Trajectorial Otto calculus*

Ioannis Karatzas† Walter Schachermayer‡ Bertram Tschiderer§

31st May 2019

Abstract. We revisit the variational characterization of diffusion as entropic gradient flux, established by Jordan, Kinderlehrer, and Otto in [**?**], and provide for it a probabilistic interpretation based on stochastic calculus. It was shown in [**?**] that, for diffusions of Langevin-Smoluchowski type, the Fokker-Planck probability density flow minimizes 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 both the forward and, most transparently, the backward, directions of time, using a very direct perturbation analysis. By averaging our trajectorial results with respect to the underlying measure on path space, we establish the minimum rate of entropy dissipation along the Fokker-Planck flow and measure exactly the deviation from this minimum that corresponds to any given perturbation. As a bonus of the perturbation analysis, we derive the so-called HWI inequality relating relative entropy (H), Wasserstein distance (W) and relative Fisher information (I).

Key Words and Phrases: Entropy, Fisher information, Wasserstein distance, optimal transport, gradient flux, diffusion, time reversal, functional inequalities.

2010 Mathematics Subject Classification: 60G44, 60H30, 60J60, 82C31, 94A17.

---

*We thank Luigi Ambrosio, Mathias Beiglböck, Max Fathi, Ivan Gentil, David Kinderlehrer, Daniel Lacker, Michel Ledoux, Jan Maas, Felix Otto, Chris Rogers, Oleg Szehr, and Lanston Lane Chun Yeung for their advice and comments during the preparation of this paper. Special thanks go to Luigi Ambrosio, Michel Ledoux and Jan Maas for their expert guidance, which helped us navigate several difficult narrows successfully.

Preliminary versions of this work, under the titles "Applying Itô calculus to Otto calculus" and "Pathwise Otto calculus", were posted on arXiv on 21 November 2018 (arXiv:1811.08686v1) and 6 February 2019 (arXiv:1811.08686v2), respectively.

I. Karatzas acknowledges support from the National Science Foundation (NSF) under grant NSF-DMS-14-05210. W. Schachermayer and B. Tschiderer acknowledge support by the Austrian Science Fund (FWF) under grant P28861. W. Schachermayer additionally appreciates support by the Vienna Science and Technology Fund (WWTF) through projects MA14-008 and MA16-021.

Most of this work was done in Fall 2018, when W. Schachermayer was visiting the Mathematics Department at Columbia University as Minerva Foundation Fellow.

†Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, USA
(email: ikl@columbia.edu);
and INTECH Investment Management, One Palmer Square, Suite 441, Princeton, NJ 08542, USA
(email: ikaratzas@intechjanus.com).

‡Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
(email: walter.schachermayer@univie.ac.at);
and Department of Mathematics, Columbia University, 2990 Broadway, New York, NY 10027, USA.

§Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
(email: bertram.tschiderer@gmail.com).

Contents

1 Introduction

We provide a trajectorial interpretation of a seminal result by Jordan, Kinderlehrer, and Otto [**?**], and present a proof based on stochastic calculus. 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 Otto's characterization of diffusions of Langevin-Smoluchowski type as entropic gradient fluxes in Wasserstein space, and provide stronger, trajectorial versions of these results. For consistency and better readability we adopt the setting and notation of [**?**], and even copy some paragraphs of this paper almost verbatim in the remainder of this introductory section.

Following the lines of [**?**] we thus consider a Fokker-Planck equation of the form

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

with initial condition

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

Here, ρ 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, $\beta > 0$ is a real constant, and ρ^0 is a probability density on $\mathbb{R}^n$. The solution $\rho(t,x)$ of (**??**) with initial condition (**??**) stays non-negative and conserves its mass, which means that the spatial integral

$$\int_{\mathbb{R}^n} \rho(t,x)\,\mathrm{d}x \tag{1.3}$$

is independent of the time parameter $t \geqslant 0$ and is thus equal to $\int \rho^0\,\mathrm{d}x = 1$. Therefore, $\rho(t,\,\cdot\,)$ must be a probability density on $\mathbb{R}^n$ for every fixed time $t \geqslant 0$.

As in [**?**] we note that the Fokker-Planck equation (**??**) with initial condition (**??**) is inherently related to the stochastic differential equation of Langevin-Smoluchowski type [**?**, **?**, **?**, **?**]

$$\mathrm{d}X(t) = -\nabla\Psi\bigl(X(t)\bigr)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W(t), \qquad X(0) = X^0. \tag{1.4}$$

In the equation above, $(W(t))_{t\geqslant 0}$ is an n-dimensional Brownian motion started from 0, 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 ρ^0 and, unless specified otherwise, the reference measure will always be Lebesgue measure on $\mathbb{R}^n$. Then $\rho(t,\,\cdot\,)$, the solution of (**??**) with initial condition (**??**), gives at any given time $t \geqslant 0$ the probability density function of the random variable $X(t)$ from (**??**).

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

$$Z(\beta) = \int_{\mathbb{R}^n} \mathrm{e}^{-\beta\Psi(x)}\,\mathrm{d}x \tag{1.5}$$

is finite and there exists a unique stationary solution of the Fokker-Planck equation (**??**); namely, the probability density ρ_s of the Gibbs distribution given by [**?**, **?**, **?**]

$$\rho_s(x) = \bigl(Z(\beta)\bigr)^{-1}\,\mathrm{e}^{-\beta\Psi(x)} \tag{1.6}$$

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

In [**?**] it is shown that the stationary probability density ρ_s satisfies the following variational principle: it minimizes the *free energy functional*

$$F(\rho) = E(\rho) + \beta^{-1}\,S(\rho) \tag{1.7}$$

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

$$E(\rho) := \int_{\mathbb{R}^n} \Psi \rho \, \mathrm{d}x \tag{1.8}$$

models the potential energy, whereas the internal energy is given by the negative of the Gibbs-Boltzmann entropy functional

$$S(\rho) := \int_{\mathbb{R}^n} \rho \log \rho \, \mathrm{d}x. \tag{1.9}$$

Similarly as in [**?**, Theorem 5.1] we consider the following regularity assumptions.

Assumptions 1.1.

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

(ii) The distribution of $X(0)$ in (**??**) has a probability density function $\rho^0(x)$ with respect to Lebesgue measure on $\mathbb{R}^n$, which has finite second moment as well as finite free energy, i.e.,

$$\int_{\mathbb{R}^n} \rho^0(x) \, |x|^2 \, \mathrm{d}x < \infty \qquad \text{and} \qquad F(\rho^0) \in \mathbb{R}. \tag{1.10}$$

In [**?**] 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. However, we shall impose the following, additional assumptions.

Assumptions 1.2 (Regularity assumptions for the trajectorial results of the present paper)**.** In addition to conditions **??** and **??** of Assumptions **??**, 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

$$\big\langle x, \nabla \Psi(x) \big\rangle_{\mathbb{R}^n} \geqslant -c \, |x|^2 \tag{1.11}$$

for all $x \in \mathbb{R}^n$ with $|x| \geqslant R$.

(iv) The potential Ψ is sufficiently well-behaved to guarantee that the solution of (**??**) is well-defined for all $t \geqslant 0$, and that the solution $(t, x) \mapsto \rho(t, x)$ of (**??**) with initial condition (**??**) 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 $(t, x) \mapsto \nabla \log \rho(t, x)$ 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 [**?**] showing that this is indeed the case.

For the formulation of Theorem **??** 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 the perturbed potential $\Psi + B$ satisfies condition **??**.

The Assumptions **??** 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 $\alpha > 0$. Furthermore, they guarantee that the second-moment condition in (**??**) propagates in time, i.e.,

$$\int_{\mathbb{R}^n} \rho(t, x) \, |x|^2 \, \mathrm{d}x < \infty, \qquad t \geqslant 0; \tag{1.12}$$

see Lemma **??** below. It is important to point out, that these assumptions do not rule out the case when the constant $Z(\beta)$ in (**??**) is infinite; thus, they allow for cases (such as $\Psi = 0$) in which the stationary probability density function ρ_s does not exist. In fact, in [**?**] the authors point out explicitly that, even when the stationary probability density ρ_s is not defined, the free energy (**??**) of a density $\rho(t,x)$ satisfying the Fokker-Planck equation (**??**) with initial condition (**??**) can be defined, provided that the free energy $F(\rho^0)$ is finite.

Assumptions 1.3 (Regularity assumptions regarding the Wasserstein distance)**.** In order to compute explicitly the metric derivative of the quadratic Wasserstein distance along the Fokker-Planck probability density flow, which is the purpose of Section **??**, we require in addition to conditions **??** – **??** of Assumptions **??** that:

(vi) For every $t \geqslant 0$, there exists a sequence of functions $\big(\varphi_m(t,\cdot)\big)_{m\geqslant 1} \subseteq \mathcal{C}_c^\infty(\mathbb{R}^n;\mathbb{R})$, whose gradients $\big(\nabla\varphi_m(t,\cdot)\big)_{m\geqslant 1}$ converge in the norm of $L^2(P(t))$ to the time-dependent velocity field $v(t,\,\cdot\,)$ of (**??**), as $m \to \infty$. Here, $P(t)$ denotes the probability measure on the Borel sets of $\mathbb{R}^n$ with density $\rho(t,\,\cdot\,)$.

Remark 1.4. The last-mentioned requirement guarantees, for every $t \geqslant 0$, that the time-dependent 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 [**?**, Definition 8.4.1]. For the details we refer to Section **??**, in particular, the display (**??**). However, we do not know whether this condition **??** in Assumptions **??** is actually an additional requirement, or whether it is automatically satisfied in our setting. But as this issue only affects the Wasserstein distance and has no relevance for our novel trajectorial results, which constitute the main point of this work, we will not pursue this problem any further.

The condition **??** in Assumptions **??** is satisfied by simple potentials such as for example $\Psi \equiv 0$ or $\Psi(x) = |x|^2/4$. More generally, potentials with a curvature lower bound $\mathrm{Hess}(\Psi) \geqslant \kappa\, I_n$, for some $\kappa \in \mathbb{R}$ (as in (**??**) below), for instance the double-well potential $\Psi(x) = (x^2 - \alpha^2)^2$ on the real line, satisfy this condition; this follows from [**?**, Theorem 10.4.13]. ◇

1.1 Preview

We set up in Section **??** our model for the Langevin-Smoluchowski diffusion and introduce its fundamental quantities, such as the current and the invariant distribution of particles, the resulting likelihood ratio process, as well as the associated concepts of free energy, relative entropy, and relative Fisher information.

Section **??** presents our basic results. These include Theorem **??**, which computes in terms of the relative Fisher information the rate of relative entropy decay in the ambient Wasserstein space of probability density functions with finite second moment; as well as its "perturbed" counterpart, Theorem **??**. We compute explicitly the difference between these perturbed and unperturbed rates, and show that it is always non-negative, in fact strictly positive unless the perturbation and the gradient of the log-likelihood ratio are collinear. This way, the Langevin-Smoluchowski diffusion emerges as the *steepest descent* (or "gradient flux") of the relative entropy functional with respect to the Wasserstein metric.

We also show that both Theorems **??** and **??** follow as very simple consequences of their stronger, trajectorial versions, Theorems **??** and **??**, respectively. These latter are the main results of this work; they provide very detailed descriptions of the semimartingale dynamics for the relative entropy process, in both its "pure" and "perturbed" forms. Such descriptions are most transparent when time is *reversed*, so we choose to present them primarily in this context. Several important consequences and ramifications of Theorems **??**, **??** are developed in Subsections **??** and **??**, including a derivation of the famous HWI inequality (Otto and Villani [**?**, **?**, **?**], Cordero-Erausquin [**?**]) that relates relative entropy (H) to Wasserstein distance (W) and to relative Fisher information (I).

Most of the detailed arguments and proofs are collected in Section **??** and in the appendices. In particular, Appendix **??** presents a completely self-contained account of time reversal for Itô diffusion processes. The necessary background on optimal Wasserstein transport is recalled in Section **??**.

2 The stochastic approach

Thus far, we have been mostly quoting from [**?**]. We adopt now a more probabilistic point of view, and translate our setting into the language of stochastic processes and probability measures. For notational convenience, and without loss of generality, we fix the constant $\beta > 0$ to equal 2, so that the stochastic differential equation (**??**) becomes

$$\mathrm{d}X(t) = -\nabla\Psi\big(X(t)\big)\,\mathrm{d}t + \mathrm{d}W(t), \qquad t \geqslant 0. \tag{2.1}$$

Let $P(0)$ be a probability measure on the Borel sets of $\mathbb{R}^n$ with density function $p^0(x) := \rho^0(x)$. We shall study the stochastic differential equation (**??**) with initial probability distribution $P(0)$.

While we do make an effort to follow the setting and notation of [**?**] as closely as possible, our notation here differs slightly from [**?**]. To conform with our probabilistic approach, we shall use from now onward the familiar letters p^0 and $p(0,\,\cdot\,)$ rather than ρ^0 and $\rho(0,\,\cdot\,)$.

The initial probability measure $P(0)$ on $\mathbb{R}^n$ with density function $p(0,\,\cdot\,)$, induces a probability measure $\mathbb{P}$ on the 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 (**??**) 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 $p(t) \equiv p(t,\,\cdot\,)$ the corresponding probability density function, at each time $t \geqslant 0$. This function solves the equation (**??**).

We shall see in Appendix **??** that, in conjunction with the second-moment condition in (**??**), the drift condition (**??**) guarantees finite second moments of the probability density functions $p(t)$ at all times $t \geqslant 0$; equivalently, membership of the probability distribution $P(t)$ in the space $\mathscr{P}_2(\mathbb{R}^n)$ of definition (**??**) in Section **??**, for all $t \geqslant 0$. This property also holds when the potential Ψ is replaced by $\Psi + B$ as in condition **??** of Assumptions **??**; see Lemma **??**.

Lemma 2.1. *Under the* Assumptions **??**, *the Langevin-Smoluchowski diffusion equation* (**??**) *with initial distribution* $P(0)$ *admits a pathwise unique, strong solution, which satisfies* $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$ *for all* $t \geqslant 0$.

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

$$\frac{\mathrm{d}P(t)}{\mathrm{dQ}}\big(X(t)\big) = \ell\big(t, X(t)\big), \qquad \text{where} \quad \ell(t,x) := \frac{p(t,x)}{q(x)} = p(t,x)\,\mathrm{e}^{2\Psi(x)} \tag{2.2}$$

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

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

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

$$H\big(P(t)\,|\,\mathrm{Q}\big) := \mathbb{E}_{\mathbb{P}}\big[\log \ell\big(t, X(t)\big)\big] = \int_{\mathbb{R}^n} \log\left(\frac{p(t,x)}{q(x)}\right) p(t,x)\,\mathrm{d}x, \qquad t \geqslant 0, \tag{2.4}$$

$$I\big(P(t)\,|\,\mathrm{Q}\big) := \mathbb{E}_{\mathbb{P}}\Big[\,\big|\nabla \log \ell\big(t, X(t)\big)\big|^2\,\Big], \qquad t \geqslant 0. \tag{2.5}$$

Remark 2.2. Following the approach of [**?**, Section 2], we show in Appendix **??** that the relative entropy $H(P \,|\, \mathrm{Q})$ is well-defined for every probability measure P in $\mathscr{P}_2(\mathbb{R}^n)$ and takes values in $(-\infty, \infty]$. ◇

The following well-known identity (**??**) states that the relative entropy $H(P(t) \,|\, \mathrm{Q})$ is equal to the free energy $F(p(t, \,\cdot\,))$, up to a multiplicative factor of 2, for all $t \geqslant 0$. In light of condition **??** in Assumptions **??**, this identity implies $H(P(0) \,|\, \mathrm{Q}) \in \mathbb{R}$, so the quantity in (**??**) is finite for $t = 0$; thus, on account of (**??**) below, finite also for $t > 0$.

Lemma 2.3. *Under the* Assumptions **??**, *and along the curve of probability measures* $(P(t))_{t \geqslant 0}$, *the free energy functional in* (**??**) *and the relative entropy in* (**??**) *are related for each* $t \geqslant 0$ *through the equation*

$$2\, F\big(p(t, \,\cdot\,)\big) = H\big(P(t) \,|\, \mathrm{Q}\big). \tag{2.6}$$

Proof. Indeed,

$$\mathbb{E}_{\mathbb{P}}\big[\log \ell(t, X(t))\big] = \mathbb{E}_{\mathbb{P}}\Big[\log\Big(\mathrm{e}^{2\Psi(X(t))}\, p(t, X(t))\Big)\Big] = \mathbb{E}_{\mathbb{P}}\big[2\,\Psi\big(X(t)\big)\big] + \mathbb{E}_{\mathbb{P}}\big[\log p(t, X(t))\big] \tag{2.7}$$

$$= 2\int_{\mathbb{R}^n} \Psi(x)\, p(t,x)\, \mathrm{d}x + \int_{\mathbb{R}^n} p(t,x)\, \log p(t,x)\, \mathrm{d}x, \tag{2.8}$$

which equals $2\, F(p(t, \,\cdot\,))$. □

The identity (**??**) shows that studying the decay of the free energy $F(p(t, \,\cdot\,))$, is equivalent to studying the decay of the relative entropy $H(P(t) \,|\, \mathrm{Q})$, a key aspect of thermodynamics.

Remark 2.4. In conjunction with (**??**), the condition $F(p(0, \,\cdot\,)) \in \mathbb{R}$ in (**??**), and (**??**) – (**??**), the decrease of the relative entropy established in (**??**) shows that $\mathbb{E}_{\mathbb{P}}\big[\Psi(X(t))\big]$ is finite for all $t \geqslant 0$. Thus, if the potential Ψ dominates a quadratic, we deduce that $\mathbb{E}_{\mathbb{P}}\big[|X(t)|^2\big] < \infty$, i.e., $P(t) \in \mathscr{P}_2(\mathbb{R}^n)$, also holds for all $t \geqslant 0$, without invoking the coercivity condition (**??**). ◇

3 The theorems

As already indicated in (**??**) and (**??**), the probability density function $p\colon [0, \infty) \times \mathbb{R}^n \to [0, \infty)$ solves the Fokker-Planck or forward Kolmogorov [**?**] equation [**?**, **?**, **?**, **?**]

$$\partial_t p(t,x) = \operatorname{div}\big(\nabla\Psi(x)\, p(t,x)\big) + \tfrac{1}{2}\Delta p(t,x), \qquad (t,x) \in (0,\infty) \times \mathbb{R}^n, \tag{3.1}$$

with initial condition

$$p(0,x) = p^0(x), \qquad x \in \mathbb{R}^n. \tag{3.2}$$

By contrast, the function $q(\,\cdot\,)$ does not depend on the temporal variable, and solves the stationary version of the forward Kolmogorov equation (**??**), namely

$$0 = \operatorname{div}\big(\nabla\Psi(x)\, q(x)\big) + \tfrac{1}{2}\Delta q(x), \qquad x \in \mathbb{R}^n. \tag{3.3}$$

In light of Lemma **??**, the object of interest in [**?**] is to relate the decay of the *relative entropy functional*

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P \,|\, \mathrm{Q}) \in (-\infty, \infty] \tag{3.4}$$

along the curve $(P(t))_{t \geqslant 0}$, to the quadratic Wasserstein distance W_2 defined in (**??**) of Section **??**. We resume the remarkable relation between these two quantities in the following two theorems; these provide a way to quantify the relationship between displacement in the ambient space (the denominator of the expression in (**??**)) and fluctuations of the free energy, or equivalently of the relative entropy (the numerator in the expression (**??**)). The proofs will be given in Subsection **??** below.

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

$$\lim_{t\to t_0} \frac{H(P(t)\,|\,\mathrm{Q}) - H(P(t_0)\,|\,\mathrm{Q})}{t - t_0} = -\tfrac{1}{2}\, I(P(t_0)\,|\,\mathrm{Q}), \tag{3.5}$$

as well as the local behavior of the quadratic Wasserstein distance

$$\lim_{t\to t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \tfrac{1}{2}\, \sqrt{I(P(t_0)\,|\,\mathrm{Q})}, \tag{3.6}$$

so that

$$\lim_{t\to t_0} \left(\operatorname{sgn}(t - t_0) \cdot \frac{H(P(t)\,|\,\mathrm{Q}) - H(P(t_0)\,|\,\mathrm{Q})}{W_2(P(t), P(t_0))} \right) = -\sqrt{I(P(t_0)\,|\,\mathrm{Q})}. \tag{3.7}$$

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

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

$$I(P(t_0)\,|\,\mathrm{Q}) = \mathbb{E}_{\mathbb{P}}\left[\, \frac{\left|\nabla\ell(t_0, X(t_0))\right|^2}{\ell(t_0, X(t_0))^2} \,\right] = \int_{\mathbb{R}^n} \left| \frac{\nabla p(t_0, x)}{p(t_0, x)} + 2\,\nabla\Psi(x) \right|^2 p(t_0, x)\,\mathrm{d}x. \tag{3.8}$$

Remark 3.2. Under the Assumptions **??** it is perfectly possible for the relative Fisher information $I(P(t_0)\,|\,\mathrm{Q})$ to be infinite at $t_0 = 0$. For instance, think of $p(0,\,\cdot\,)$ as the indicator function of a subset of $\mathbb{R}^n$ with Lebesgue measure equal to 1.

