920** of *Lecture Notes in Math.*, page 212. Springer, Berlin, Heidelberg, 1982.

[Wil91] D. Williams. *Probability with Martingales*. Cambridge Math. Textbooks. Cambridge University Press, 1991.

[WZ23] J. Wiesel and E. Zhang. An optimal transport-based characterization of convex order. *Depend. Model.*, **11**(1):20230102, 2023.

Concluding discussion

This dissertation establishes a trajectorial approach to the characterization of diffusion processes as gradient flows in Wasserstein space. The probabilistic viewpoint leads to consequences and applications for related classical topics such as entropy dissipation, variational characterizations, and functional inequalities. For different models of diffusions, appropriate probabilistic frameworks are defined in order to provide trajectorial interpretations of these topics. Given that our approach is based on stochastic analysis and time-reversal, the corresponding results are particularly relevant for an audience with a background in probability theory. More precisely, the gradient flow property is established via a perturbation analysis by means of stochastic calculus, where either the drift or volatility coefficient of the underlying diffusion is used as a control variable. The language of probability theory then allows for a rather intuitive interpretation of gradient flows as curves of maximal slope for the corresponding relative entropy functional on the quadratic Wasserstein space.

Another advantage of our probabilistic approach is that it is robust in the sense that it allows for various models of diffusions. In Parts I and II, where diffusions of Langevin–Smoluchowski type are studied, our methods of choice are classical Itô calculus and time-reversal. In contrast, Part III deals with McKean–Vlasov diffusions, which are nonlocal (or nonlinear) in the sense that the drift term depends on the distribution of the state variable. As a consequence, the Itô theory needs to be replaced by Lions' differential calculus over Wasserstein space. A further difficulty is that, even when the stationary distribution of the McKean–Vlasov diffusion exists, it does not have a closed-form expression and it is not even unique in general. Contrary to Parts I – III, where the diffusion coefficient is constant, in Part IV we consider time-homogeneous diffusion processes with variable diffusion coefficients. In this setting, the diffusion processes can be viewed as gradient flows in quadratic Wasserstein spaces induced by a Riemannian metric, where the underlying metric tensor is the inverse of the volatility matrix.

While we are able to explicitly compute the trajectorial rate of relative entropy dissipation on the level of stochastic processes in all the settings described above, this is not the case for the Wasserstein distance itself. We leave the derivation of the Wasserstein metric derivative using probabilistic tools, as well as trajectorial interpretations in this context, as open questions.

The second topic of this thesis deals with the martingale version of the Benamou–Brenier problem of classical optimal transport in continuous time. The main result of Part VI provides a characterization of optimizers as martingales induced by gradients of convex functions, similar to the characterization of optimizers in the classical transport problem for quadratic distance cost. This result gives a definitive structural description

of optimizers in terms of Bass martingales, in the case of an irreducible pair of time-marginal distributions. We also establish in Part VI a novel dual formulation of the martingale Benamou–Brenier problem in terms of convex functions.

In Part VII we introduce the so-called Bass functional. We show the equivalence between the existence of minimizers of the Bass functional and the existence of a Bass martingale with prescribed marginals. This complements the convex duality approach of Part VI with a purely variational perspective. We also prove the displacement convexity of the Bass functional along certain generalized geodesics in the quadratic Wasserstein space.

While leading to several important consequences, at the same time the results of Parts VI and VII raise some interesting open problems, for example, a complete characterization of the general, non-irreducible case in $\mathbb{R}^d$. We will treat this problem in the follow-up work [ST24]. More precisely, we will show that there is a paving of $\mathbb{R}^d$ into relatively open convex sets such that the stretched Brownian motion decomposes into a (possibly uncountable) family of Bass martingales on these sets. This paving coincides with the irreducible convex paving introduced by H. De March and N. Touzi in the article [DMT19].

Several open questions also arise in the context of extending the results of Parts VI and VII beyond the Gaussian case. Part VIII offers some first results into this direction. In particular, we introduce the notion of q-Bass martingales in discrete time, and give sufficient conditions for their existence. We leave the definition of general q-Bass martingales in continuous time as an open question. Finally, we raise the question under which assumptions a q-Bass martingale with given marginals μ and ν exists. Given an irreducible pair (μ, ν), we conjecture that a q-Bass martingale from μ to ν exists, provided that q is absolutely continuous and satisfies suitable integrability conditions.

Bertram Tschiderer

Curriculum vitae sites.google.com/view/bertramtschiderer/

Education

Since 2019 **Doctoral Program in Mathematics**, *University of Vienna*
2016–2019 **MSc in Mathematics**, *University of Vienna*
2012–2015 **BSc in Technical Mathematics**, *University of Innsbruck*

Publications

1. B. Tschiderer and L.C. Yeung. A trajectorial approach to relative entropy dissipation of McKean–Vlasov diffusions: gradient flows and HWBI inequalities. *Bernoulli*, **29**(1):725–756, 2023.

2. I. Karatzas and B. Tschiderer. A variational characterization of Langevin–Smoluchowski diffusions. In G. Yin and T. Zariphopoulou, editors, *Stochastic Analysis, Filtering, and Stochastic Optimization — A Commemorative Volume to Honor Mark H.A. Davis's Contributions*, pages 239–265. Springer, Cham, 2022.

3. 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*, **66**(4):668–707, 2022.

Preprints

1. B. Tschiderer. q-Bass martingales. arXiv:2402.05669, 2024.

2. B. Tschiderer. Diffusion processes as Wasserstein gradient flows via stochastic control of the volatility matrix. arXiv:2310.18678, 2023.

3. J. Backhoff-Veraguas, W. Schachermayer, and B. Tschiderer. The Bass functional of martingale transport. arXiv:2309.11181, 2023.

4. J. Backhoff-Veraguas, M. Beiglböck, W. Schachermayer, and B. Tschiderer. The structure of martingale Benamou–Brenier in $\mathbb{R}^d$. arXiv:12306.11019, 2023.

Unpublished

1. I. Karatzas, W. Schachermayer, and B. Tschiderer. Trajectorial Otto calculus. arXiv:1811.08686, 2020.

In preparation

1. W. Schachermayer and B. Tschiderer. The decomposition of stretched Brownian motion into Bass martingales.