For future reference, we denote by N the set of exceptional points $t_0 \geqslant 0$ for which the right-sided limiting assertion

$$\lim_{t\downarrow t_0} \frac{H(P(t)\,|\,\mathrm{Q}) - H(P(t_0)\,|\,\mathrm{Q})}{t - t_0} = -\tfrac{1}{2}\, I(P(t_0)\,|\,\mathrm{Q}) \tag{3.9}$$

fails. According to Theorem **??**, this exceptional set N has zero Lebesgue measure. ◇

The remarkable insight of [**?**] states that the slope in (**??**) in the direction of the curve $(P(t))_{t\geqslant 0}$ is, in fact, the slope of *steepest descent* for the relative entropy functional (**??**) at the point $P = P(t_0)$. 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 of class $\mathcal{C}^\infty(\mathbb{R}^n; \mathbb{R})$ with compact support, as in condition **??** of Assumptions **??**. This vector field β will serve as a perturbation, and we consider the thus perturbed Fokker-Planck equation

$$\partial_t p^\beta(t, x) = \operatorname{div}\Big(\big(\nabla\Psi(x) + \beta(x)\big)\, p^\beta(t, x) \Big) + \tfrac{1}{2}\Delta p^\beta(t, x), \qquad (t, x) \in (t_0, \infty) \times \mathbb{R}^n \tag{3.10}$$

with initial condition

$$p^\beta(t_0, x) = p(t_0, x), \qquad x \in \mathbb{R}^n. \tag{3.11}$$

We denote by $\mathbb{P}^\beta$ the probability measure on the 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

$$\mathrm{d}X(t) = -\Big(\nabla\Psi\big(X(t)\big) + \beta\big(X(t)\big) \Big)\,\mathrm{d}t + \mathrm{d}W^\beta(t), \qquad t \geqslant t_0 \tag{3.12}$$

with initial probability distribution $P(t_0)$. Here, the process $(W^\beta(t))_{t\geqslant t_0}$ is a 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 $t \geqslant t_0$; once again, the corresponding probability density function $p^\beta(t) \equiv p^\beta(t, \cdot\,)$ solves the equation (**??**) subject to the initial condition (**??**).

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

Lemma 3.3. *Under the* Assumptions **??**, *let* $t_0 \geqslant 0$. *Then the perturbed diffusion equation* (**??**) *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 flux, or *steepest descent*, property of the flow $(P(t))_{t\geqslant 0}$ generated by the Langevin-Smoluchowski diffusion (**??**) in the ambient space of probability measures $\mathscr{P}_2(\mathbb{R}^n)$ endowed with the Wasserstein metric.

Theorem 3.4. *Under the* Assumptions **??**, *the following assertions hold for every point* $t_0 \geqslant 0$ *at which the right-sided limiting identity* (**??**) *is valid* (*i.e., every* $t_0 \in \mathbb{R}_+ \setminus N$):
The $\mathbb{R}^n$*-valued random vectors*

$$a := \nabla \log \ell(t_0, X(t_0)) = \nabla \log p(t_0, X(t_0)) + 2\,\nabla\Psi\big(X(t_0)\big)\,, \qquad b := \beta(X(t_0)) \tag{3.13}$$

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

$$\lim_{t\downarrow t_0} \frac{H\big(P^\beta(t)\,|\,\mathrm{Q}\big) - H\big(P^\beta(t_0)\,|\,\mathrm{Q}\big)}{t - t_0} = -\tfrac{1}{2}\, I\big(P(t_0)\,|\,\mathrm{Q}\big) - \langle a, b\rangle_{L^2(\mathbb{P})} = -\tfrac{1}{2}\,\langle a, a + 2b\rangle_{L^2(\mathbb{P})}. \tag{3.14}$$

Furthermore, the local behavior of the quadratic Wasserstein distance (**??**) *in this perturbed context is given by*

$$\lim_{t\downarrow t_0} \frac{W_2\big(P^\beta(t), P^\beta(t_0)\big)}{t - t_0} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}. \tag{3.15}$$

Combining (**??**) *with* (**??**), *and assuming* $\|a + 2b\|_{L^2(\mathbb{P})} > 0$, *we have*

$$\lim_{t\downarrow t_0} \frac{H\big(P^\beta(t)\,|\,\mathrm{Q}\big) - H\big(P^\beta(t_0)\,|\,\mathrm{Q}\big)}{W_2\big(P^\beta(t), P^\beta(t_0)\big)} = -\left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}, \tag{3.16}$$

and therefore

$$\lim_{t\downarrow t_0} \left(\frac{H\big(P^\beta(t)\,|\,\mathrm{Q}\big) - H\big(P^\beta(t_0)\,|\,\mathrm{Q}\big)}{W_2\big(P^\beta(t), P^\beta(t_0)\big)} - \frac{H\big(P(t)\,|\,\mathrm{Q}\big) - H\big(P(t_0)\,|\,\mathrm{Q}\big)}{W_2\big(P(t), P(t_0)\big)} \right) \tag{3.17}$$

$$= \|a\|_{L^2(\mathbb{P})} - \left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \tag{3.18}$$

Remark 3.5. On the strength of the Cauchy-Schwarz inequality, the expression in (**??**) is non-negative, and vanishes if and only if a and b are collinear. 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 (**??**) is strictly positive. In other words, the slope quantified by the first term of the difference (**??**), is then strictly bigger than the (negative) slope expressed by the second term of (**??**). ⋄

These two theorems are essentially well known. They build upon a vast amount of previous work. In the quadratic case $\Psi(x) = |x|^2/4$, i.e., when the process $(X(t))_{t\geqslant 0}$ in (**??**) is Ornstein-Uhlenbeck with invariant measure in (**??**) standard Gaussian, the relation

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, H\big(P(t)\,|\,\mathrm{Q}\big) = -\tfrac{1}{2}\, I\big(P(t)\,|\,\mathrm{Q}\big) \tag{3.19}$$

has been known since [**?**] 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 [**?**] by Cover and Thomas for an account of the results by Barron, Blachman, Brown, Linnik, Rényi, Shannon, Stam and many others in this vein, as well as to the book [**?**] by Villani. See also the paper by Carlen and Soffer [**?**] and the book by O. Johnson [**?**] on the relation of (**??**) to the central limit theorem, and the paper by Fathi [**?**] for connections with large deviations.

In (**??**), the de Bruijn identity (**??**) is established for more general measures Q, those that satisfy Assumptions **??**; in a similar vein, see also the seminal work [**?**] by Bakry and Émery.

The paper [**?**] 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 $(P(t,\,\cdot\,))_{t\geqslant 0}$ as a gradient flux. This relation was further investigated by Otto in the paper [**?**], where the theory now known as *"Otto calculus"* was developed.

The statements of our Theorems **??**, **??** complement the existing results in some details, e.g., the precise form (**??**), measuring the difference of the two slopes appearing in (**??**). The main novelty of our approach, however, will only become apparent with the formulation of the trajectorial Theorems **??**, **??** below.

We shall thus investigate Theorems **??** and **??** in a trajectorial fashion, by considering the *relative entropy process*

$$\log \ell(t, X(t)) = \log\left(\frac{p(t, X(t))}{q(X(t))}\right) = \log p(t, X(t)) + 2\,\Psi\big(X(t)\big)\,, \qquad 0 \leqslant t \leqslant T, \tag{3.20}$$

along each trajectory of the canonical coordinate process $(X(t))_{0\leqslant t\leqslant T}$, and calculating its dynamics (stochastic differential) under the probability measure $\mathbb{P}$. The expectation with respect to $\mathbb{P}$ of this quantity is, of course, the relative entropy in (**??**).

A decisive tool in the analysis of the relative entropy process (**??**) is to reverse time, and use a remarkable insight due to Fontbona and Jourdain [**?**]. These authors consider the canonical coordinate process $(X(t))_{0\leqslant t\leqslant T}$ on the 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 $(X(T-s))_{0\leqslant s\leqslant T}$; it is then notationally convenient to consider a finite time interval $[0,T]$, rather than $\mathbb{R}_+$. Of course, this does not restrict the generality of the arguments.

At this stage it is 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

$$\mathcal{F}(t) = \sigma\big(X(u)\colon\, 0 \leqslant u \leqslant t\big), \qquad t \geqslant 0; \tag{3.21}$$

and we call $(\mathcal{F}(t))_{t\geqslant 0}$ the "filtration generated by $(X(t))_{t\geqslant 0}$". Likewise, we let $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ be the "filtration generated by the time-reversed canonical coordinate process $(X(T-s))_{0\leqslant s\leqslant T}$" in the same sense as before. In particular,

$$\mathcal{G}(T-s) = \sigma\big(X(T-u)\colon\, 0 \leqslant u \leqslant s\big), \qquad 0 \leqslant s \leqslant T, \tag{3.22}$$

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, consult Section 2.7 in [**?**]; in particular, Problems 7.1 – 7.6 and Proposition 7.7.

3.1 Main results

The following two Theorems **??** and **??** are the main new results of this paper. They can be regarded as trajectorial versions of Theorems **??** and **??**, whose proofs will follow from Theorems **??** and **??** simply by taking expectations. Similar trajectorial approaches have already been applied successfully to the theory of optimal stopping in [**?**], to Doob's martingale inequalities in [**?**], and to the Burkholder-Davis-Gundy inequality in [**?**].

The significance of Theorem **??** right below, is that the trade-off between the decay of relative entropy and the "Wasserstein transportation cost", both of which are characterized in terms of the cumulative relative Fisher information process, is valid not only in expectation, but also along (almost) each trajectory, provided we run time in the reverse direction.[1]

Theorem 3.6. *Under the* Assumptions **??**, *we let* $T > 0$ *and define the cumulative relative Fisher information process, accumulated from the right, as*

$$\begin{aligned} F(T-s) &:= \int_0^s \frac{1}{2} \frac{\bigl|\nabla \ell\bigl(T-u, X(T-u)\bigr)\bigr|^2}{\ell\bigl(T-u, X(T-u)\bigr)^2} \, \mathrm{d}u \\ &= \int_0^s \frac{1}{2} \biggl| \frac{\nabla p\bigl(T-u, X(T-u)\bigr)}{p\bigl(T-u, X(T-u)\bigr)} + 2\, \nabla \Psi\bigl(X(T-u)\bigr) \biggr|^2 \, \mathrm{d}u \end{aligned} \tag{3.23}$$

for $0 \leqslant s \leqslant T$. *Then* $\mathbb{E}_{\mathbb{P}}[F(0)] = \frac{1}{2} \int_0^T I\bigl(P(t) \,|\, \mathrm{Q}\bigr) \, \mathrm{d}t < \infty$, *and the process*

$$M(T-s) := \Bigl(\log \ell\bigl(T-s, X(T-s)\bigr) - \log \ell\bigl(T, X(T)\bigr) \Bigr) - F(T-s) \tag{3.24}$$

for $0 \leqslant s \leqslant T$, *is a square-integrable martingale of the backwards filtration* $(\mathcal{G}(T-s))_{0 \leqslant s \leqslant T}$ *under the probability measure* $\mathbb{P}$. *More explicitly, the martingale of* (**??**) *can be represented as*

$$M(T-s) = \int_0^s \biggl\langle \frac{\nabla \ell\bigl(T-u, X(T-u)\bigr)}{\ell\bigl(T-u, X(T-u)\bigr)} \,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u) \biggr\rangle_{\mathbb{R}^n}, \qquad 0 \leqslant s \leqslant T, \tag{3.25}$$

where the stochastic process $\bigl(\overline{W}^{\mathbb{P}}(T-s)\bigr)_{0 \leqslant s \leqslant T}$ *is a* $\mathbb{P}$*-Brownian motion of the backwards filtration* $(\mathcal{G}(T-s))_{0 \leqslant s \leqslant T}$. *In particular, the quadratic variation of the martingale of* (**??**) *is given by the non-decreasing process in* (**??**), *up to a multiplicative factor of* $1/2$.

Remark 3.7. The finiteness of the expression $\mathbb{E}_{\mathbb{P}}[F(0)] = \frac{1}{2} \int_0^T I\bigl(P(t) \,|\, \mathrm{Q}\bigr) \, \mathrm{d}t$, in conjunction with the representation (**??**), shows that the martingale of (**??**) is bounded in $L^2(\mathbb{P})$. ⋄

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

$$\ell^{\beta}(t,x) := \frac{p^{\beta}(t,x)}{q(x)} = p^{\beta}(t,x)\, \mathrm{e}^{2\Psi(x)} \,, \qquad (t,x) \in [t_0, \infty) \times \mathbb{R}^n. \tag{3.26}$$

Its stochastic analogue is the *perturbed relative entropy process*

$$\log \ell^{\beta}\bigl(t, X(t)\bigr) = \log \left(\frac{p^{\beta}\bigl(t, X(t)\bigr)}{q\bigl(X(t)\bigr)} \right) = \log p^{\beta}\bigl(t, X(t)\bigr) + 2\, \Psi\bigl(X(t)\bigr) \,, \qquad t_0 \leqslant t \leqslant T. \tag{3.27}$$

[1] As David Kinderlehrer kindly pointed out to the second named author, the implicit Euler scheme used in [**?**] also reflects the idea of going back in time at each step of the discretization.

Theorem 3.8. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *We define the perturbed cumulative relative Fisher information process, accumulated from the right, as*

$$F^\beta(T-s) := \int_0^s \left(\frac{1}{2} \frac{\left|\nabla \ell^\beta(T-u, X(T-u))\right|^2}{\ell^\beta(T-u, X(T-u))^2} + \Big(\langle \beta \,,\, 2\,\nabla\Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \Big)(X(T-u)) \right) \mathrm{d}u \tag{3.28}$$

for $0 \leqslant s \leqslant T - t_0$. *Then* $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0)\big] < \infty$, *and the process*

$$M^\beta(T-s) := \Big(\log \ell^\beta(T-s, X(T-s)) - \log \ell^\beta(T, X(T)) \Big) - F^\beta(T-s) \tag{3.29}$$

for $0 \leqslant s \leqslant T - t_0$, *is a square-integrable martingale of the backwards filtration* $(\mathcal{G}(T-s))_{0 \leqslant s \leqslant T-t_0}$ *under the probability measure* $\mathbb{P}^\beta$. *More explicitly, the martingale of* (**??**) *can be represented as*

$$M^\beta(T-s) = \int_0^s \left\langle \frac{\nabla \ell^\beta(T-u, X(T-u))}{\ell^\beta(T-u, X(T-u))} \,,\, \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-u) \right\rangle_{\mathbb{R}^n}, \qquad 0 \leqslant s \leqslant T - t_0, \tag{3.30}$$

where the stochastic process $\big(\overline{W}^{\mathbb{P}^\beta}(T-s)\big)_{0 \leqslant s \leqslant T-t_0}$ *is a* $\mathbb{P}^\beta$*-Brownian motion of the backwards filtration* $(\mathcal{G}(T-s))_{0 \leqslant s \leqslant T-t_0}$.

Remark 3.9. The representation (**??**), in conjunction with the finiteness of $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(0)\big]$, shows that the martingale of (**??**) is bounded in $L^2(\mathbb{P}^\beta)$. ◇

3.2 Important consequences

We state now several important consequences of these two basic results, Theorems **??** and **??**. In particular, we indicate how the corresponding assertions in the earlier Theorems **??**, **??** follow directly from these results by taking expectations.

Corollary 3.10 (Dissipation of relative entropy)**.** *Under the* Assumptions **??**, *we have for all* $t, t_0 \geqslant 0$ *the relative entropy identity*

$$H\big(P(t) \,|\, \mathrm{Q}\big) - H\big(P(t_0) \,|\, \mathrm{Q}\big) = \mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{\ell(t, X(t))}{\ell(t_0, X(t_0))} \right) \right] = \mathbb{E}_{\mathbb{P}}\left[\int_{t_0}^t \left(-\frac{1}{2} \frac{\left|\nabla \ell(u, X(u))\right|^2}{\ell(u, X(u))^2} \right) \mathrm{d}u \right]. \tag{3.31}$$

Furthermore, we have for Lebesgue-almost every $t_0 \geqslant 0$ *the* generalized de Bruijn identity

$$\lim_{t \to t_0} \frac{H\big(P(t) \,|\, \mathrm{Q}\big) - H\big(P(t_0) \,|\, \mathrm{Q}\big)}{t - t_0} = -\tfrac{1}{2}\, \mathbb{E}_{\mathbb{P}}\left[\ \frac{\left|\nabla \ell(t_0, X(t_0))\right|^2}{\ell(t_0, X(t_0))^2} \ \right], \tag{3.32}$$

as well as the local behavior of the quadratic Wasserstein distance

$$\lim_{t \to t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \tfrac{1}{2} \left(\mathbb{E}_{\mathbb{P}}\left[\ \frac{\left|\nabla \ell(t_0, X(t_0))\right|^2}{\ell(t_0, X(t_0))^2} \ \right] \right)^{1/2}. \tag{3.33}$$

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

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

According to the Lebesgue differentiation theorem, the monotone function $t \mapsto H(P(t)\,|\,\mathrm{Q})$ is differentiable for Lebesgue-almost every $t_0 \geqslant 0$, in which case (??) leads to the identity (??).

The limiting behavior of the Wasserstein distance (??), for Lebesgue-almost every $t_0 \geqslant 0$, is well known and carefully worked out in [?]; see Section ?? below for the details. In Theorem ?? we will prove the last-mentioned assertion of Corollary ??, claiming that the validity of (??) for some $t_0 \geqslant 0$ implies that the limiting assertion (??) also holds for the same point t_0. □

Proof of Theorem ?? from Theorem ??: Recalling the definition of the relative Fisher information (??) as well as (??), we realize that the limiting assertions (??) and (??) in Theorem ?? correspond to the limits (??) and (??) in the just proved Corollary ??. If $t_0 \geqslant 0$ is chosen so that the limit (??) exists, the last part of Corollary ?? tells us that then the limit (??) exists as well. Therefore, we can divide the first of these limits by the second, in order to obtain the limiting identity (??) of Theorem ?? for Lebesgue-almost every $t_0 \geqslant 0$. □

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

Corollary 3.11 (Dissipation of relative entropy under perturbations)**.** *Under the* Assumptions ??*, we have for all $t \geqslant t_0 \geqslant 0$ the relative entropy identity*

$$\begin{aligned} H(P^\beta(t)\,|\,\mathrm{Q}) - H(P^\beta(t_0)\,|\,\mathrm{Q}) &= \mathbb{E}_{\mathbb{P}^\beta}\left[\log\left(\frac{\ell^\beta(t,X(t))}{\ell^\beta(t_0,X(t_0))}\right)\right] \\ &= \mathbb{E}_{\mathbb{P}^\beta}\left[\int_{t_0}^t\left(-\frac{1}{2}\frac{\left|\nabla\ell^\beta(u,X(u))\right|^2}{\ell^\beta(u,X(u))^2} + \Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n}\Big)(X(u))\right)\mathrm{d}u\right]. \end{aligned} \tag{3.34}$$

Furthermore, for every point $t_0 \geqslant 0$ at which the right-sided limiting assertion (??) is valid (i.e., every $t_0 \in \mathbb{R}_+ \setminus N$), we have also the limiting identities

$$\lim_{t\downarrow t_0}\frac{H(P^\beta(t)\,|\,\mathrm{Q}) - H(P^\beta(t_0)\,|\,\mathrm{Q})}{t-t_0} = \mathbb{E}_{\mathbb{P}}\left[-\frac{1}{2}\frac{\left|\nabla\ell(t_0,X(t_0))\right|^2}{\ell(t_0,X(t_0))^2} + \Big(\operatorname{div}\beta - \big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n}\Big)(X(t_0))\right], \tag{3.35}$$

as well as

$$\lim_{t\downarrow t_0}\frac{W_2(P^\beta(t),P^\beta(t_0))}{t-t_0} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\,\left|\frac{\nabla\ell(t_0,X(t_0))}{\ell(t_0,X(t_0))} + 2\,\beta(X(t_0))\right|^2\,\right]\right)^{1/2}. \tag{3.36}$$

Proof of Corollary ?? from Theorem ??: Taking expectations under the probability measure $\mathbb{P}^\beta$ and using the martingale property of the process in (??) for $T \geqslant t \geqslant t_0$, leads to the identity (??).

In order to derive the limiting identity (??) from the equation (??), some care is needed to show that (??) is valid for every time $t_0 \geqslant 0$ which is not an exceptional point excluded by Theorem ??, or equivalently by Corollary ??. More precisely, if $t_0 \geqslant 0$ is chosen so that the right-sided limit (??) can be derived from (??) in Corollary ?? (i.e., if $t_0 \in \mathbb{R}_+ \setminus N$), we have to show that for the same point t_0 the perturbed equation (??) leads to the identity (??). Colloquially speaking, we want to show that the generalized de Bruijn identity (??) is stable under perturbations; see in this context also Remark ?? below.

We shall verify in Lemma ?? of Subsection ?? 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

$$\left|\frac{\ell^\beta(t,x)}{\ell(t,x)} - 1\right| = \left|\frac{p^\beta(t,x)}{p(t,x)} - 1\right| \leqslant C\,(t-t_0)\,, \qquad (t,x) \in [t_0,T]\times\mathbb{R}^n \tag{3.37}$$

as well as

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_{t_0}^{t}\,\bigg|\nabla\log\bigg(\frac{\ell^{\beta}(u,X(u))}{\ell(u,X(u))}\bigg)\bigg|^{2}\,\mathrm{d}u\bigg]\leqslant C\,(t-t_0)^2\,,\qquad t_0\leqslant t\leqslant T. \tag{3.38}$$

We turn now to the derivation of (**??**) from (**??**). First, as 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 clearly

$$\lim_{t\downarrow t_0}\frac{1}{t-t_0}\,\mathbb{E}_{\mathbb{P}^{\beta}}\bigg[\int_{t_0}^{t}\Big(\operatorname{div}\beta-\big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n}\Big)(X(u))\,\mathrm{d}u\bigg]=\mathbb{E}_{\mathbb{P}^{\beta}}\Big[\Big(\operatorname{div}\beta-\big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n}\Big)(X(t_0))\Big] \tag{3.39}$$

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 (**??**) 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 (**??**), as required.

Regarding the remaining term on the right-hand side of (**??**), it can be seen by applying (**??**) and (**??**), that the equality

$$\lim_{t\downarrow t_0}\frac{1}{t-t_0}\,\mathbb{E}_{\mathbb{P}^{\beta}}\bigg[\int_{t_0}^{t}\bigg(-\frac{1}{2}\frac{\big|\nabla\ell^{\beta}(u,X(u))\big|^2}{\ell^{\beta}(u,X(u))^2}\bigg)\,\mathrm{d}u\bigg]=\lim_{t\downarrow t_0}\frac{1}{t-t_0}\,\mathbb{E}_{\mathbb{P}}\bigg[\int_{t_0}^{t}\bigg(-\frac{1}{2}\frac{\big|\nabla\ell(u,X(u))\big|^2}{\ell(u,X(u))^2}\bigg)\,\mathrm{d}u\bigg] \tag{3.40}$$

holds as long as $t_0\geqslant 0$ is chosen so that one of the limits exists; for the details we refer to [**?**, Section 3.1]. In other words, the existence and equality of the limits in (**??**) is guaranteed if and only if $t_0\in\mathbb{R}_+\setminus N$. It develops that both limits in (**??**) exist if $t_0\geqslant 0$ is not contained in the exceptional set N of zero Lebesgue measure, and their common value is

$$-\tfrac{1}{2}\,I(P(t_0)\,|\,\mathrm{Q})=-\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\,\frac{\big|\nabla\ell(t_0,X(t_0))\big|^2}{\ell(t_0,X(t_0))^2}\,\bigg]; \tag{3.41}$$

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

As regards the final assertion (**??**), we note that, by analogy with (**??**), the limiting behavior of the Wasserstein distance (**??**), for Lebesgue-almost every $t_0\geqslant 0$, is well known [**?**]; for the details we refer to Section **??** below. More precisely, it will follow from Theorem **??** that the limiting assertion

$$\lim_{t\downarrow t_0}\frac{W_2(P(t),P(t_0))}{t-t_0}=\tfrac{1}{2}\,\sqrt{I(P(t_0)\,|\,\mathrm{Q})} \tag{3.42}$$

is valid for all $t_0\in\mathbb{R}_+\setminus N$. Once again, concerning the relation between the limits in (**??**) and (**??**) pertaining to the Wasserstein distance, we discern a similar pattern as in the case of the generalized de Bruijn identity. In fact, Theorem **??** will tell us that the perturbed Wasserstein limit (**??**) also holds for every $t_0\in\mathbb{R}_+\setminus N$. In other words, the local behavior of the quadratic Wasserstein distance is stable under perturbations as well; we shall come back to this point in Remark **??** below.

Summing up, if $t_0\in\mathbb{R}_+\setminus N$, i.e., the limiting identity (**??**) holds, the limiting assertions (**??**) and (**??**) are valid as well. Hence, except for the set N of zero Lebesgue measure in Remark **??**, we have shown the validity of (**??**) and (**??**), thus completing the proof of Corollary **??**. □

Proof of Theorem **??** *from Theorems* **??**, **??**: Let $t_0\in\mathbb{R}_+\setminus N$, i.e., such that the limiting assertion (**??**), and as a consequence also (**??**), are valid (these are the right-sided limits corresponding to (**??**),

(??) in Corollary ?? of Theorem ??). Then the limiting identities (??), (??) from Corollary ?? of Theorem ?? are valid as well. Recalling the abbreviations $a = \nabla \log \ell(t_0, X(t_0))$ and $b = \beta(X(t_0))$ in (??), we summarize now the identities just mentioned as

$$\lim_{t \downarrow t_0} \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2} \, \|a\|^2_{L^2(\mathbb{P})}, \tag{3.43}$$

$$\lim_{t \downarrow t_0} \frac{W_2\big(P(t), P(t_0)\big)}{t - t_0} = \tfrac{1}{2} \, \|a\|_{L^2(\mathbb{P})}, \tag{3.44}$$

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) \,|\, \mathrm{Q}) - H(P^\beta(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2} \, \big\langle a, a + 2b \big\rangle_{L^2(\mathbb{P})}, \tag{3.45}$$

$$\lim_{t \downarrow t_0} \frac{W_2\big(P^\beta(t), P^\beta(t_0)\big)}{t - t_0} = \tfrac{1}{2} \, \|a + 2b\|_{L^2(\mathbb{P})}. \tag{3.46}$$

Indeed, the equations (??), (??), and (??) correspond precisely to (??), (??), and (??), respectively. As for (??), we note that, according to equation (??) of Corollary ??, the limit in (??) equals

$$-\tfrac{1}{2} \, \|a\|^2_{L^2(\mathbb{P})} + \mathbb{E}_{\mathbb{P}}\Big[\Big(\operatorname{div} \beta - 2 \, \big\langle \beta, \nabla \Psi \big\rangle_{\mathbb{R}^n} \Big)\big(X(t_0)\big)\Big]. \tag{3.47}$$

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

$$\mathbb{E}_{\mathbb{P}}\Big[\Big(\operatorname{div} \beta - \big\langle \beta \,,\, 2 \, \nabla \Psi \big\rangle_{\mathbb{R}^n} \Big)\big(X(t_0)\big)\Big] = -\langle a, b \rangle_{L^2(\mathbb{P})}. \tag{3.48}$$

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

$$\int_{\mathbb{R}^n} \Big(\operatorname{div} \beta(x) - \big\langle \beta(x) \,,\, 2 \, \nabla \Psi(x) \big\rangle_{\mathbb{R}^n} \Big) \, p(t_0, x) \, \mathrm{d}x. \tag{3.49}$$

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

$$-\int_{\mathbb{R}^n} \Big\langle \beta(x) \,,\, \nabla \log p(t_0, x) + 2 \, \nabla \Psi(x) \Big\rangle_{\mathbb{R}^n} \, p(t_0, x) \, \mathrm{d}x, \tag{3.50}$$

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

The limiting identities (??) – (??) now clearly imply the assertions of Theorem ??. □

The following two results, Propositions ?? and ??, are trajectorial versions of Corollaries ?? and ??, respectively. They compute the rate of temporal change of relative entropy for the equation (??) and for its perturbed version (??), respectively, in the more precise trajectorial manner of Theorems ??, ??.

Proposition 3.12 (Trajectorial rate of relative entropy dissipation)**.** *Under the* Assumptions ??*, let* $t_0 \geqslant 0$ *be such that the generalized de Bruijn identity* (??) *does hold. Then the relative entropy process*

(**??**) *satisfies, with* $T > t_0$*, the following trajectorial relations:*

$$\lim_{s\uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}}\Big[\log \ell(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell(T-s, X(T-s))}{T-t_0-s} \tag{3.51}$$

$$= \lim_{s\downarrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}}\Big[\log \ell(T-s, X(T-s)) \mid \mathcal{G}(t_0)\Big] - \log \ell(t_0, X(t_0))}{s-(T-t_0)} \tag{3.52}$$

$$= \frac{1}{2}\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2} = \frac{1}{2}\Big|\frac{\nabla p(t_0, X(t_0))}{p(t_0, X(t_0))} + 2\,\nabla\Psi\big(X(t_0)\big)\Big|^2, \tag{3.53}$$

where the limits (**??**) *and* (**??**) *exist in the norm of* $L^1(\mathbb{P})$*.*

Remark 3.13. The limiting assertions (**??**) – (**??**) of Proposition **??** are the conditional trajectorial versions of the generalized de Bruijn identity (**??**). ⋄

Proof of Proposition **??** *from Theorem* **??***:* Let $t_0 \geqslant 0$ be such that the generalized de Bruijn identity (**??**) from Corollary **??** of Theorem **??** is valid, and select $T > t_0$. The martingale property of the process in (**??**) allows us to write the numerator in (**??**) as

$$\mathbb{E}_{\mathbb{P}}\Big[F(t_0) - F(T-s) \mid \mathcal{G}(T-s)\Big], \qquad 0 \leqslant s \leqslant T-t_0, \tag{3.54}$$

in the notation of (**??**). Similarly, the numerator in (**??**) equals $\mathbb{E}_{\mathbb{P}}\big[F(T-s) - F(t_0) \mid \mathcal{G}(t_0)\big]$, $T - t_0 \leqslant s \leqslant T$. By analogy with the derivation of (**??**) from (**??**), where we calculated real-valued expectations, we rely on the Lebesgue differentiation theorem to obtain the corresponding results (**??**) – (**??**) for conditional expectations. Using the left-continuity of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$, we can invoke the measure-theoretic result in Proposition **??** of Appendix **??**, which establishes the claims (**??**) – (**??**) pertaining to conditional expectations. □

Proposition 3.14 (Trajectorial rate of relative entropy dissipation under perturbations)**.** *Under the* Assumptions **??***, let* $t_0 \in \mathbb{R}_+ \setminus N$*. Then the relative entropy process* (**??**) *and its perturbed version* (**??**) *satisfy, with* $T > t_0$*, the following trajectorial relations:*

$$\begin{aligned}\lim_{s\uparrow T-t_0} &\frac{\mathbb{E}_{\mathbb{P}^\beta}\Big[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ &= \frac{1}{2}\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2} - \operatorname{div}\beta\big(X(t_0)\big) + \Big\langle \beta\big(X(t_0)\big)\,,\, 2\,\nabla\Psi\big(X(t_0)\big)\Big\rangle_{\mathbb{R}^n},\end{aligned} \tag{3.55}$$

as well as

$$\begin{aligned}\lim_{s\uparrow T-t_0} &\frac{\mathbb{E}_{\mathbb{P}}\Big[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s)\Big] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ &= \frac{1}{2}\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2} - \Big(\operatorname{div}\beta\big(X(t_0)\big) + \Big\langle \beta\big(X(t_0)\big)\,,\, \nabla \log p(t_0, X(t_0))\Big\rangle_{\mathbb{R}^n}\Big),\end{aligned} \tag{3.56}$$

and

$$\begin{aligned}\lim_{s\uparrow T-t_0} &\frac{\log \ell^\beta(T-s, X(T-s)) - \log \ell(T-s, X(T-s))}{T-t_0-s} \\ &= \operatorname{div}\beta\big(X(t_0)\big) + \Big\langle \beta\big(X(t_0)\big)\,,\, \nabla \log p(t_0, X(t_0))\Big\rangle_{\mathbb{R}^n},\end{aligned} \tag{3.57}$$

where the limits in (**??**) – (**??**) *exist in the norms of both* $L^1(\mathbb{P})$ *and* $L^1(\mathbb{P}^\beta)$.

Remark 3.15. It is perhaps noteworthy that the three limiting expressions in (**??**), (**??**) and (**??**) are quite different from each other. The first limiting assertion (**??**) of Proposition **??** is the conditional trajectorial version of the perturbed de Bruijn identity (**??**). We also note that in fact the third limiting assertion (**??**) is valid for all $t_0 > 0$. ◇

Proof of the assertion (**??**) *in Proposition* **??**, *from Theorem* **??**: Let $t_0 \in \mathbb{R}_+ \setminus N$, i.e., so that the right-sided limiting assertion (**??**) is valid, and select $T > t_0$. In (**??**) from Corollary **??** of Theorem **??** we have seen that the limits in (**??**) and (**??**) have the same exceptional sets, hence also the limiting identity (**??**) holds. Now, for such t_0, we show the limiting assertion (**??**) in the same way as the assertion (**??**) in the proof of Proposition **??** above. Indeed, this time we invoke the $\mathbb{P}^\beta$-martingale property of the process in (**??**), and write the numerator in the first line of (**??**) as $\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0) - F^\beta(T-s) \,\big|\, \mathcal{G}(T-s)\big]$, $0 \leqslant s \leqslant T - t_0$, in the notation of (**??**).

Applying Proposition **??** of Appendix **??** in this situation proves the limiting identity (**??**) with respect to the norm of $L^1(\mathbb{P}^\beta)$. As we shall see in Lemma **??** of Subsection **??** 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 (**??**)). Hence, convergence in $L^1(\mathbb{P})$ is equivalent to convergence in $L^1(\mathbb{P}^\beta)$. This establishes the $L^1(\mathbb{P})$-convergence of (**??**), which completes the proof of the limiting assertion (**??**).

The proofs of the limiting assertions (**??**) and (**??**) are postponed to Subsection **??**. □

3.3 Ramifications

Theorem **??** and, in particular, its equation (**??**) above, show — at least on a formal level — that the functional

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P \,|\, \mathrm{Q}) - H(P(0) \,|\, \mathrm{Q}) \tag{3.58}$$

can be linearly approximated in the neighborhood of $P(0)$ by the functional

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto \langle a, c\rangle_{L^2(\mathbb{P})}, \qquad \text{where} \quad c = -\tfrac{a}{2} - b \tag{3.59}$$

as in (**??**) with $t_0 = 0$ and $a = \nabla \log \ell(0, X(0))$, $b = \beta(X(0))$. As it turns out, formula (**??**) is closely related to the sharpened form of the *HWI inequality* due to Otto and Villani [**?**] (see also Cordero-Erausquin [**?**] and [**?**, p. 650]); we explain presently how.

Consider the starting time $t_0 = 0$ and the curve $(P^\beta(t))_{t\geqslant 0}$ as in Theorems **??** and **??**, for a fixed perturbation β as above, and suppose that this t_0 is not an exceptional point in the preceding limiting assertions. Let us fix $P_1 \in \mathscr{P}_2(\mathbb{R}^n)$ and study the "tangent" $(P_t)_{0\leqslant t\leqslant 1}$ to the curve $(P^\beta(t))_{t\geqslant 0}$ at the point $P^\beta(0) = P(0) = P_0$ in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$, and analyze the behavior of the relative entropy functional (**??**) along the curve $(P_t)_{0\leqslant t\leqslant 1}$. Here $(P_t)_{0\leqslant t\leqslant 1}$ is understood to be a "straight line" in $\mathscr{P}_2(\mathbb{R}^n)$; i.e., using the terminology of McCann [**?**], as the *displacement interpolation* between the elements P_0 and P_1 in $\mathscr{P}_2(\mathbb{R}^n)$.

Once we have identified this tangent $(P_t)_{0\leqslant t\leqslant 1}$, it is geometrically obvious — at least on an intuitive level — that the slope of the relative entropy functional (**??**) along the straight line $(P_t)_{0\leqslant t\leqslant 1}$ should be equal to the slope along $(P^\beta(t))_{t\geqslant 0}$ at the touching point $P^\beta(0) = P_0$. This slope is given by (**??**), and we shall verify in the following Lemma **??** and Proposition **??** that — under suitable regularity assumptions — it coincides with the slope along $(P_t)_{0\leqslant t\leqslant 1}$ as identified by Otto and Villani in [**?**] and Cordero-Erausquin in [**?**].

To work out the connection between (**??**) and [**?**], [**?**] we shall turn things upside down; i.e., we first define the tangent $(P_t)_{0\leqslant t\leqslant 1}$, and then find the corresponding perturbation β so that the curve

$(P^\beta(t))_{t\geqslant 0}$ indeed has $(P_t)_{0\leqslant t\leqslant 1}$ as tangent at the point $P^\beta(0) = P_0$. In this manner, we shall treat β more as an element of "control", than a perturbation.

Fix an element $P \in \mathscr{P}_2(\mathbb{R}^n)$, and let $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ be such that $T(x) := x + \gamma(x)$ transports $P_0 = P(0)$ to $P_1 = P$ optimally with respect to the quadratic Wasserstein distance, i.e., $T_\#(P_0) = P_1$ and $\|\gamma\|_{L^2(P_0)} = W_2(P_0, P_1)$; see also (**??**) and (**??**) below. Here and throughout, $T_\#(P_0)$ denotes the pushforward measure of P_0 by the map T and is given by $(T_\#(P_0))(B) = P_0(T^{-1}(B))$ for every Borel set $B \subseteq \mathbb{R}^n$. Then (**??**), (**??**) suggest that the displacement interpolation $(P_t)_{0\leqslant t\leqslant 1}$ between the two probability measures $P_0 = P(0)$ and $P_1 = P$, to be defined in (**??**) below, is tangent to the curve $(P^\beta(t))_{t\geqslant 0}$ as in Theorems **??** and **??**, if γ and β are related via

$$\gamma(x) = -\tfrac{1}{2}\,\nabla \log \ell(0,x) - \beta(x), \qquad x \in \mathbb{R}^n. \tag{3.60}$$

Then the random variable c of (**??**) becomes $c = \gamma(X(0))$.

We formalize these intuitive geometric insights in the subsequent Lemma **??**, which provides the analogue of (**??**) for the displacement interpolation flow $(P_t)_{0\leqslant t\leqslant 1}$ of (**??**). To this end, we impose temporarily the following strong regularity conditions. As it will turn out in the proof of Proposition **??**, these will not restrict, eventually, the generality of the argument.

Assumptions 3.16 (Regularity assumptions of Lemma **??**)**.** In addition to the conditions **??** – **??** of Assumptions **??**, we also impose that:

(vii) 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 on the interior of their respective supports. Hence there is a map $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ of the form $\gamma(x) = \nabla(G(x) - |x|^2/2)$ 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

$$T_t^\gamma(x) := x + t\cdot\gamma(x) = (1-t)\cdot x + t\cdot \nabla G(x) \qquad \text{and} \qquad P_t := (T_t^\gamma)_\#(P_0) = P_0 \circ (T_t^\gamma)^{-1} \tag{3.61}$$

for $0 \leqslant t \leqslant 1$, and $T_1^\gamma = \nabla G$; 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

$$W_2(P_0, P_t) = t\sqrt{\int_{\mathbb{R}^n} |x - \nabla G(x)|^2 \,\mathrm{d}P_0(x)} = t\,\|\gamma\|_{L^2(P_0)}, \qquad 0 \leqslant t \leqslant 1. \tag{3.62}$$

The (P_0-almost everywhere) unique gradient $T := \nabla G$ of a convex function pushing forward P_0 to P_1, i.e., $T_\#(P_0) = P_1$, and having the optimality property (**??**) with respect to the quadratic Wasserstein distance, is called the *Brenier map*; see [**?**, Theorem 2.12].

Remark 3.17. For the existence and uniqueness of the optimal transport map $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ we refer to [**?**, Theorem 2.12], and for its smoothness to [**?**, Theorem 4.14] as well as [**?**, Remarks 4.15]. These results are known collectively under the rubric of *Brenier's theorem* [**?**]. ◇

Remark 3.18. We remark at this point, that we have chosen the subscript notation for P_t in order to avoid confusion with the probability measure $P(t)$ from our Section **??** here. While $P_0 = P(0)$, the flow $(P_t)_{0\leqslant t\leqslant 1}$ from P_0 to P_1 will have otherwise very little to do with the flow $(P(t))_{t\geqslant 0}$ from $P(0)$ to Q appearing in Theorems **??** and **??** (except for the tangential relation at $P_0 = P(0)$). Similarly, the likelihood ratio function

$$\ell_t(x) = \frac{p_t(x)}{q(x)}, \qquad (t,x) \in [0,1]\times\mathbb{R}^n, \tag{3.63}$$

is different from $\ell(t,\,\cdot\,)$, as now $p_t(\,\cdot\,)$ is the density function of the probability measure P_t. ◇

Let us now return to our general theme, where we consider the potential Ψ and the (possibly only σ-finite) measure Q on the Borel sets of $\mathbb{R}^n$ with density $q(x) = \mathrm{e}^{-2\Psi(x)}$ for $x \in \mathbb{R}^n$.

Lemma 3.19. *Under the* Assumptions **??**, *let X_0 be a random variable with probability distribution $P_0 = P(0)$, defined on some probability space $(S, \mathcal{S}, \nu)$. Then we have*

$$\lim_{t \downarrow 0} \frac{H(P_t \,|\, \mathrm{Q}) - H(P_0 \,|\, \mathrm{Q})}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)}. \tag{3.64}$$

Remark 3.20. The relative entropy $H(P \,|\, \mathrm{Q})$ is well-defined for every $P \in \mathscr{P}_2(\mathbb{R}^n)$, and takes values in $(-\infty, \infty]$; see Appendix **??**. As the displacement interpolation $(P_t)_{0 \leqslant t \leqslant 1}$ is the constant-speed geodesic joining the probability measures P_0 and P_1 in $\mathscr{P}_2(\mathbb{R}^n)$, we see that the relative entropy $H(P_t \,|\, \mathrm{Q})$ is well-defined for every $t \in [0, 1]$. $\diamond$

We relegate the proof of Lemma **??**, which follows a similar (but simpler) line of reasoning as the proof of Theorem **??**, to Appendix **??**. Combining Lemma **??** with well-known arguments, in particular, with a fundamental result on displacement convexity due to McCann [**?**], we derive now the HWI inequality of Otto and Villani [**?**] and Cordero-Erausquin [**?**]; see also [**?**, p. 650]. This result relates the fundamental quantities of relative entropy (H), Wasserstein distance (W) and relative Fisher information (I).

Proposition 3.21 (HWI inequality [**?**])**.** *Under the* Assumptions **??**, *we set $P_0 = P(0)$, fix $P_1 \in \mathscr{P}_2(\mathbb{R}^n)$ with finite relative entropy $H(P_1 \,|\, \mathrm{Q})$, and let γ be as in* (**??**). *We suppose in addition that the potential $\Psi \colon \mathbb{R}^n \to [0, \infty)$ satisfies a curvature lower bound*

$$\mathrm{Hess}(\Psi) \geqslant \kappa\, I_n, \tag{3.65}$$

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

$$H(P_0 \,|\, \mathrm{Q}) - H(P_1 \,|\, \mathrm{Q}) \leqslant -\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1), \tag{3.66}$$

where the random variable X_0, the likelihood ratio function ℓ_0, and the probability measure ν, are as in Lemma **??**.

Remark 3.22. Let us stress that we do not require Q to be a probability measure in the formulation of the HWI inequality (**??**). $\diamond$

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

$$-\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} \leqslant \|\nabla \log \ell_0(X_0)\|_{L^2(\nu)}\; \|\gamma(X_0)\|_{L^2(\nu)}, \tag{3.67}$$

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

$$I(P_0 \,|\, \mathrm{Q}) = \mathbb{E}_\nu\Big[|\nabla \log \ell_0(X_0)|^2 \Big] = \|\nabla \log \ell_0(X_0)\|^2_{L^2(\nu)}, \tag{3.68}$$

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

$$\|\gamma(X_0)\|_{L^2(\nu)} = W_2(P_0, P_1) \tag{3.69}$$

as in (**??**), along with the inequality

$$-\big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)} \leqslant \sqrt{I(P_0 \,|\, \mathrm{Q})}\; W_2(P_0, P_1). \tag{3.70}$$

Inserting (**??**) into (**??**) we obtain the usual form of the HWI inequality

$$H(P_0 \,|\, \mathrm{Q}) - H(P_1 \,|\, \mathrm{Q}) \leqslant W_2(P_0, P_1)\, \sqrt{I(P_0 \,|\, \mathrm{Q})} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1). \tag{3.71}$$

When there is a non-trivial angle between $\nabla \log \ell_0(X_0)$ and $\gamma(X_0)$ in $L^2(\nu)$, the inequality (**??**) gives a sharper bound than (**??**). We refer to the original paper [**?**], as well as to [**?**], [**?**, Chapter 5], [**?**, p. 650] and the recent paper [**?**], for a detailed discussion of the HWI inequality.

Remark 3.23. Let us suppose now that the strong non-degeneracy condition (**??**) holds with $\kappa > 0$, and that Q is a probability measure in $\mathscr{P}_2(\mathbb{R}^n)$. Then the inequality (**??**) contains as special cases the *Talagrand* [**?**] and *logarithmic Sobolev* [**?**] *inequalities*, namely

$$W_2^2(P, \mathrm{Q}) \leqslant \tfrac{2}{\kappa}\, H(P \,|\, \mathrm{Q}), \qquad H(P \,|\, \mathrm{Q}) \leqslant \tfrac{1}{2\kappa}\, I(P \,|\, \mathrm{Q}), \tag{3.72}$$

respectively; just by reading (**??**) first with $(P_0, P_1) = (\mathrm{Q}, P)$, then with $(P_0, P_1) = (P, \mathrm{Q})$ and applying Young's inequality $xy \leqslant x^2/2 + y^2/2$, which is valid for all $x, y \in \mathbb{R}$. On the other hand, and now in the context of Section **??**, the second inequality in (**??**) leads, in conjunction with the generalized de Bruijn identity (**??**) and (**??**), to

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, H\big(P(t) \,|\, \mathrm{Q}\big) \leqslant -\kappa\, H\big(P(t) \,|\, \mathrm{Q}\big), \tag{3.73}$$

and thence to the Bakry-Émery [**?**] *exponential temporal dissipation of the relative entropy*

$$H\big(P(t) \,|\, \mathrm{Q}\big) \leqslant H\big(P(t_0) \,|\, \mathrm{Q}\big)\, \mathrm{e}^{-\kappa(t-t_0)}, \qquad t \geqslant t_0 \tag{3.74}$$

as well as of the Wasserstein distance $W_2(P(t), \mathrm{Q})$ on account of (**??**). For an exposition of the Bakry-Émery theory, which derives also the *exponential temporal dissipation of the relative Fisher information* in the context of Section **??**, we refer to [**?**] and [**?**]. ⋄

The inequality (**??**) is yet another manifestation of the interplay between displacement in the ambient space of probability measures (the quantity $W_2(P_0, P_1)$) and fluctuations of the relative entropy (the quantity $H(P_0 \,|\, \mathrm{Q}) - H(P_1 \,|\, \mathrm{Q})$) as governed by the square root of the Fisher information $\sqrt{I(P_0 \,|\, \mathrm{Q})}$, very much in the spirit of (**??**).

Proof of Proposition **??**. As elaborated in [**?**, Section 9.4] we may assume without loss of generality that P_0 and P_1 satisfy the strong regularity Assumptions **??**. For the existence and smoothness of the optimal transport map γ we refer to Remark **??**.

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 $0 \leqslant t \leqslant 1$. We show that the displacement convexity results of McCann [**?**] imply

$$f''(t) \geqslant \kappa\, W_2^2(P_0, P_1), \qquad 0 \leqslant t \leqslant 1. \tag{3.75}$$

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

$$g(t) := \int_{\mathbb{R}^n} p_t(x) \log p_t(x)\, \mathrm{d}x, \qquad h(t) := 2 \int_{\mathbb{R}^n} \Psi(x)\, p_t(x)\, \mathrm{d}x, \qquad 0 \leqslant t \leqslant 1, \tag{3.76}$$

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

$$g''(t) \geqslant 0, \qquad h''(t) \geqslant \kappa\, W_2^2(P_0, P_1), \qquad 0 \leqslant t \leqslant 1. \tag{3.77}$$

By analogy with Lemma **??** we have $f = g + h$, and conclude that the relative entropy function f is κ-uniformly displacement convex, i.e., its second derivative satisfies (**??**).

We appeal now to Lemma **??**, according to which we have

$$f'(0+) = \lim_{t \downarrow 0} \frac{f(t) - f(0)}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\, \gamma(X_0) \big\rangle_{L^2(\nu)}. \tag{3.78}$$

In conjunction with (**??**) and (**??**), the formula $f(1) = f(0) + f'(0+) + \int_0^1 (1-t) f''(t)\, \mathrm{d}t$ now yields (**??**). □

4 Details and proofs

In this section we complete the proofs of Corollary **??** and Proposition **??**, and provide the proofs of our main results, Theorems **??** and **??**. In fact, all we have to do in order to prove these latter theorems is to apply Itô's formula so as to calculate the dynamics, i.e., the stochastic differentials, of the "pure" and "perturbed" relative entropy processes

$$\log \ell(t, X(t)) = \log\left(\frac{p(t, X(t))}{q(X(t))}\right), \quad t \geqslant 0, \qquad\qquad \log \ell^{\beta}(t, X(t)) = \log\left(\frac{p^{\beta}(t, X(t))}{q(X(t))}\right), \quad t \geqslant t_0 \tag{4.1}$$

under the measures $\mathbb{P}$ and $\mathbb{P}^{\beta}$, respectively. We may (and shall) do this in both the forward and, most importantly, the backward, directions of time.

However, such a brute-force approach does not provide any hint as to why we obtain the remarkable form of the drift term of the time-reversed relative entropy process

$$\log \ell\big(T - s, X(T - s)\big) = \log\left(\frac{p(T - s, X(T - s))}{q(X(T - s))}\right), \qquad 0 \leqslant s \leqslant T, \tag{4.2}$$

as stated in Theorem **??**, namely

$$\mathrm{d} \log \ell\big(T - s, X(T - s)\big) = \left\langle \frac{\nabla \ell(T - s, X(T - s))}{\ell(T - s, X(T - s))}, \mathrm{d}\overline{W}^{\mathbb{P}}(T - s) \right\rangle_{\mathbb{R}^n} + \frac{1}{2} \frac{\big|\nabla \ell(T - s, X(T - s))\big|^2}{\ell\big(T - s, X(T - s)\big)^2} \, \mathrm{d}s, \tag{4.3}$$

for $0 \leqslant s \leqslant T$, with respect to the backwards filtration $(\mathcal{G}(T - s))_{0 \leqslant s \leqslant T}$. Therefore, in order to motivate and illustrate the derivation of the dynamics (**??**), we first impose the additional assumption $\mathrm{Q}(\mathbb{R}^n) < \infty$ (which precludes the case $\Psi \equiv 0$), so as to conform to the setting of [**?**]. This is done in Subsection **??**, which serves purely as motivation; in the remainder of this paper we do not rely on the assumption $\mathrm{Q}(\mathbb{R}^n) < \infty$.

4.1 Some preliminaries

Our first task is to calculate the dynamics of the time-reversed relative entropy process (**??**) under the probability measure $\mathbb{P}$. In order to do this, we start by calculating the stochastic differential of the time-reversed canonical coordinate process $(X(T - s))_{0 \leqslant s \leqslant T}$ under $\mathbb{P}$, which is a well-known and classical theme; see e.g. [**?**, **?**], [**?**], [**?**], [**?**], and [**?**]. For the convenience of the reader we present the theory of time reversal for diffusion processes in Appendix **??**. The idea of time reversal goes back to the ideas of Boltzmann [**?**, **?**, **?**] and Schrödinger [**?**, **?**], as well as Kolmogorov [**?**]. In fact, as we shall recall in Appendix **??**, 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 [**?**, **?**] from 1900; at least, when admitting a good portion of wisdom of hindsight.

Recall that the probability measure $\mathbb{P}$ was defined on the 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 (**??**) with initial probability distribution $P(0)$ for $X(0)$ under $\mathbb{P}$. In other words, the process

$$W(t) = X(t) - X(0) + \int_0^t \nabla \Psi\big(X(u)\big) \, \mathrm{d}u, \qquad t \geqslant 0 \tag{4.4}$$

defines a Brownian motion of the forward filtration $(\mathcal{F}(t))_{t \geqslant 0}$ under the probability measure $\mathbb{P}$, where the integral in (**??**) is to be understood in a pathwise Riemann-Stieltjes sense. Passing to the reverse direction of time, the following classical result is well known to hold under the Assumptions **??**.

Proposition 4.1. *Under the* Assumptions **??**, *we let* $T > 0$. *The process*

$$\overline{W}^{\mathbb{P}}(T-s) := W(T-s) - W(T) - \int_0^s \nabla \log p\big(T-u, X(T-u)\big)\,\mathrm{d}u \tag{4.5}$$

for $0 \leqslant s \leqslant T$, *is a Brownian motion of the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ *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*

$$\mathrm{d}X(T-s) = \Big(\nabla \log p\big(T-s, X(T-s)\big) + \nabla\Psi\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) \tag{4.6}$$

$$= \Big(\nabla \log \ell\big(T-s, X(T-s)\big) - \nabla\Psi\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}}(T-s), \tag{4.7}$$

for $0 \leqslant s \leqslant T$, *with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$.

We provide proofs and references for this result in Theorems **??** and **??** of Appendix **??**.

Before proving Theorem **??** in Subsection **??** — as already announced — we digress now to present the following didactic, illuminating and important special case.

4.2 The case $\mathrm{Q}(\mathbb{R}^n) < \infty$

We shall impose, for the purposes of the present subsection only, the additional assumption

$$\mathrm{Q}(\mathbb{R}^n) = \int_{\mathbb{R}^n} \mathrm{e}^{-2\Psi(x)}\,\mathrm{d}x < \infty. \tag{4.8}$$

Under this assumption, the measure Q on the Borel sets of $\mathbb{R}^n$, introduced in Section **??**, is finite and can thus be re-normalized, so as to become a probability measure. In this manner, it induces a probability measure $\mathbb{Q}$ on the path space $\Omega = \mathcal{C}(\mathbb{R}_+;\mathbb{R}^n)$, under which the canonical coordinate process $(X(t,\omega))_{t\geqslant 0} = (\omega(t))_{t\geqslant 0}$ satisfies the stochastic equation (**??**) with initial probability distribution Q for $X(0)$. And because this distribution is invariant, it is also the distribution of $X(t)$ under $\mathbb{Q}$ for every $t \geqslant 0$.

For the present authors, the eye-opener leading to (**??**) was the subsequent remarkable insight due to Fontbona and Jourdain [**?**]. This provided us with much of the original motivation to start this line of research. The result right below holds in much greater generality (essentially one only needs the Markovian structure of the process $(X(t))_{t\geqslant 0}$) but we only state it in the present setting given by (**??**) under the Assumptions **??** and $\mathrm{Q}(\mathbb{R}^n) = 1$ in (**??**).

Theorem 4.2 (Fontbona-Jourdain theorem [**?**])**.** *Under the* Assumptions **??** *and* $\mathrm{Q}(\mathbb{R}^n) = 1$, *we fix* $T \in (0,\infty)$. *The time-reversed likelihood ratio process*

$$\ell\big(T-s, X(T-s)\big) = \frac{p\big(T-s, X(T-s)\big)}{q\big(X(T-s)\big)}\,, \qquad 0 \leqslant s \leqslant T \tag{4.9}$$

is a martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ *under the probability measure* $\mathbb{Q}$.

Corollary 4.3. *Under the* Assumptions **??** *and* $\mathrm{Q}(\mathbb{R}^n) = 1$, *we fix* $T \in (0,\infty)$. *The time-reversed process*

$$\ell\big(T-s, X(T-s)\big) \cdot \log \ell\big(T-s, X(T-s)\big)\,, \qquad 0 \leqslant s \leqslant T \tag{4.10}$$

is a submartingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ *under the probability measure* $\mathbb{Q}$. *In particular, we have*

$$H\big(P(t)\,|\,\mathrm{Q}\big) \leqslant H\big(P(0)\,|\,\mathrm{Q}\big), \qquad 0 \leqslant t \leqslant T. \tag{4.11}$$

Proof. This is an immediate consequence of Theorem **??**, Jensen's inequality for conditional expectations, and the convexity of the function $f(x) = x \log x$, $x > 0$. □

For the convenience of the reader we recall in Appendix **??** the surprisingly straightforward proof of Theorem **??**. Since this result states that the time-reversed likelihood ratio process (**??**) is a $\mathbb{Q}$-martingale with respect to the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$, we will first state the analogue of Proposition **??** in terms of the probability measure $\mathbb{Q}$ on the path space $\Omega = \mathcal{C}(\mathbb{R}_+;\mathbb{R}^n)$, which is induced by the invariant probability distribution Q on $\mathbb{R}^n$.

Proposition 4.4. *Under the* Assumptions **??** *and* $\mathrm{Q}(\mathbb{R}^n) = 1$*, we fix* $T \in (0,\infty)$*. The process*

$$\overline{W}^{\mathbb{Q}}(T-s) := W(T-s) - W(T) + 2\int_0^s \nabla\Psi\big(X(T-u)\big)\,\mathrm{d}u \tag{4.12}$$

for $0 \leqslant s \leqslant T$*, is a Brownian motion of the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ *under the probability measure* $\mathbb{Q}$*. Moreover, the time-reversed canonical coordinate process* $(X(T-s))_{0\leqslant s\leqslant T}$ *satisfies the stochastic differential equation*

$$\mathrm{d}X(T-s) = -\nabla\Psi\big(X(T-s)\big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{Q}}(T-s), \tag{4.13}$$

for $0 \leqslant s \leqslant T$*, with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$*.*

Again, for the proof of this result, we refer to Theorems **??** and **??** of Appendix **??**.

In the following lemma we determine the drift term that allows one to pass from the $\mathbb{Q}$-Brownian motion $\big(\overline{W}^{\mathbb{Q}}(T-s)\big)_{0\leqslant s\leqslant T}$ to the $\mathbb{P}$-Brownian motion $\big(\overline{W}^{\mathbb{P}}(T-s)\big)_{0\leqslant s\leqslant T}$, and vice versa.

Lemma 4.5. *Under the* Assumptions **??** *and* $\mathrm{Q}(\mathbb{R}^n) = 1$*, we fix* $T \in (0,\infty)$*. For* $0 \leqslant s \leqslant T$*, we have*

$$\mathrm{d}\overline{W}^{\mathbb{Q}}(T-s) = \nabla \log \ell\big(T-s, X(T-s)\big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}}(T-s). \tag{4.14}$$

Proof. One just has to compare the equations (**??**) and (**??**). □

The next corollary is a direct consequence of Theorem **??**, Proposition **??** and Itô's formula.

Corollary 4.6. *Under the* Assumptions **??** *and* $\mathrm{Q}(\mathbb{R}^n) = 1$*, we fix* $T \in (0,\infty)$*. The time-reversed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\mathrm{d}\ell\big(T-s, X(T-s)\big) = \Big\langle \nabla\ell\big(T-s, X(T-s)\big)\,,\, \mathrm{d}\overline{W}^{\mathbb{Q}}(T-s)\Big\rangle_{\mathbb{R}^n} \tag{4.15}$$

and

$$\mathrm{d}\log\ell\big(T-s, X(T-s)\big) = \left\langle \frac{\nabla\ell\big(T-s, X(T-s)\big)}{\ell\big(T-s, X(T-s)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{Q}}(T-s)\right\rangle_{\mathbb{R}^n} - \frac{1}{2}\frac{\big|\nabla\ell\big(T-s, X(T-s)\big)\big|^2}{\ell\big(T-s, X(T-s)\big)^2}\,\mathrm{d}s, \tag{4.16}$$

respectively, for $0 \leqslant s \leqslant T$*, with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$*.*

Proof. To prove (**??**), the decisive insight is provided by Theorem **??** due to Fontbona and Jourdain [**?**]. This implies that the drift term in (**??**) must vanish, so that it suffices to calculate the diffusion term in front of $\mathrm{d}\overline{W}^{\mathbb{Q}}(T-s)$ in (**??**); using (**??**), this is a straightforward task.

We note that the vanishing of the drift term in (**??**) can also be obtained from applying Itô's formula to the process (**??**), using (**??**) as well as the backwards Kolmogorov equation (**??**) for the likelihood ratio function $\ell(t,x)$ and observing that all these terms cancel out. But such a procedure does not provide a hint as to why this miracle happens.

Having said this, we apply Itô's formula to the process (**??**) and use Theorem **??** to obtain (**??**). Assertion (**??**) follows once again by applying Itô's formula to the logarithm of the process (**??**), and using the dynamics (**??**). □

We have now all the ingredients in order to compute, under the additional assumption $\mathrm{Q}(\mathbb{R}^n) = 1$, the dynamics of the time-reversed relative entropy process (**??**) under the probability measure $\mathbb{P}$. Indeed, substituting (**??**) into the stochastic equation (**??**), we see that the process (**??**) satisfies the crucial stochastic differential equation (**??**), for $0 \leqslant s \leqslant T$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$.

4.3 The proof of Theorem ??

We drop now the assumption $\mathrm{Q}(\mathbb{R}^n) < \infty$, and write the Fokker-Planck equation (**??**) as

$$\partial_t p(t,x) = \tfrac{1}{2}\Delta p(t,x) + \big\langle \nabla p(t,x), \nabla\Psi(x)\big\rangle_{\mathbb{R}^n} + p(t,x)\,\Delta\Psi(x), \qquad t > 0. \tag{4.17}$$

The probability density function $p(t,x)$ can be represented in the form

$$p(t,x) = \ell(t,x)\,q(x) = \ell(t,x)\,\mathrm{e}^{-2\Psi(x)}, \qquad t \geqslant 0, \tag{4.18}$$

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

$$\partial_t \ell(t,x) = \tfrac{1}{2}\Delta\ell(t,x) - \big\langle \nabla\ell(t,x), \nabla\Psi(x)\big\rangle_{\mathbb{R}^n}, \qquad t > 0, \tag{4.19}$$

a feature consonant with the fact that the dynamics of the likelihood ratio process are most transparent under *time reversal*. This equation will allow us to develop an alternative way of deriving the dynamics (**??**) and proving Theorem **??**, which does not rely on assumption (**??**) and uses exclusively the probability measure $\mathbb{P}$, as follows.

Proof of Theorem ??. We have to show that the stochastic process $(M(T-s))_{0\leqslant s\leqslant T}$ in (**??**) is a $\mathbb{P}$-martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$, is bounded in $L^2(\mathbb{P})$, and admits the integral representation (**??**).

Step 1. Applying Itô's formula to the time-reversed likelihood ratio process (**??**), and using (**??**) as well as the backwards Kolmogorov equation (**??**) for the likelihood ratio function $\ell(t,x)$, we obtain the stochastic differential equation

$$\frac{\mathrm{d}\ell\big(T-s, X(T-s)\big)}{\ell\big(T-s, X(T-s)\big)} = \left\langle \frac{\nabla\ell\big(T-s, X(T-s)\big)}{\ell\big(T-s, X(T-s)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s)\right\rangle_{\mathbb{R}^n} + \frac{\big|\nabla\ell\big(T-s, X(T-s)\big)\big|^2}{\ell\big(T-s, X(T-s)\big)^2}\,\mathrm{d}s \tag{4.20}$$

as well as its logarithmic version

$$\mathrm{d}\log\ell\big(T-s, X(T-s)\big) = \Big\langle \nabla\log\ell\big(T-s, X(T-s)\big)\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s)\Big\rangle_{\mathbb{R}^n} + \tfrac{1}{2}\big|\nabla\log\ell\big(T-s, X(T-s)\big)\big|^2\mathrm{d}s, \tag{4.21}$$

for $0 \leqslant s \leqslant T$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$. This equation right above is nothing other than the desired stochastic differential equation (**??**) from the beginning of Section **??**.

Step 2. We show first, that the process $(M(T-s))_{0\leqslant s\leqslant T}$ of (**??**), with integral representation (**??**), is a continuous local $\mathbb{P}$-martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$.

By condition **??** of Assumptions **??** the function $(0,\infty) \ni t \mapsto \nabla\log\ell(t,x)$ is continuous for every $x \in \mathbb{R}^n$. Together with the continuity of the paths of the canonical coordinate process $(X(t))_{t\geqslant 0}$, this implies that

$$\int_0^{T-\varepsilon} \frac{\big|\nabla\ell\big(T-u, X(T-u)\big)\big|^2}{\ell\big(T-u, X(T-u)\big)^2}\,\mathrm{d}u < \infty, \qquad \mathbb{P}\text{-almost surely}, \tag{4.22}$$

for every $0 < \varepsilon \leqslant T$. On account of (**??**), the sequence of stopping times (with respect to the backwards filtration)

$$\tau_n := \inf\left\{ t \geqslant 0 \colon \int_0^t \frac{\left|\nabla \ell(T-u, X(T-u))\right|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u \geqslant n \right\} \wedge T, \qquad n \in \mathbb{N}_0 \tag{4.23}$$

is non-decreasing and converges $\mathbb{P}$-almost surely to T. Hence, according to (**??**) and the definition in (**??**), the stopped process $(M^{\tau_n}(T-s))_{0\leqslant s\leqslant T}$ is a uniformly integrable $\mathbb{P}$-martingale of the backwards filtration, for every $n \in \mathbb{N}_0$.

Step 3. To show that, in fact, the process $(M(T-s))_{0\leqslant s\leqslant T}$ is a true $\mathbb{P}$-martingale, we have to rely on the finite free energy condition (**??**) which, in the light of Lemma **??**, asserts that the relative entropy $H(P(0)\,|\,\mathrm{Q})$ is finite.

Taking expectations with respect to the probability measure $\mathbb{P}$ in (**??**) at time $s = \tau_n$, and using that by Step 2 the stopped process $(M^{\tau_n}(T-s))_{0\leqslant s\leqslant T}$ is a true $\mathbb{P}$-martingale with respect to the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$, we get

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{\tau_n} \frac{1}{2} \frac{\left|\nabla \ell(T-u, X(T-u))\right|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u \right] = H\big(P(T-\tau_n)\,|\,\mathrm{Q}\big) - H\big(P(T)\,|\,\mathrm{Q}\big) \tag{4.24}$$

$$\leqslant H\big(P(0)\,|\,\mathrm{Q}\big) - H\big(P(T)\,|\,\mathrm{Q}\big), \tag{4.25}$$

for every $n \in \mathbb{N}_0$. The inequality in (**??**) is justified by the decrease of the relative entropy function $t \mapsto H(P(t)\,|\,\mathrm{Q})$, which we have from (**??**). However, in Corollary **??** we had assumed that Q is a probability measure, which is in general not the case under the Assumptions **??**. Nevertheless, the σ-finite measure Q and the stochastic differential equation (**??**) induce a σ-finite measure $\mathbb{Q}$ on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$, which is invariant in the sense that at all times $t \geqslant 0$ its marginal distributions are equal to Q. For this path measure $\mathbb{Q}$, conditional expectations with respect to the canonical forward and backwards filtrations are well-defined, see [**?**]. As a consequence, the martingale assertion of Theorem **??** makes sense also in this σ-finite setting; for the details we refer to [**?**, Chapter 1]. Since Jensen's inequality is also valid for conditional expectations with respect to the σ-finite path measure $\mathbb{Q}$ with marginals Q, Corollary **??** remains true under the Assumptions **??** without the additional requirement $\mathrm{Q}(\mathbb{R}^n) = 1$.

Now, as we have shown the inequality (**??**), we pass to the limit as $n \to \infty$. Since the non-decreasing sequence of stopping times $(\tau_n)_{n\geqslant 0}$ from (**??**) converges $\mathbb{P}$-almost surely to T, we deduce from (**??**) and the monotone convergence theorem that

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{T} \frac{1}{2} \frac{\left|\nabla \ell(T-u, X(T-u))\right|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u \right] \leqslant H\big(P(0)\,|\,\mathrm{Q}\big) - H\big(P(T)\,|\,\mathrm{Q}\big) < \infty, \tag{4.26}$$

because the initial relative entropy $H(P(0)\,|\,\mathrm{Q})$ is finite by assumption and $H(P(T)\,|\,\mathrm{Q})$ cannot take the value $-\infty$; see Lemma **??** and Appendix **??**. From (**??**) we finally obtain that the stochastic integral in (**??**) defines an $L^2(\mathbb{P})$-bounded martingale for $0 \leqslant s \leqslant T$.

Summing up, we conclude that the process $(M(T-s))_{0\leqslant s\leqslant T}$ is a $L^2(\mathbb{P})$-bounded martingale of the backwards filtration, satisfying (**??**). This completes the proof of Theorem **??**. □

4.4 The proof of Theorem ??

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

$$\ell^\beta(T-s, X(T-s)) = \frac{p^\beta(T-s, X(T-s))}{q(X(T-s))}\,, \qquad 0 \leqslant s \leqslant T - t_0, \tag{4.27}$$

and its logarithm.

By analogy with Proposition **??**, the following result is well known to hold under suitable regularity conditions, such as Assumptions **??**. Recall that $(W^\beta(t))_{t\geqslant t_0}$ denotes the $\mathbb{P}^\beta$-Brownian motion (in the forward direction of time) defined in (**??**).

Proposition 4.7. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *The process*

$$\overline{W}^{\mathbb{P}^\beta}(T-s) := W^\beta(T-s) - W^\beta(T) - \int_0^s \nabla \log p^\beta(T-u, X(T-u))\,\mathrm{d}u \tag{4.28}$$

for $0 \leqslant s \leqslant T-t_0$, *is a Brownian motion of the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$ *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*

$$\mathrm{d}X(T-s) = \Big(\nabla \log p^\beta(T-s, X(T-s)) + \big(\nabla\Psi + \beta\big)\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s) \tag{4.29}$$

$$= \Big(\nabla \log \ell^\beta(T-s, X(T-s)) - \big(\nabla\Psi - \beta\big)\big(X(T-s)\big)\Big)\,\mathrm{d}s + \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s), \tag{4.30}$$

for $0 \leqslant s \leqslant T-t_0$, *with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$.

We note next, how the Brownian motions $\big(\overline{W}^{\mathbb{P}^\beta}(T-s)\big)_{0\leqslant s\leqslant T-t_0}$ and $\big(\overline{W}^{\mathbb{P}}(T-s)\big)_{0\leqslant s\leqslant T-t_0}$, in reverse-time, are related.

Lemma 4.8. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *For* $0 \leqslant s \leqslant T-t_0$, *we have*

$$\mathrm{d}\big(\overline{W}^{\mathbb{P}} - \overline{W}^{\mathbb{P}^\beta}\big)(T-s) = \left(\beta\big(X(T-s)\big) + \nabla \log\left(\frac{p^\beta(T-s, X(T-s))}{p(T-s, X(T-s))}\right)\right)\mathrm{d}s \tag{4.31}$$

$$= \left(\beta\big(X(T-s)\big) + \nabla \log\left(\frac{\ell^\beta(T-s, X(T-s))}{\ell(T-s, X(T-s))}\right)\right)\mathrm{d}s. \tag{4.32}$$

Proof. It suffices to compare the equation (**??**) with (**??**). □

Remark 4.9. Later we shall apply Lemma **??** to the situation when s is close to $T-t_0$. In this case the logarithmic gradients in (**??**) and (**??**) will become small in view of $p^\beta(t_0, \cdot\,) = p(t_0, \cdot\,)$, so that these logarithmic gradients will disappear in the limit $s \uparrow T-t_0$; see also Lemma **??** below. By contrast, the term $\beta(X(T-s))$ will not disappear in the limit $s \uparrow T-t_0$. Rather, it will tend to the random variable $\beta(X(t_0))$, which plays an important role in distinguishing between (**??**) and (**??**) in Proposition **??**. ◇

Next, by analogy with Subsection **??**, for $t > t_0$, we write the perturbed Fokker-Planck equation (**??**) as

$$\partial_t p^\beta(t,x) = \tfrac{1}{2}\Delta p^\beta(t,x) + \big\langle \nabla p^\beta(t,x)\,, \nabla\Psi(x) + \beta(x)\big\rangle_{\mathbb{R}^n} + p^\beta(t,x)\,\big(\Delta\Psi(x) + \operatorname{div}\beta(x)\big). \tag{4.33}$$

Using the relation

$$p^\beta(t,x) = \ell^\beta(t,x)\,q(x) = \ell^\beta(t,x)\,\mathrm{e}^{-2\Psi(x)}, \qquad t \geqslant t_0, \tag{4.34}$$

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

$$\begin{aligned}\partial_t \ell^\beta(t,x) = \tfrac{1}{2}\Delta\ell^\beta(t,x) + \big\langle \nabla\ell^\beta(t,x)\,,\, \beta(x) - \nabla\Psi(x)\big\rangle_{\mathbb{R}^n} \\ + \ell^\beta(t,x)\Big(\operatorname{div}\beta(x) - \big\langle \beta(x)\,,\, 2\,\nabla\Psi(x)\big\rangle_{\mathbb{R}^n}\Big), \qquad t > t_0;\end{aligned} \tag{4.35}$$

this is the analogue of the backwards Kolmogorov equation (**??**) in this "perturbed" context, and reduces to (**??**) when $\beta \equiv 0$.

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

Lemma 4.10. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *The time-reversed perturbed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\begin{aligned}\frac{\mathrm{d}\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))} &= \Big(\big\langle \beta\,,\, 2\,\nabla\Psi\big\rangle_{\mathbb{R}^n} - \operatorname{div}\beta\Big)\big(X(T-s)\big)\,\mathrm{d}s \\ &\quad + \frac{\big|\nabla\ell^\beta(T-s,X(T-s))\big|^2}{\ell^\beta(T-s,X(T-s))^2}\,\mathrm{d}s \,+\, \left\langle \frac{\nabla\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s)\right\rangle_{\mathbb{R}^n}\end{aligned} \tag{4.36}$$

and

$$\begin{aligned}\mathrm{d}\log\ell^\beta(T-s,X(T-s)) &= \Big(\big\langle \beta\,,\, 2\,\nabla\Psi\big\rangle_{\mathbb{R}^n} - \operatorname{div}\beta\Big)\big(X(T-s)\big)\,\mathrm{d}s \\ &\quad + \frac{1}{2}\frac{\big|\nabla\ell^\beta(T-s,X(T-s))\big|^2}{\ell^\beta(T-s,X(T-s))^2}\,\mathrm{d}s \,+\, \left\langle \frac{\nabla\ell^\beta(T-s,X(T-s))}{\ell^\beta(T-s,X(T-s))}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-s)\right\rangle_{\mathbb{R}^n},\end{aligned} \tag{4.37}$$

respectively, for $0 \leqslant s \leqslant T - t_0$, *with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$.

Proof. The equations (**??**), (**??**) follow from Itô's formula together with (**??**), (**??**). □

We have assembled now all the ingredients for the proof of Theorem **??**.

Proof of Theorem ??. Formally, the stochastic differential in (**??**) of the time-reversed perturbed likelihood ratio process (**??**) amounts to the conclusions (**??**) – (**??**) of Theorem **??**. But we still have to substantiate the claim, that the stochastic process $(M^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ defined in (**??**) with representation (**??**), indeed yields a $\mathbb{P}^\beta$-martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$, which is bounded in $L^2(\mathbb{P}^\beta)$.

Step 1. By analogy with Step 2 in the proof of Theorem **??**, we show first, that the stochastic process $(M^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ of (**??**), with integral representation (**??**), is a continuous local $\mathbb{P}^\beta$-martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$.

By condition **??** of Assumptions **??** the function $(t_0,\infty) \ni t \mapsto \nabla\log\ell^\beta(t,x)$ is continuous for every $x \in \mathbb{R}^n$. Together with the continuity of the paths of the canonical coordinate process $(X(t))_{t\geqslant 0}$, this implies that

$$\int_0^{T-t_0-\varepsilon} \frac{\big|\nabla\ell^\beta(T-u,X(T-u))\big|^2}{\ell^\beta(T-u,X(T-u))^2}\,\mathrm{d}u < \infty, \qquad \mathbb{P}^\beta\text{-almost surely,} \tag{4.38}$$

for every $0 < \varepsilon \leqslant T - t_0$. On account of (**??**), the sequence of stopping times (with respect to the backwards filtration)

$$\tau_n^\beta := \inf\left\{ t \geqslant 0\colon \int_0^t \frac{\big|\nabla\ell^\beta(T-u,X(T-u))\big|^2}{\ell^\beta(T-u,X(T-u))^2}\,\mathrm{d}u \,\geqslant n \right\} \wedge (T-t_0), \qquad n \in \mathbb{N}_0 \tag{4.39}$$

is non-decreasing and converges $\mathbb{P}^\beta$-almost surely to $T-t_0$. Hence, according to (**??**) and the definition in (**??**), the stopped process $\big((M^\beta)^{\tau_n^\beta}(T-s)\big)_{0\leqslant s\leqslant T-t_0}$ is a uniformly integrable $\mathbb{P}^\beta$-martingale of the backwards filtration, for every $n \in \mathbb{N}_0$.

Step 2. We show that, in fact, the process $(M^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ is a true $\mathbb{P}^\beta$-martingale.

Taking expectations with respect to the probability measure $\mathbb{P}^\beta$ in (**??**) at time $s = \tau_n^\beta$, and using that by Step 1 the stopped process $\big((M^\beta)^{\tau_n^\beta}(T-s)\big)_{0\leqslant s\leqslant T-t_0}$ is a true $\mathbb{P}^\beta$-martingale with respect to

the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$, we obtain

$$H\big(P^\beta(T-\tau_n^\beta)\,|\,\mathrm{Q}\big) - H\big(P^\beta(T)\,|\,\mathrm{Q}\big) \;= \mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{\tau_n^\beta} \frac{1}{2}\frac{\big|\nabla\ell^\beta\big(T-u,X(T-u)\big)\big|^2}{\ell^\beta\big(T-u,X(T-u)\big)^2}\,\mathrm{d}u\right] \tag{4.40}$$

$$+\;\mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{\tau_n^\beta}\Big(\big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n} - \operatorname{div}\beta\Big)\big(X(T-u)\big)\,\mathrm{d}u\right], \tag{4.41}$$

for every $n\in\mathbb{N}_0$. According to condition **??** of Assumptions **??**, the perturbation $\beta\colon\mathbb{R}^n\to\mathbb{R}^n$ is the gradient of a potential $B\colon\mathbb{R}^n\to\mathbb{R}$ of class $\mathcal{C}_c^\infty(\mathbb{R}^n;\mathbb{R})$. Therefore, for every $n\in\mathbb{N}_0$, the absolute value of the expectation in (**??**) can be bounded by

$$C_1 := \mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{T-t_0}\Big|\big\langle\beta\,,\,2\,\nabla\Psi\big\rangle_{\mathbb{R}^n} - \operatorname{div}\beta\Big|\big(X(T-u)\big)\,\mathrm{d}u\right] < \infty. \tag{4.42}$$

We denote by Q^β the σ-finite measure on the Borel sets of $\mathbb{R}^n$, whose density with respect to Lebesgue measure is

$$q^\beta(x) := \mathrm{e}^{-2(\Psi+B)(x)}, \qquad x\in\mathbb{R}^n. \tag{4.43}$$

Now, by analogy with (**??**) and the discussion of the σ-finite case in Step 3 from the proof of Theorem **??**, we have

$$H\big(P^\beta(t)\,|\,\mathrm{Q}^\beta\big) \leqslant H\big(P^\beta(t_0)\,|\,\mathrm{Q}^\beta\big), \qquad t_0\leqslant t\leqslant T. \tag{4.44}$$

Note that for this decrease of the relative entropy function $t\mapsto H(P^\beta(t)\,|\,\mathrm{Q}^\beta)$ to be valid, it is necessary to "take into account the perturbation for the flow as well as for the reference measure". However, in view of the left-hand side of (**??**), we are interested in the behavior of the function $t\mapsto H(P^\beta(t)\,|\,\mathrm{Q})$. In order to compare these two, we define the constant

$$C_2 := 2\cdot\max_{x\in\mathbb{R}^n}|B(x)| < \infty. \tag{4.45}$$

It is then straightforward to derive the estimate

$$H\big(P^\beta(t)\,|\,\mathrm{Q}\big) - C_2 \;\leqslant\; H\big(P^\beta(t)\,|\,\mathrm{Q}^\beta\big) \;\leqslant\; H\big(P^\beta(t)\,|\,\mathrm{Q}\big) + C_2, \qquad t_0\leqslant t\leqslant T. \tag{4.46}$$

Combining (**??**), (**??**) and using that $P^\beta(t_0)=P(t_0)$, we get

$$H\big(P^\beta(t)\,|\,\mathrm{Q}\big) \;\leqslant\; H\big(P(t_0)\,|\,\mathrm{Q}\big) + 2C_2, \qquad t_0\leqslant t\leqslant T. \tag{4.47}$$

Consequently, the left-hand side of (**??**) can be dominated by

$$C_3 := H\big(P(t_0)\,|\,\mathrm{Q}\big) + 2C_2 - H\big(P^\beta(T)\,|\,\mathrm{Q}\big) < \infty. \tag{4.48}$$

This constant cannot be equal to $+\infty$ because $H(P(t_0)\,|\,\mathrm{Q})$ is finite by the same argument as in the proof of Theorem **??**, and the relative entropy $H(P^\beta(T)\,|\,\mathrm{Q})$ cannot take the value $-\infty$; see Lemma **??** and Appendix **??**. Altogether, we obtain from (**??**), (**??**) and the above considerations that

$$\mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{\tau_n^\beta}\frac{1}{2}\frac{\big|\nabla\ell^\beta\big(T-u,X(T-u)\big)\big|^2}{\ell^\beta\big(T-u,X(T-u)\big)^2}\,\mathrm{d}u\right] \leqslant C_1 + C_3, \tag{4.49}$$

for every $n\in\mathbb{N}_0$. Since the non-decreasing sequence of stopping times $(\tau_n^\beta)_{n\geqslant 0}$ from (**??**) converges $\mathbb{P}^\beta$-almost surely to $T-t_0$ as $n\to\infty$, we deduce from (**??**) and the monotone convergence theorem that

$$\mathbb{E}_{\mathbb{P}^\beta}\big[F^\beta(t_0)\big] = \mathbb{E}_{\mathbb{P}^\beta}\left[\int_0^{T-t_0}\frac{1}{2}\frac{\big|\nabla\ell^\beta\big(T-u,X(T-u)\big)\big|^2}{\ell^\beta\big(T-u,X(T-u)\big)^2}\,\mathrm{d}u\right] \leqslant C_1 + C_3 < \infty. \tag{4.50}$$

From (**??**) we finally obtain that the stochastic integral in (**??**) defines an $L^2(\mathbb{P}^\beta)$-bounded martingale for $0 \leqslant s \leqslant T - t_0$.

Summing up, we conclude that the process $(M^\beta(T-s))_{0\leqslant s\leqslant T-t_0}$ is a $L^2(\mathbb{P}^\beta)$-bounded martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$, admitting the representation (**??**). This completes the proof of Theorem **??**. □

4.5 Some useful lemmas

In this subsection we collect some useful results needed in order to justify the claims (**??**), (**??**) made in the course of the proof of Corollary **??**, and to complete the proof of Proposition **??** in Subsection **??**.

First, let us recall the probability density function $(t,x) \mapsto p(t,x)$ from (**??**), its perturbed version $(t,x) \mapsto p^\beta(t,x)$ from (**??**), and the respective likelihood ratios

$$\ell(t,x) = \frac{p(t,x)}{q(x)}, \qquad \ell^\beta(t,x) = \frac{p^\beta(t,x)}{q(x)} \tag{4.51}$$

from (**??**), (**??**), respectively. We introduce also the "perturbed-to-unperturbed" ratio

$$Y^\beta(t,x) := \frac{\ell^\beta(t,x)}{\ell(t,x)} = \frac{p^\beta(t,x)}{p(t,x)}, \qquad (t,x) \in [t_0,\infty) \times \mathbb{R}^n. \tag{4.52}$$

We recall the backward Kolmogorov-type equations (**??**), (**??**). These lead to the partial differential equation

$$\begin{aligned} \partial_t Y^\beta(t,x) = \tfrac{1}{2}\Delta Y^\beta(t,x) + \big\langle \nabla Y^\beta(t,x)\,,\, \beta(x) + \nabla \log p(t,x) + \nabla\Psi(x)\big\rangle_{\mathbb{R}^n} \\ + Y^\beta(t,x)\Big(\operatorname{div}\beta(x) + \big\langle \beta(x)\,,\, \nabla \log p(t,x)\big\rangle_{\mathbb{R}^n}\Big), \qquad t > t_0, \end{aligned} \tag{4.53}$$

with $Y^\beta(t_0,\,\cdot\,) = 1$, for the ratio in (**??**). In conjunction with (**??**), this equation leads to the following backward dynamics.

Lemma 4.11. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *The time-reversed ratio process* $\big(Y^\beta(T-s, X(T-s))\big)_{0\leqslant s\leqslant T-t_0}$ *and its logarithm satisfy the stochastic differential equations*

$$\begin{aligned} \frac{\mathrm{d}Y^\beta\big(T-s,X(T-s)\big)}{Y^\beta\big(T-s,X(T-s)\big)} = \left\langle \frac{\nabla Y^\beta\big(T-s,X(T-s)\big)}{Y^\beta\big(T-s,X(T-s)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta\big(X(T-s)\big)\,\mathrm{d}s\right\rangle_{\mathbb{R}^n} \\ - \Big(\operatorname{div}\beta\big(X(T-s)\big) + \Big\langle \beta\big(X(T-s)\big)\,,\nabla \log p\big(T-s,X(T-s)\big)\Big\rangle_{\mathbb{R}^n}\Big)\,\mathrm{d}s \end{aligned} \tag{4.54}$$

and

$$\begin{aligned} \mathrm{d}\log Y^\beta\big(T-s,X(T-s)\big) = \left\langle \frac{\nabla Y^\beta\big(T-s,X(T-s)\big)}{Y^\beta\big(T-s,X(T-s)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta\big(X(T-s)\big)\,\mathrm{d}s\right\rangle_{\mathbb{R}^n} \\ - \Big(\operatorname{div}\beta\big(X(T-s)\big) + \Big\langle \beta\big(X(T-s)\big)\,,\nabla \log p\big(T-s,X(T-s)\big)\Big\rangle_{\mathbb{R}^n}\Big)\,\mathrm{d}s \\ - \frac{1}{2}\frac{\big|\nabla Y^\beta\big(T-s,X(T-s)\big)\big|^2}{Y^\beta\big(T-s,X(T-s)\big)^2}\,\mathrm{d}s, \end{aligned} \tag{4.55}$$

respectively, for $0 \leqslant s \leqslant T - t_0$, *with respect to the backwards filtration* $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T-t_0}$.

Remark 4.12. We can obtain the dynamics of (**??**), thus also of (**??**), directly from (**??**), (**??**), just by subtracting and using Lemma **??**. ◇

We also need a preliminary control on $Y^\beta(\,\cdot\,,\,\cdot\,)$, which is the subject of the following Lemma **??**. This will be refined in Lemma **??** below.

Lemma 4.13. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *There is a real constant* $C > 1$ *such that*

$$\frac{1}{C} \leqslant Y^\beta(t,x) \leqslant C\,, \qquad (t,x) \in [t_0,T] \times \mathbb{R}^n. \tag{4.56}$$

Proof. In the forward direction of time, the canonical coordinate process $(X(t))_{t_0 \leqslant t \leqslant T}$ on the path space $\Omega = \mathcal{C}([t_0,T];\mathbb{R}^n)$ satisfies the stochastic equations (**??**) and (**??**) 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 (**??**) can be represented as

$$W(t) - W(t_0) = W^\beta(t) - W^\beta(t_0) - \int_{t_0}^t \beta\big(X(u)\big)\,\mathrm{d}u, \qquad t_0 \leqslant t \leqslant T, \tag{4.57}$$

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

$$Z(t) := \frac{\mathrm{d}\mathbb{P}^\beta}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}(t)} = \exp\bigg(- \int_{t_0}^t \Big\langle \beta\big(X(u)\big)\,,\,\mathrm{d}W(u)\Big\rangle_{\mathbb{R}^n} - \tfrac{1}{2}\int_{t_0}^t \big|\beta\big(X(u)\big)\big|^2\,\mathrm{d}u\bigg). \tag{4.58}$$

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 (**??**) with respect to the probability measure $\mathbb{P}$, where we condition on $X(t) = x$; to wit,

$$Y^\beta(t,x) = \mathbb{E}_{\mathbb{P}}\big[Z(t)\,|\,X(t) = x\big]\,, \qquad (t,x) \in [t_0,T] \times \mathbb{R}^n. \tag{4.59}$$

Therefore, in order to obtain the estimate (**??**), 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 (**??**) 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 (**??**), 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 (**??**), we can express the stochastic integral appearing in (**??**) as

$$\int_{t_0}^t \Big\langle \beta\big(X(u)\big)\,,\,\mathrm{d}W(u)\Big\rangle_{\mathbb{R}^n} = B\big(X(t)\big) - B\big(X(t_0)\big) + \int_{t_0}^t \Big(\big\langle \beta\,,\,\nabla\Psi\big\rangle_{\mathbb{R}^n} - \tfrac{1}{2}\,\mathrm{div}\,\beta\Big)\big(X(u)\big)\,\mathrm{d}u \tag{4.60}$$

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

The following Lemma **??** provides the crucial estimates (**??**) and (**??**), needed in the proof of Corollary **??** from Theorem **??**, and of Proposition **??**.

Lemma 4.14. *Under the* Assumptions **??**, *we let* $t_0 \geqslant 0$ *and* $T > t_0$. *There is a constant* $C > 0$ *such that*

$$\big|Y^\beta(T-s,x) - 1\big| \leqslant C\,(T - t_0 - s), \tag{4.61}$$

as well as

$$\mathbb{E}_{\mathbb{P}}\bigg[\int_s^{T-t_0} \Big|\nabla \log Y^\beta\big(T-u,X(T-u)\big)\Big|^2\,\mathrm{d}u \;\bigg|\; X(T-s) = x\bigg] \leqslant C\,(T - t_0 - s)^2, \tag{4.62}$$

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*

$$\lim_{s \uparrow T - t_0} \frac{\log Y^\beta(T-s,x)}{T - t_0 - s} = \operatorname{div} \beta(x) + \Big\langle \beta(x)\,,\, \nabla \log p(t_0,x) \Big\rangle_{\mathbb{R}^n}, \tag{4.63}$$

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

Remark 4.15. The pointwise limiting assertion (**??**) is the deterministic analogue of the trajectorial relation (**??**) from Proposition **??**. In Subsection **??** below we will prove that the limiting assertion (**??**) holds in the norm of L^1 under both $\mathbb{P}$ and $\mathbb{P}^\beta$, and is valid for all $t_0 > 0$. ⋄

Proof. As $\log Y^\beta = \log \ell^\beta - \log \ell$, we obtain from (**??**), (**??**) and (**??**) that the martingale part of the process in (**??**) is bounded in $L^2(\mathbb{P})$, i.e.,

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{T-t_0} \frac{\big|\nabla Y^\beta(T-u, X(T-u))\big|^2}{Y^\beta(T-u, X(T-u))^2}\, \mathrm{d}u \right] < \infty. \tag{4.64}$$

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

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^{T-t_0} \Big|\nabla Y^\beta(T-u, X(T-u))\Big|^2\, \mathrm{d}u \right] < \infty. \tag{4.65}$$

According to (**??**), the time-reversed ratio process $\big(Y^\beta(T-s, X(T-s))\big)_{0 \leqslant s \leqslant T-t_0}$ satisfies the stochastic differential equation

$$\begin{aligned} \mathrm{d}Y^\beta(T-s, X(T-s)) &= \Big\langle \nabla Y^\beta(T-s, X(T-s))\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-s) - \beta\big(X(T-s)\big)\,\mathrm{d}s \Big\rangle_{\mathbb{R}^n} \\ &\quad - Y^\beta(T-s, X(T-s))\bigg(\operatorname{div}\beta\big(X(t-s)\big) + \Big\langle \beta\big(X(T-s)\big)\,,\, \nabla \log p\big(T-s, X(T-s)\big) \Big\rangle_{\mathbb{R}^n} \bigg)\, \mathrm{d}s \end{aligned} \tag{4.66}$$

for $0 \leqslant s \leqslant T - t_0$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leqslant s \leqslant T-t_0}$. In view of (**??**), the martingale part in (**??**) is bounded in $L^2(\mathbb{P})$. As regards the drift terms 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 terms are bounded, i.e., the constant

$$C_1 := \sup_{\substack{t_0 \leqslant t \leqslant T \\ y \in \mathbb{R}^n}} \left| - Y^\beta(t,y) \bigg(\operatorname{div}\beta(y) + \Big\langle \beta(y)\,,\, \nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)} \Big\rangle_{\mathbb{R}^n} \bigg) \right| \tag{4.67}$$

is finite, and the processes

$$Y^\beta\big(T-s, X(T-s)\big) + C_1\, s \qquad \text{and} \qquad Y^\beta\big(T-s, X(T-s)\big) - C_1\, s \tag{4.68}$$

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

$$\Big| Y^\beta(T-s, x) - \mathbb{E}_{\mathbb{P}}\Big[Y^\beta\big(t_0, X(t_0)\big) \;\Big|\; X(T-s) = x \Big] \Big| \leqslant C_1\,(T - t_0 - s) \tag{4.69}$$

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

$$\big| Y^\beta(T-s, x) - 1 \big| \leqslant C_1\,(T - t_0 - s). \tag{4.70}$$

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

$$\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\left[\int_s^{T-t_0} \left|\nabla \log Y^\beta\big(T-u, X(T-u)\big)\right|^2 \mathrm{d}u \;\middle|\; X(T-s) = x\right] \tag{4.71}$$

is equal to

$$\log Y^\beta(T-s,x) - Y^\beta(T-s,x) + 1 + \mathbb{E}_{\mathbb{P}}\left[\int_s^{T-t_0} G\big(T-u, X(T-u)\big)\,\mathrm{d}u \;\middle|\; X(T-s) = x\right], \tag{4.72}$$

where we have set

$$G(t,y) := \big(Y^\beta(t,y) - 1\big)\left(\operatorname{div}\beta(y) + \left\langle \beta(y)\,, \nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)} \right\rangle_{\mathbb{R}^n}\right) \tag{4.73}$$

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

$$C_2 := \sup_{\substack{t_0 \leqslant t \leqslant T \\ y \in \mathbb{R}^n}} \left|\operatorname{div}\beta(y) + \left\langle \beta(y)\,, \nabla \log p(t,y) + \frac{\nabla Y^\beta(t,y)}{Y^\beta(t,y)} \right\rangle_{\mathbb{R}^n}\right| \tag{4.74}$$

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

$$\log Y^\beta(T-s,x) - Y^\beta(T-s,x) + 1 \leqslant 0. \tag{4.75}$$

This implies that the expression of (**??**) is bounded by $C_1\,C_2\,(T-t_0-s)^2$, which establishes the second estimate (**??**). We also note that the elementary inequality (**??**) in conjunction with the estimate (**??**) shows that

$$\log Y^\beta(T-s,x) \leqslant C_1\,(T-t_0-s) \tag{4.76}$$

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 (**??**) is uniformly bounded on $[0, T-t_0] \times \mathbb{R}^n$.

Regarding the limiting assertion (**??**), 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 (**??**). On account of (**??**), the stochastic integral with respect to the $\mathbb{P}$-Brownian motion $(\overline{W}^{\mathbb{P}}(T-s))_{0\leqslant s\leqslant T}$ in (**??**) vanishes. Dividing by $T-t_0-s$ and passing to the limit as $s \uparrow T-t_0$, we can use the estimate (**??**) to observe that the expression in the third line of (**??**) 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 (**??**) can be bounded by

$$\frac{1}{T-t_0-s}\int_s^{T-t_0} \left|\nabla \log Y^\beta\big(T-u, X(T-u)\big)\right| \cdot \left|\beta\big(X(T-u)\big)\right| \mathrm{d}u. \tag{4.77}$$

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

$$\lim_{s \uparrow T-t_0} \mathbb{E}_{\mathbb{P}}\left[\frac{1}{T-t_0-s}\int_s^{T-t_0} \left|\nabla \log Y^\beta\big(T-u, X(T-u)\big)\right| \cdot \left|\beta\big(X(T-u)\big)\right| \mathrm{d}u \;\middle|\; X(T-s) = x\right] = 0. \tag{4.78}$$

Finally, using continuity and uniform boundedness once again, the conditional expectations of the normalized integrals over the second line of (**??**) converge to the right-hand side of (**??**), as claimed. □

Remark 4.16 (Stability of the entropy limits under perturbations). The above Lemma **??** justifies the estimates (**??**) and (**??**), which we have used in the proof of Corollary **??**. They were the crucial ingredients in the effort to show that the exceptional set for the limiting assertion (**??**) does not change when passing from the unperturbed to the perturbed equation (**??**). It is now time to come back to this technical issue.

As a general observation, we stress that no worries about limits of difference quotients arise as long as we remain in the realm of an *integral formulation* of our results, as opposed to passing to a *differential formulation*. It is precisely the spirit of our basic trajectorial Theorems **??** and **??**, that they are naturally formulated in *integral terms*.

We also note that the problem of exceptional points does not arise if we impose regularity assumptions strong enough, so that the limiting assertions (**??**) and (**??**) are valid *for all* $t_0 > 0$, or even *for all* $t_0 \geqslant 0$, instead of *for Lebesgue-almost every* $t_0 \geqslant 0$. For example, this follows if we impose, in addition to Assumptions **??**, the *a priori* assumption that the relative Fisher information function $t \mapsto I(P(t) \,|\, \mathrm{Q})$ is continuous on $(0, \infty)$, or continuous on $[0, \infty)$, respectively.

Having made these general observations, let us now be more technical and have a precise look at the exceptional sets in the framework of the regularity codified by Assumptions **??**. In the proof of Corollary **??** we have deduced from the Lebesgue differentiation theorem that the generalized de Bruijn identity (**??**), (**??**) is valid outside a set of Lebesgue measure zero. In particular, let us recall from Remark **??** that N denotes the set of exceptional points $t_0 \geqslant 0$ for which the right-sided limiting assertion (**??**), i.e.,

$$\lim_{t \downarrow t_0} \frac{H(P(t) \,|\, \mathrm{Q}) - H(P(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2} \, \|a\|^2_{L^2(\mathbb{P})}, \tag{4.79}$$

fails. We have shown in the equations (**??**) and (**??**) from the proof of Corollary **??** that the limiting assertion (**??**), (**??**), i.e.,

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) \,|\, \mathrm{Q}) - H(P^\beta(t_0) \,|\, \mathrm{Q})}{t - t_0} = -\tfrac{1}{2} \, \langle a, a + 2b \rangle_{L^2(\mathbb{P})}, \tag{4.80}$$

is valid if and only if $t_0 \in \mathbb{R}_+ \setminus N$. In other words, the limits in (**??**) and (**??**) have the same exceptional set N of zero Lebesgue measure. Put another way, the entropy limit (**??**) is stable under perturbations, in the sense that the corresponding perturbed entropy limit (**??**) continues to be valid for the same points $t_0 \in \mathbb{R}_+ \setminus N$.

Furthermore, in the proofs of Propositions **??** and **??** we have seen that the limiting assertions therein are valid, respectively, for those points $t_0 \geqslant 0$ for which the generalized de Bruijn identity (**??**) does hold, and for $t_0 \in \mathbb{R}_+ \setminus N$. ◇

Remark 4.17 (Stability of the Wasserstein limits under perturbations). Let us now pass to the limits of the difference quotients pertaining to the Wasserstein distance. We fix some $t_0 \in \mathbb{R}_+ \setminus N$ so that the limiting assertion (**??**), and as a consequence also (**??**), are valid. Then the unperturbed Wasserstein limit

$$\lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0} = \tfrac{1}{2} \, \|a\|_{L^2(\mathbb{P})} \tag{4.81}$$

of (**??**) is valid as well. This is remarkable, because a priori there is no significant relation between relative entropy and Wasserstein distance, except for the fact that in the limit the relative Fisher information appears on the right-hand sides of both (**??**) and (**??**). Even more, the unperturbed Wasserstein limit (**??**), (**??**), i.e.,

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \tfrac{1}{2} \, \|a + 2b\|_{L^2(\mathbb{P})}, \tag{4.82}$$

also holds for this point $t_0 \in \mathbb{R}_+ \setminus N$. In other words, the Wasserstein limits are stable under perturbations in the same manner as the entropy limits are.

Summing up, not only do the limiting assertions (**??**) and (**??**) hold for every $t_0 \in \mathbb{R}_+ \setminus N$, but so do also the limiting assertions (**??**) and (**??**) pertaining to the Wasserstein distance. We will prove these results in Theorems **??** and **??** of Section **??**. ⋄

4.6 Completing the proof of Proposition ??

On account of the preparations in Subsection **??** above, we are now able to complete the proof of Proposition **??**, by establishing the remaining limiting assertions (**??**) and (**??**) therein.

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

$$\lim_{s \uparrow T - t_0} \frac{\log Y^\beta\big(T - s, X(T-s)\big)}{T - t_0 - s}. \tag{4.83}$$

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

$$\frac{1}{T - t_0 - s} \int_s^{T-t_0} \Big(\operatorname{div} \beta\big(X(T-u)\big) + \Big\langle \beta\big(X(T-u)\big)\,, \nabla \log p\big(T-u, X(T-u)\big) \Big\rangle_{\mathbb{R}^n} \Big)\, \mathrm{d}u \tag{4.84}$$

is uniformly bounded on $[0, T-t_0] \times \operatorname{supp} \beta$. Using continuity and uniform boundedness, we conclude that (**??**) converges $\mathbb{P}$-almost surely as well as in $L^1(\mathbb{P})$ to the right-hand side of (**??**), 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

$$\frac{1}{T - t_0 - s} \int_s^{T-t_0} \left\langle \frac{\nabla Y^\beta\big(T-u, X(T-u)\big)}{Y^\beta\big(T-u, X(T-u)\big)}\,,\, \beta\big(X(T-u)\big) \right\rangle_{\mathbb{R}^n} \mathrm{d}u \tag{4.85}$$

converges to zero $\mathbb{P}$-almost surely 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

$$\frac{1}{T - t_0 - s} \int_s^{T-t_0} \frac{1}{2} \frac{\big|\nabla Y^\beta\big(T-u, X(T-u)\big)\big|^2}{Y^\beta\big(T-u, X(T-u)\big)^2}\, \mathrm{d}u \tag{4.86}$$

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

$$\mathbb{E}_{\mathbb{P}} \left[\frac{1}{T - t_0 - s} \int_s^{T-t_0} \frac{\big|\nabla Y^\beta\big(T-u, X(T-u)\big)\big|^2}{Y^\beta\big(T-u, X(T-u)\big)^2}\, \mathrm{d}u \right] \leqslant C\,(T - t_0 - s) \tag{4.87}$$

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

$$-\frac{1}{T - t_0 - s} \int_s^{T-t_0} \left\langle \frac{\nabla Y^\beta\big(T-u, X(T-u)\big)}{Y^\beta\big(T-u, X(T-u)\big)}\,,\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-u) \right\rangle_{\mathbb{R}^n}. \tag{4.88}$$

The expression (**??**) converges to zero $\mathbb{P}$-almost surely, and according to (**??**) we have

$$\mathbb{E}_{\mathbb{P}} \left[\frac{1}{(T - t_0 - s)^2} \int_s^{T-t_0} \frac{\big|\nabla Y^\beta\big(T-u, X(T-u)\big)\big|^2}{Y^\beta\big(T-u, X(T-u)\big)^2}\, \mathrm{d}u \right] \leqslant C. \tag{4.89}$$

By means of the Itô isometry, we deduce that

$$\lim_{s\uparrow T-t_0} \mathbb{E}_{\mathbb{P}}\left[\left(\frac{1}{T-t_0-s}\int_s^{T-t_0}\left\langle \frac{\nabla Y^\beta(T-u,X(T-u))}{Y^\beta(T-u,X(T-u))}\,,\,\mathrm{d}\overline{W}^{\mathbb{P}}(T-u)\right\rangle_{\mathbb{R}^n}\right)^2\right] = 0. \tag{4.90}$$

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

Summing up, we have shown that the limiting assertion (**??**) holds in the norm of $L^1(\mathbb{P})$ and is valid for all $t_0 > 0$. As we have seen in Lemma **??**, 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 the limiting assertion (**??**). □

Proof of the assertion (**??**) *in Proposition* **??***:* This is proved in very much the same way, as assertions (**??**), (**??**). The only novelty here, is the use of (**??**) to pass to the $\mathbb{P}$-Brownian motion $(\overline{W}^{\mathbb{P}}(T-s))_{0\leqslant s\leqslant T-t_0}$ from the $\mathbb{P}^\beta$-Brownian motion $(\overline{W}^{\mathbb{P}^\beta}(T-s))_{0\leqslant s\leqslant T-t_0}$, and the reliance on (**??**) to ensure that the resulting stochastic integral is a (square-integrable) $\mathbb{P}$-martingale. We leave the details to the care of the diligent reader, or refer to [**?**, Section 3.2]. □

4.7 The dynamics in the forward direction of time

For the sake of completeness, we calculate now the stochastic differentials of the relative entropy process and its perturbed counterpart of (**??**) also in the forward direction of time, under the measures $\mathbb{P}$ and $\mathbb{P}^\beta$, respectively. It will turn out that we are able to derive Theorems **??** and **??** also by applying Itô's formula in the forward direction of time. But the relations between these theorems and the stochastic differentials will become less transparent than in reverse time. In fact, additional terms will show up in the forward direction of time, which on the contrary did not appear in the backward direction. One may still take expectations, but to obtain Theorems **??** and **??** one has to argue why the expectations of these additional terms vanish. This is straightforward in the unperturbed situation of Lemma **??**, but in the perturbed context of Lemma **??** one also has to rely on integration by parts.

We first compute the differentials of the likelihood ratio process

$$\ell(t,X(t)) = \frac{p(t,X(t))}{q(X(t))}, \qquad t \geqslant 0 \tag{4.91}$$

and its logarithm in the forward direction of time. We start by recalling the backwards Kolmogorov equation (**??**). With its help, we can compute the forward dynamics of the likelihood ratio process (**??**) in the following manner.

Lemma 4.18. *Under the* Assumptions **??**, *the likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\mathrm{d}\ell(t,X(t)) = \Delta\ell(t,X(t))\,\mathrm{d}t + \left\langle \nabla\ell(t,X(t))\,,\,\mathrm{d}W(t) - 2\,\nabla\Psi(X(t))\,\mathrm{d}t\right\rangle_{\mathbb{R}^n} \tag{4.92}$$

$$= 2\,\partial_t\ell(t,X(t))\,\mathrm{d}t + \left\langle \nabla\ell(t,X(t))\,,\,\mathrm{d}W(t)\right\rangle_{\mathbb{R}^n} \tag{4.93}$$

and

$$\mathrm{d}\log \ell(t,X(t)) = \left(\frac{\Delta \ell(t,X(t))}{\ell(t,X(t))} - \left\langle \frac{\nabla \ell(t,X(t))}{\ell(t,X(t))} \,,\, 2\,\nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} \right) \mathrm{d}t \tag{4.94}$$

$$- \frac{1}{2} \frac{\left|\nabla \ell(t,X(t))\right|^2}{\ell(t,X(t))^2}\,\mathrm{d}t \,+\, \left\langle \frac{\nabla \ell(t,X(t))}{\ell(t,X(t))} \,,\, \mathrm{d}W(t) \right\rangle_{\mathbb{R}^n} \tag{4.95}$$

$$= \left(2\,\partial_t \log \ell(t,X(t)) - \tfrac{1}{2} \left|\nabla \log \ell(t,X(t))\right|^2 \right) \mathrm{d}t + \left\langle \nabla \log \ell(t,X(t)) \,,\, \mathrm{d}W(t) \right\rangle_{\mathbb{R}^n}, \tag{4.96}$$

respectively, for $t \geqslant 0$, with respect to the forward filtration $(\mathcal{F}(t))_{t \geqslant 0}$.

Proof. Applying Itô's formula and using the equations (**??**), (**??**) shows (**??**), (**??**). One more application of Itô's formula leads to the stochastic equations (**??**) – (**??**). □

In order to deduce Theorem **??** from Lemma **??** — at least formally — we take expectations in (**??**) – (**??**) and use (**??**) to observe that

$$\mathbb{E}_{\mathbb{P}} \left[\frac{\Delta \ell(t,X(t))}{\ell(t,X(t))} - \left\langle \frac{\nabla \ell(t,X(t))}{\ell(t,X(t))} \,,\, 2\,\nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} \right] = \mathbb{E}_{\mathbb{P}} \left[\frac{2\,\partial_t \ell(t,X(t))}{\ell(t,X(t))} \right] = 0. \tag{4.97}$$

Next, we calculate the differentials of the perturbed likelihood ratio process

$$\ell^\beta(t,X(t)) = \frac{p^\beta(t,X(t))}{q(X(t))}\,, \qquad t \geqslant t_0 \tag{4.98}$$

and its logarithm, again in the forward direction. With the help of the "perturbed" backwards Kolmogorov equation (**??**), we obtain the forward dynamics of the perturbed likelihood ratio process (**??**), as follows.

Lemma 4.19. *Under the* Assumptions **??**, *let $t_0 \geqslant 0$. The perturbed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\frac{\mathrm{d}\ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} = \left(\frac{\Delta \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} - \left\langle \frac{\nabla \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} \,,\, 2\,\nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} \right) \mathrm{d}t \tag{4.99}$$

$$+ \Big(\operatorname{div} \beta - \langle \beta \,,\, 2\,\nabla \Psi \rangle_{\mathbb{R}^n} \Big)(X(t))\,\mathrm{d}t \,+\, \left\langle \frac{\nabla \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} \,,\, \mathrm{d}W^\beta(t) \right\rangle_{\mathbb{R}^n} \tag{4.100}$$

$$= \Big(2\,\partial_t \log \ell^\beta(t,X(t)) - \left\langle \nabla \log \ell^\beta(t,X(t)) \,,\, \beta(X(t)) \right\rangle_{\mathbb{R}^n} \Big)\,\mathrm{d}t \tag{4.101}$$

$$+ \Big(\langle \beta \,,\, 2\,\nabla \Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \Big)(X(t))\,\mathrm{d}t \,+\, \left\langle \nabla \log \ell^\beta(t,X(t)) \,,\, \mathrm{d}W^\beta(t) \right\rangle_{\mathbb{R}^n} \tag{4.102}$$

and

$$\mathrm{d}\log \ell^\beta(t,X(t)) = \left(\frac{\Delta \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} - \left\langle \frac{\nabla \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} \,,\, 2\,\nabla\Psi\big(X(t)\big) \right\rangle_{\mathbb{R}^n} \right) \mathrm{d}t \tag{4.103}$$

$$+ \left(\Big(\operatorname{div}\beta - \big\langle \beta \,,\, 2\,\nabla\Psi \big\rangle_{\mathbb{R}^n} \Big)\big(X(t)\big) - \frac{1}{2} \frac{\big|\nabla \ell^\beta(t,X(t))\big|^2}{\ell^\beta(t,X(t))^2} \right) \mathrm{d}t + \left\langle \frac{\nabla \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} \,,\, \mathrm{d}W^\beta(t) \right\rangle_{\mathbb{R}^n} \tag{4.104}$$

$$= \Big(2\,\partial_t \log \ell^\beta(t,X(t)) - \Big\langle \nabla \log \ell^\beta(t,X(t)) \,,\, \beta\big(X(t)\big) \Big\rangle_{\mathbb{R}^n} \Big)\, \mathrm{d}t \tag{4.105}$$

$$+ \left(\Big(\big\langle \beta \,,\, 2\,\nabla\Psi \big\rangle_{\mathbb{R}^n} - \operatorname{div}\beta \Big)\big(X(t)\big) - \tfrac{1}{2} \big|\nabla \log \ell^\beta(t,X(t))\big|^2 \right) \mathrm{d}t + \Big\langle \nabla \log \ell^\beta(t,X(t)) \,,\, \mathrm{d}W^\beta(t) \Big\rangle_{\mathbb{R}^n}, \tag{4.106}$$

respectively, for $t \geqslant t_0$, with respect to the forward filtration $(\mathcal{F}(t))_{t \geqslant t_0}$.

Proof. Using (**??**), (**??**) and Itô's formula, we obtain the stochastic equations (**??**) – (**??**). □

The perturbed situation is not as nice as the unperturbed one, since according to (**??**) we have

$$\Delta \ell^\beta(t,x) - \big\langle \nabla \ell^\beta(t,x) \,,\, 2\,\nabla\Psi(x) \big\rangle_{\mathbb{R}^n} \neq 2\,\partial_t \ell^\beta(t,x) \tag{4.107}$$

in general. However, integrating by parts shows that

$$\mathbb{E}_{\mathbb{P}} \left[\frac{\Delta \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} - \left\langle \frac{\nabla \ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} \,,\, 2\,\nabla\Psi\big(X(t)\big) \right\rangle_{\mathbb{R}^n} \right] = 0. \tag{4.108}$$

Hence taking expectations in (**??**), (**??**) allows to derive Theorem **??** from Lemma **??**, at least on a formal level. But as opposed to the backward direction of time, the identity (**??**) does not hold any more when we take expectations conditionally on $X(t)$.

5 The Wasserstein transport

For the convenience of the reader we review in Subsections **??** and **??** some well-known results on quadratic Wasserstein transport [**?**, **?**], in order to establish the limits (**??**) and (**??**) and complete the proofs of Corollaries **??** and **??**.

As we make precise statements regarding the points $t_0 \geqslant 0$ at which the the limiting assertions (**??**) and (**??**) are valid (recall Remark **??** at this point), we provide detailed proofs of the relevant Theorems **??**, **??** in Subsection **??**.

5.1 Basic notation and terminology

We recall below the definitions of the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$, and of the quadratic Wasserstein distance W_2. We follow the setting of [**?**], from where we borrow most of the notation and terminology used in this section. Thus, for unexplained notions and definitions, the reader may consult this beautiful book.

We denote by $\mathscr{P}(\mathbb{R}^n)$ the collection of probability measures on the Borel sets $\mathscr{B}(\mathbb{R}^n)$ of $\mathbb{R}^n$. The *quadratic Wasserstein space* $\mathscr{P}_2(\mathbb{R}^n)$ is the subset of $\mathscr{P}(\mathbb{R}^n)$ consisting of the probability measures on $\mathscr{B}(\mathbb{R}^n)$ with finite second moment, i.e.,

$$\mathscr{P}_2(\mathbb{R}^n) := \left\{ P \in \mathscr{P}(\mathbb{R}^n) \colon \ \int_{\mathbb{R}^n} |x|^2 \,\mathrm{d}P(x) < \infty \right\}. \tag{5.1}$$

If $p\colon \mathbb{R}^n \to [0,\infty)$ is a probability density function on $\mathbb{R}^n$, we can identify it with the probability measure $P \in \mathscr{P}(\mathbb{R}^n)$ having density p with respect to Lebesgue measure on $\mathbb{R}^n$. In particular, if p is a probability density with finite second moment, i.e.,

$$\int_{\mathbb{R}^n} |x|^2\, p(x)\,\mathrm{d}x < \infty, \tag{5.2}$$

then we can identify p with an element of $\mathscr{P}_2(\mathbb{R}^n)$.

We denote by $\Gamma(P_1,P_2)$ the collection of Kantorovich transport plans, that is, probability measures $\boldsymbol{\gamma}$ in $\mathscr{P}(\mathbb{R}^n\times\mathbb{R}^n)$ with given marginals $P_1,P_2\in\mathscr{P}(\mathbb{R}^n)$. More precisely, if $\pi^i\colon \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ are the canonical projections, then $\pi^i_\#\boldsymbol{\gamma} = P_i$, for $i\in\{1,2\}$. The *quadratic Wasserstein distance* between two probability measures $P_1,P_2\in\mathscr{P}_2(\mathbb{R}^n)$ is defined by

$$W_2^2(P_1,P_2) := \inf\Big\{ \int_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\,\mathrm{d}\boldsymbol{\gamma}(x,y)\colon\ \boldsymbol{\gamma}\in\Gamma(P_1,P_2)\Big\}. \tag{5.3}$$

The quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$, endowed with the quadratic Wasserstein distance W_2 just introduced, is a Polish space [**?**, Proposition 7.1.5].

5.2 The metric derivative of curves in the Wasserstein space

In the present section we consider the solution $(p(t))_{t\geqslant 0}$ of the Fokker-Planck equation (**??**) *with initial condition* (**??**) *as a curve in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$.* This is justified by Lemma **??**, on the basis of which the Assumptions **??** guarantee that $p(t)\in\mathscr{P}_2(\mathbb{R}^n)$ for all $t\geqslant 0$. For fixed $T\in(0,\infty)$, we define now the time-dependent velocity field

$$[0,T]\times\mathbb{R}^n \ni (t,x) \longmapsto v(t,x) := -\Big(\frac{1}{2}\frac{\nabla p(t,x)}{p(t,x)} + \nabla\Psi(x)\Big) = -\frac{1}{2}\frac{\nabla \ell(t,x)}{\ell(t,x)} \in \mathbb{R}^n. \tag{5.4}$$

Then the Fokker-Planck equation (**??**), satisfied by the curve $(p(t))_{0\leqslant t\leqslant T}$ of probability density functions in $\mathscr{P}_2(\mathbb{R}^n)$, can be cast as a *continuity equation*, namely,

$$\partial_t p(t,x) + \operatorname{div}\big(v(t,x)\,p(t,x)\big) = 0, \qquad (t,x)\in(0,T]\times\mathbb{R}^n. \tag{5.5}$$

According to (**??**), we have

$$2\int_0^T \Big(\int_{\mathbb{R}^n} |v(t,x)|^2\,p(t,x)\,\mathrm{d}x\Big)\,\mathrm{d}t < \infty, \tag{5.6}$$

since by definition of the velocity field $v(t)\equiv v(t,\,\cdot\,)$ the expressions in (**??**) and (**??**) are simply the same.

As $(p(t))_{0\leqslant t\leqslant T}$ is a curve in the Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (**??**) as well as the integrability condition (**??**), we can invoke Theorem 8.3.1 in [**?**]. This result relates absolutely continuous curves in $\mathscr{P}_2(\mathbb{R}^n)$ to the continuity equation. In particular, its second implication tells us that the curve $(p(t))_{0\leqslant t\leqslant T}$ is absolutely continuous [**?**, Definition 1.1.1]. As a consequence, for Lebesgue-almost every $t_0\in[0,T]$, its *metric derivative* [**?**, Theorem 1.1.2]

$$|p'|(t_0) := \lim_{t\to t_0} \frac{W_2(p(t),p(t_0))}{|t-t_0|} \tag{5.7}$$

exists. Furthermore, [**?**, Theorem 8.3.1] provides for Lebesgue-almost every $t_0\in[0,T]$ the estimate

$$|p'|(t_0) \leqslant \|v(t_0)\|_{L^2(P(t_0))}. \tag{5.8}$$

On the other hand, according to condition **??** in Assumptions **??**, the time-dependent gradient vector field $v\colon [0,T]\times\mathbb{R}^n\to\mathbb{R}^n$ of (**??**) is an element of the tangent space [**?**, Definition 8.4.1] of $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P(t)\in\mathscr{P}_2(\mathbb{R}^n)$, i.e.,

$$v(t,\,\cdot\,)\in \mathrm{Tan}_{P(t)}\mathscr{P}_2(\mathbb{R}^n) := \overline{\{\nabla\varphi\colon\ \varphi\in\mathcal{C}_c^\infty(\mathbb{R}^n;\mathbb{R})\}}^{L^2(P(t))}. \tag{5.9}$$

Since $(p(t))_{0\leqslant t\leqslant T}$ is an absolutely continuous curve in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (**??**) for the time-dependent velocity field $v(t)$, which is tangent to the curve, we can apply Proposition 8.4.5 of [**?**]. This result characterizes tangent vectors to absolutely continuous curves, and entails for Lebesgue-almost every $t_0\in[0,T]$ the inequality

$$|p'|(t_0)\geqslant \|v(t_0)\|_{L^2(P(t_0))}. \tag{5.10}$$

Combining (**??**) and (**??**), we obtain for Lebesgue-almost every $t_0\in[0,T]$ the equality

$$|p'|(t_0)= \|v(t_0)\|_{L^2(P(t_0))}. \tag{5.11}$$

This relates the strength of the time-dependent velocity field $v(t,\,\cdot\,)$ in (**??**), to the rate of change, or metric derivative as in (**??**), of the Wasserstein distance along the curve $(p(t))_{0\leqslant t\leqslant T}$ — justifying in this manner the introduction and relevance of the Wasserstein distance in this context.

5.3 The local behavior of the Wasserstein distance

In the previous section we derived along the lines of [**?**] the explicit representation (**??**) of the metric derivative of the quadratic Wasserstein distance along the Fokker-Planck probability density flow. This limit exists for Lebesgue-almost every $t_0\geqslant 0$. In the following theorem we provide accurate information regarding the points $t_0\geqslant 0$ at which the metric derivative (**??**) exists (cf. Remark **??**), and hence record its instructive proof.

Theorem 5.1 (Local behavior of the quadratic Wasserstein distance)**.** *Under the* Assumptions **??**, *let $t_0\geqslant 0$ be such that the generalized de Bruijn identity* (**??**), (**??**) *is valid. Then we have*

$$\lim_{t\to t_0}\frac{W_2(P(t),P(t_0))}{|t-t_0|} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,|v(t_0,X(t_0))|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\sqrt{I(P(t_0)\,|\,\mathrm{Q})}. \tag{5.12}$$

Instead of Theorem **??**, we will prove the more general Theorem **??** below, which amounts to the perturbed version of Theorem **??**. The latter then simply follows by setting $\beta\equiv 0$ in the statement of Theorem **??**.

In order to formulate Theorem **??**, we consider the solution $(p^\beta(t))_{t\geqslant t_0}$ of the perturbed Fokker-Planck equation (**??**) with initial condition (**??**). Again, according to Lemma **??**, this solution can be viewed as a curve in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$. Just as before, we define the time-dependent perturbed velocity field

$$[t_0,T]\times\mathbb{R}^n\ni(t,x)\longmapsto v^\beta(t,x) := -\bigg(\frac{1}{2}\frac{\nabla p^\beta(t,x)}{p^\beta(t,x)}+\nabla\Psi(x)+\beta(x)\bigg)\in\mathbb{R}^n. \tag{5.13}$$

Then the perturbed Fokker-Planck equation (**??**), satisfied by the perturbed curve $(p^\beta(t))_{t_0\leqslant t\leqslant T}$, can once again be written as a continuity equation, to wit

$$\partial_t p^\beta(t,x)+\mathrm{div}\,(v^\beta(t,x)\,p^\beta(t,x))=0,\qquad (t,x)\in(t_0,T]\times\mathbb{R}^n. \tag{5.14}$$

At this point, let us recall that we have required the perturbation $\beta\colon \mathbb{R}^n \to \mathbb{R}^n$ to be 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 (**??**) and (**??**) are related via

$$v^\beta(t_0, x) = v(t_0, x) - \nabla B(x), \qquad x \in \mathbb{R}^n. \tag{5.15}$$

Using the regularity assumption that the potential B is of class $\mathcal{C}_c^\infty(\mathbb{R}^n; \mathbb{R})$, we conclude from (**??**) and (**??**) that the perturbed vector field $v^\beta(t_0, \cdot\,)$ is 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.,

$$v^\beta(t_0, \cdot\,) \in \mathrm{Tan}_{P^\beta(t_0)}\mathscr{P}_2(\mathbb{R}^n) = \overline{\{\nabla\varphi^\beta\colon\ \varphi^\beta \in \mathcal{C}_c^\infty(\mathbb{R}^n; \mathbb{R})\}}^{L^2(P^\beta(t_0))}. \tag{5.16}$$

After these preparations we can formulate the perturbed version of Theorem **??** as follows.

Theorem 5.2 (Local behavior of the quadratic Wasserstein distance under perturbations)**.** *Under the* Assumptions **??**, *for every $t_0 \in \mathbb{R}_+ \setminus N$ we have*

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}, \tag{5.17}$$

where $a = \nabla \log \ell(t_0, X(t_0))$ and $b = \beta(X(t_0))$.

Remark 5.3. With this notation, the rightmost side of (**??**) is $\frac{1}{2}\|a\|_{L^2(\mathbb{P})}$. Since $X(t_0)$ has the same probability distribution under $\mathbb{P}$, as it does under $\mathbb{P}^\beta$, the expectation $\mathbb{E}_{\mathbb{P}}$ appearing in (**??**) can be replaced by $\mathbb{E}_{\mathbb{P}^\beta}$. Let us also recall from Remark **??** that the exceptional set N consists of those points $t_0 \geqslant 0$ for which the limiting assertion (**??**) fails. ⋄

Proof of Theorem **??**. The second equality in (**??**) is apparent from the definition of the time-dependent perturbed velocity field $\big(v^\beta(t, \cdot\,)\big)_{t \geqslant t_0}$ from (**??**) above. The delicate point is to show that the limiting assertion (**??**) 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 (**??**) is valid. In the following steps we prove that then also the limiting assertion (**??**) does hold.

Step 1. The vector field $v^\beta(t_0, \cdot\,)$ induces a family of *linearized transport maps* $(\Upsilon_t^\beta)_{t \geqslant t_0}$ defined by

$$\Upsilon_t^\beta(x) := x + (t - t_0) \cdot v^\beta(t_0, x), \qquad x \in \mathbb{R}^n \tag{5.18}$$

in the manner of (**??**), and we denote by $P_\Upsilon^\beta(t)$ the image measure of $P^\beta(t_0) = P(t_0)$ under the transport map $\Upsilon_t^\beta\colon \mathbb{R}^n \to \mathbb{R}^n$; i.e.,

$$P_\Upsilon^\beta(t) := (\Upsilon_t^\beta)_\# P^\beta(t_0), \qquad t \geqslant t_0. \tag{5.19}$$

To motivate the subsequent arguments, let us first pretend that, for all $t > t_0$ sufficiently close to t_0, the map Υ_t^β is the *optimal quadratic Wasserstein transport* from $P^\beta(t_0)$ to $P_\Upsilon^\beta(t)$; i.e.,

$$W_2^2(P_\Upsilon^\beta(t), P^\beta(t_0)) = \mathbb{E}_{\mathbb{P}^\beta}\Big[\,\big|\Upsilon_t^\beta(X(t_0)) - X(t_0)\big|^2\,\Big] = \mathbb{E}_{\mathbb{P}}\Big[\,\big|\Upsilon_t^\beta(X(t_0)) - X(t_0)\big|^2\,\Big], \tag{5.20}$$

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 (**??**), we could conclude that

$$\lim_{t \downarrow t_0} \frac{W_2(P_\Upsilon^\beta(t), P^\beta(t_0))}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}. \tag{5.21}$$

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

$$\lim_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta_\Upsilon(t))}{t-t_0} = 0. \tag{5.22}$$

Using (**??**) and (**??**), we would now derive the desired equality (**??**). Indeed, invoking the triangle inequality for the quadratic Wasserstein distance we obtain

$$\lim_{t\downarrow t_0} \frac{W_2(P^\beta_\Upsilon(t), P^\beta(t_0))}{t-t_0} \leqslant \lim_{t\downarrow t_0} \frac{W_2(P^\beta_\Upsilon(t), P^\beta(t))}{t-t_0} + \liminf_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t-t_0}, \tag{5.23}$$

and one more application of it yields

$$\limsup_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t-t_0} \leqslant \lim_{t\downarrow t_0} \frac{W_2(P^\beta(t), P^\beta_\Upsilon(t))}{t-t_0} + \lim_{t\downarrow t_0} \frac{W_2(P^\beta_\Upsilon(t), P^\beta(t_0))}{t-t_0}. \tag{5.24}$$

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 Υ^β_t defined in (**??**) of Step 1 should be optimal with respect to quadratic Wasserstein transportation costs; i.e., by *Brenier's theorem* [**?**], 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 Υ^β_t as in Step 1 by localizing the vector field $v^\beta(t_0, \cdot\,)$ as well as the transport maps $(\Upsilon^\beta_t)_{t\geqslant t_0}$ to compact subsets of $\mathbb{R}^n$ (Steps 2 – 4); and that, after these localizations have been carried out, an analogue of the identity (**??**) also holds, allowing us to complete the argument (Steps 5 – 7).

To this end, we first 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, compactly supported potential $B\colon \mathbb{R}^n \to \mathbb{R}$. Thus the vector field $v^\beta(t_0, \cdot\,)$ from (**??**), (**??**) can be represented as a gradient, namely

$$v^\beta(t_0, x) = -\nabla\Big(\tfrac{1}{2}\log \ell(t_0, x) + B(x)\Big), \qquad x \in \mathbb{R}^n. \tag{5.25}$$

Even more, according to (**??**), it is an element of the tangent space

$$\mathrm{Tan}_{P^\beta(t_0)}\mathscr{P}_2(\mathbb{R}^n) = \overline{\{\nabla\varphi^\beta\colon\ \varphi^\beta \in \mathcal{C}^\infty_c(\mathbb{R}^n;\mathbb{R})\}}^{L^2(P^\beta(t_0))} \tag{5.26}$$

of the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ at the point $P^\beta(t_0) \in \mathscr{P}_2(\mathbb{R}^n)$. Therefore we can choose a sequence of potential functions $\big(\varphi^\beta_m(t_0, \cdot\,)\big)_{m\geqslant 1} \subseteq \mathcal{C}^\infty_c(\mathbb{R}^n;\mathbb{R})$ such that

$$\lim_{m\to\infty} \mathbb{E}_{\mathbb{P}}\Big[\ \Big|v^\beta\big((t_0, X(t_0)\big) - \nabla\varphi^\beta_m(t_0, X(t_0))\Big|^2\ \Big] = 0. \tag{5.27}$$

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

$$v^\beta_m(t_0, x) := \nabla\varphi^\beta_m(t_0, x), \qquad x \in \mathbb{R}^n. \tag{5.28}$$

By construction, these have compact support and approximate the gradient vector field $v^\beta(t_0, \cdot\,)$ in $L^2(P(t_0))$, as in (**??**).

Finally, for every $m \in \mathbb{N}$, the localized gradient vector field $v^\beta_m(t_0, \cdot\,)$ induces a family of *localized linear transports* $(\Upsilon^{\beta,m}_t)_{t\geqslant t_0}$ defined by analogy with (**??**) via

$$\Upsilon^{\beta,m}_t(x) := x + (t-t_0)\cdot v^\beta_m(t_0, x), \qquad x \in \mathbb{R}^n. \tag{5.29}$$

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

$$P_\Upsilon^{\beta,m}(t) := (\Upsilon_t^{\beta,m})_\# P^\beta(t_0), \qquad t \geqslant t_0. \tag{5.30}$$

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

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

$$\Upsilon_t^{\beta,m}(x) = \nabla\Big(\tfrac{1}{2}|x|^2 + (t - t_0) \cdot \varphi_m^\beta(t_0, x)\Big), \qquad x \in \mathbb{R}^n. \tag{5.31}$$

Therefore, it remains to show that the function $\frac{1}{2}|\cdot|^2 + (t - t_0) \cdot \varphi_m^\beta(t_0, \cdot\,)$ is convex for every $m \in \mathbb{N}$ and $t > t_0$ sufficiently close to t_0. The Hessian matrix of this function is given by

$$I_n + (t - t_0) \cdot \mathrm{Hess}(\varphi_m^\beta(t_0, x)), \qquad x \in \mathbb{R}^n. \tag{5.32}$$

In order to deduce the desired convexity, we have to verify that the Hessian matrix of (**??**) is positive semidefinite for all $t > t_0$ sufficiently near t_0, uniformly in $x \in \mathbb{R}^n$. To see this, let us fix $m \in \mathbb{N}$. Now the identity matrix I_n is positive definite and the Hessian matrix of the function $\varphi_m^\beta(t_0, \cdot\,)$ is symmetric. Furthermore, we recall that the smooth function $\varphi_m^\beta(t_0, \cdot\,)$ has *compact support*, which is crucial in order to justify the present argument. Checking the defining condition guaranteeing positive semidefiniteness of the matrix in (**??**) for unit vectors and using compactness as well as continuity, we obtain the existence of some $\varepsilon_m > 0$ such that, for all $t > t_0$ with $|t - t_0| < \varepsilon_m$, the Hessian matrix of (**??**) is positive semidefinite (in fact, positive definite).

Step 4. From Step 3 we know that, for every $m \in \mathbb{N}$, there exists some $\varepsilon_m > 0$ such that for all $t > t_0$ with $|t - t_0| < \varepsilon_m$, the localized map $\Upsilon_t^{\beta,m}$ is the optimal transport from $P^\beta(t_0)$ to $P_\Upsilon^{\beta,m}(t)$ with respect to quadratic Wasserstein costs. Therefore, we can apply the considerations of Step 1 to the optimal map $\Upsilon_t^{\beta,m}$ in (**??**), and deduce that

$$\lim_{t \downarrow t_0} \frac{W_2\big(P_\Upsilon^{\beta,m}(t), P^\beta(t_0)\big)}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v_m^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} \tag{5.33}$$

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

$$\lim_{m \to \infty} \lim_{t \downarrow t_0} \frac{W_2\big(P_\Upsilon^{\beta,m}(t), P^\beta(t_0)\big)}{t - t_0} = \Big(\mathbb{E}_{\mathbb{P}}\Big[\,\big|v^\beta(t_0, X(t_0))\big|^2\,\Big]\Big)^{1/2} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P})}. \tag{5.34}$$

From the inequalities (**??**) and (**??**) of Step 1 (with $P_\Upsilon^{\beta,m}(t)$ instead of $P_\Upsilon^\beta(t)$) it follows that, in order to conclude (**??**), it remains to establish the analogue

$$\lim_{m \to \infty} \lim_{t \downarrow t_0} \frac{W_2\big(P^\beta(t), P_\Upsilon^{\beta,m}(t)\big)}{t - t_0} = 0 \tag{5.35}$$

of the identity (**??**).

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

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,\Phi_t^\beta = v^\beta(t, \Phi_t^\beta) \quad \text{ for all } t \geqslant t_0\,, \qquad \Phi_{t_0}^\beta = \mathrm{Id}_{\mathbb{R}^n}. \tag{5.36}$$

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

The localized linear transports $\Upsilon_t^{\beta,m} \colon \mathbb{R}^n \to \mathbb{R}^n$ defined in (**??**) of Step 2 transport $P^\beta(t_0)$ to $P_\Upsilon^{\beta,m}(t)$, see (**??**). As $P^\beta(t_0)$ and $P_\Upsilon^{\beta,m}(t)$ have full support and are absolutely continuous with respect to Lebesgue measure, the inverse map $(\Upsilon_t^{\beta,m})^{-1} \colon \mathbb{R}^n \to \mathbb{R}^n$ is well-defined and satisfies

$$((\Upsilon_t^{\beta,m})^{-1})_\# P_\Upsilon^{\beta,m}(t) = P^\beta(t_0), \qquad t \geqslant t_0. \tag{5.37}$$

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

$$\lim_{m\to\infty} \lim_{t\downarrow t_0} \frac{1}{t-t_0} \left(\mathbb{E}_{\mathbb{P}_\Upsilon^{\beta,m}} \Big[\, \Big| \mathfrak{X}_t^{\beta,m}(X(t)) - X(t) \Big|^2 \, \Big] \right)^{1/2} = 0\,, \tag{5.38}$$

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

$$\mathfrak{X}_t^{\beta,m} := \Phi_t^\beta \circ (\Upsilon_t^{\beta,m})^{-1}, \qquad t \geqslant t_0 \tag{5.39}$$

for this job, recall that $(\Upsilon_t^{\beta,m})^{-1}$ transports $P_\Upsilon^{\beta,m}(t)$ to $P^\beta(t_0)$ while Φ_t^β transports $P^\beta(t_0)$ to $P^\beta(t)$, and conclude that $\mathfrak{X}_t^{\beta,m}$ of (**??**) transports $P_\Upsilon^{\beta,m}(t)$ to $P^\beta(t)$; thus, we have

$$\mathbb{E}_{\mathbb{P}_\Upsilon^{\beta,m}} \Big[\, \Big| \mathfrak{X}_t^{\beta,m}(X(t)) - X(t) \Big|^2 \, \Big] = \mathbb{E}_{\mathbb{P}} \Big[\, \Big| \Phi_t^\beta(X(t_0)) - \Upsilon_t^{\beta,m}(X(t_0)) \Big|^2 \, \Big]. \tag{5.40}$$

Combining (**??**) and (**??**), we see that we have to establish

$$\lim_{m\to\infty} \lim_{t\downarrow t_0} \frac{1}{(t-t_0)^2} \, \mathbb{E}_{\mathbb{P}} \Big[\, \Big| \Phi_t^\beta(X(t_0)) - \Upsilon_t^{\beta,m}(X(t_0)) \Big|^2 \, \Big] = 0. \tag{5.41}$$

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

$$\tfrac{1}{2} \, |\Phi_t^\beta(x) - \Upsilon_t^{\beta,m}(x)|^2 \leqslant (t-t_0)^2 \cdot |v^\beta(t_0,x) - v_m^\beta(t_0,x)|^2 \tag{5.42}$$

$$+ \Big| (\Phi_t^\beta(x) - x) - (t-t_0) \cdot v^\beta(t_0,x) \Big|^2. \tag{5.43}$$

Therefore, in order to establish (**??**), it suffices to show the limiting assertions (**??**) and (**??**) below; these correspond to (**??**) and (**??**), respectively.

The first limiting identity

$$\lim_{m\to\infty} \mathbb{E}_{\mathbb{P}} \Big[\, \Big| v^\beta\big((t_0, X(t_0)\big) - v_m^\beta(t_0, X(t_0)) \Big|^2 \, \Big] = 0 \tag{5.44}$$

we already have from (**??**), (**??**) of Step 2.

Step 6. Our final task is to justify that

$$\lim_{t\downarrow t_0} \mathbb{E}_{\mathbb{P}} \Big[\, \Big| \tfrac{1}{t-t_0} \Big(\Phi_t^\beta(X(t_0)) - X(t_0) \Big) - v^\beta(t_0, X(t_0)) \Big|^2 \, \Big] = 0. \tag{5.45}$$

To this end, we first note that by (**??**) we have the identity

$$\Phi_t^\beta(x) = x + \int_{t_0}^t v^\beta\big(u, \Phi_u^\beta(x)\big)\,\mathrm{d}u, \qquad x \in \mathbb{R}^n, \tag{5.46}$$

for all $t \geqslant t_0$. On account of it we see that the expectation in (**??**) is equal to

$$\mathbb{E}_{\mathbb{P}}\left[\, \left| \frac{1}{t-t_0} \int_{t_0}^t v^\beta\Big(u, \Phi_u^\beta\big(X(t_0)\big)\Big)\,\mathrm{d}u - v^\beta\big(t_0, X(t_0)\big) \right|^2 \,\right]. \tag{5.47}$$

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

$$\lim_{t \downarrow t_0} \mathbb{E}_{\mathbb{P}^\beta}\left[\, \left| \frac{1}{t-t_0} \int_{t_0}^t v^\beta\big(u, X(u)\big)\,\mathrm{d}u - v^\beta\big(t_0, X(t_0)\big) \right|^2 \,\right] = 0. \tag{5.48}$$

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

$$\left(\left| \frac{1}{t-t_0} \int_{t_0}^t v^\beta\big(u, X(u)\big)\,\mathrm{d}u - v^\beta\big(t_0, X(t_0)\big) \right|^2 \right)_{t \geqslant t_0} \tag{5.49}$$

converges $\mathbb{P}^\beta$-almost surely to zero, as $t \downarrow t_0$. In order to show that their expectations also converge to zero, i.e., that (**??**) does hold, we have to verify that the family of (**??**) 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

$$\left| \frac{1}{t-t_0} \int_{t_0}^t v^\beta\big(u, X(u)\big)\,\mathrm{d}u \right|^2 \leqslant \frac{1}{t-t_0} \int_{t_0}^t \big|v^\beta\big(u, X(u)\big)\big|^2\,\mathrm{d}u, \qquad t \geqslant t_0 \tag{5.50}$$

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

$$\left(\frac{1}{t-t_0} \int_{t_0}^t \big|v^\beta\big(u, X(u)\big)\big|^2\,\mathrm{d}u \right)_{t \geqslant t_0}. \tag{5.51}$$

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

$$\left(\frac{1}{t-t_0} \int_{t_0}^t \frac{\big|\nabla \ell^\beta\big(u, X(u)\big)\big|^2}{\ell^\beta\big(u, X(u)\big)^2}\,\mathrm{d}u \right)_{t \geqslant t_0}. \tag{5.52}$$

To establish this uniform integrability, it suffices to show that the family of random variables in (**??**) converges in $L^1(\mathbb{P}^\beta)$. Now by continuity the family of (**??**) converges $\mathbb{P}^\beta$-almost surely to $|\nabla \log \ell(t_0, X(t_0))|^2$. Hence in view of *Scheffé's lemma* (Lemma **??**) 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 (**??**), (**??**) from the proof of Corollary **??**, which gives us

$$\lim_{t \downarrow t_0} \mathbb{E}_{\mathbb{P}^\beta}\left[\frac{1}{t-t_0} \int_{t_0}^t \frac{\big|\nabla \ell^\beta\big(u, X(u)\big)\big|^2}{\ell^\beta\big(u, X(u)\big)^2}\,\mathrm{d}u \right] = \mathbb{E}_{\mathbb{P}}\left[\, \frac{\big|\nabla \ell\big(t_0, X(t_0)\big)\big|^2}{\ell\big(t_0, X(t_0)\big)^2} \,\right], \tag{5.53}$$

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

Summing up, in light of (**??**), (**??**) from Step 5, the limiting assertions (**??**) and (**??**) imply the limiting behavior (**??**). According to the results of Steps 4 and 5, the latter also entails the validity of the limiting identity (**??**), which completes the proof of Theorem **??**. □

Equipped with Theorem **??**, we can now easily deduce Theorem **??**.

Proof of Theorem **??**. The second equality in (**??**) follows from the representation of the relative Fisher information in (**??**) and the definition of the time-dependent velocity field $\big(v(t,\,\cdot\,)\big)_{t\geqslant t_0}$ in (**??**). The first equality in (**??**) is a direct consequence of Theorem **??**. One just has to set $\beta \equiv 0$ in (**??**). However, the careful reader might note that the limit in (**??**) is only from the right, while the limit in (**??**) is two-sided. But the only reason why we considered right-sided limits in Theorem **??** 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 **??**. □

Appendices

A Bachelier's work relating Brownian motion to the heat equation

In this section, which is only of historical interest, we point out that Bachelier already had some thoughts on *"horizontal transport of probability measures"* in his dissertation *"Théorie de la spéculation"* [**?**, **?**], which he defended in 1900.

In this work he was the first to consider a *mathematical model* of Brownian motion. Bachelier argued using infinitesimals by visualizing Brownian motion $(W(t))_{t\geqslant 0}$ as an infinitesimal version of a random walk. Suppose that the grid in space is given by

$$\ldots,\ x_{n-2},\ x_{n-1},\ x_n,\ x_{n+1},\ x_{n+2},\ \ldots \tag{A.1}$$

having the same (infinitesimal) distance $\Delta x = x_n - x_{n-1}$, for all n, and such that at time t these points have (infinitesimal) probabilities

$$\ldots,\ p^t_{n-2},\ p^t_{n-1},\ p^t_n,\ p^t_{n+1},\ p^t_{n+2},\ \ldots \tag{A.2}$$

under the random walk under consideration. What are the probabilities

$$\ldots,\ p^{t+\Delta t}_{n-2},\ p^{t+\Delta t}_{n-1},\ p^{t+\Delta t}_n,\ p^{t+\Delta t}_{n+1},\ p^{t+\Delta t}_{n+2},\ \ldots \tag{A.3}$$

of these points at time $t + \Delta t$?

The random walk moves half of the mass p^t_n, sitting on x_n at time t, to the point x_{n+1}. En revanche, it moves half of the mass p^t_{n+1}, sitting on x_{n+1} at time t, to the point x_n. The net difference between $p^t_n/2$ and $p^t_{n+1}/2$, which Bachelier has no scruples to identify with

$$-\tfrac{1}{2}\,(p^t)'(x_n)\,\Delta x = -\tfrac{1}{2}\,(p^t)'(x_{n+1})\,\Delta x, \tag{A.4}$$

is therefore transported from the interval $(-\infty, x_n]$ to $[x_{n+1}, \infty)$. In Bachelier's own words, this is very nicely captured by the following passage of his thesis:

"Each price x during an element of time radiates towards its neighboring price an amount of probability proportional to the difference of their probabilities. I say proportional because it is necessary to account for the relation of Δx to Δt. The above law can, by analogy with certain physical theories, be called the law of radiation or diffusion of probability."

Passing formally to the continuous limit and denoting by

$$P(t,x) = \int_{-\infty}^{x} p(t,z)\,\mathrm{d}z \tag{A.5}$$

the distribution function associated to the Gaussian density function $p(t,x)$, Bachelier deduces in an intuitively convincing way the relation

$$\frac{\partial P}{\partial t} = \frac{1}{2}\frac{\partial p}{\partial x}, \tag{A.6}$$

where we have normalized the relation between Δx and Δt to obtain the constant $1/2$. By differentiating (**??**) with respect to x one obtains the usual heat equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2} \tag{A.7}$$

for the density function $p(t,x)$. Of course, the heat equation was known to Bachelier, and he notes regarding (**??**): *"C'est une équation de Fourier."*

But let us still remain with the form (**??**) of the heat equation and analyze its message in terms of *"horizontal transport of probability measures"*. To accomplish the movement of mass $-\frac{1}{2}\,p'(t,x)\,\mathrm{d}x$ from $(-\infty,x]$ to $[x,\infty)$ one is naturally led to define the flow induced by the velocity field

$$v(t,x) := -\frac{1}{2}\,\frac{p'(t,x)}{p(t,x)}, \tag{A.8}$$

which has the natural interpretation as the "speed" of the transport induced by $p(t,x)$. We thus encounter *in nuce* the ubiquitous "score function" $\nabla p(t,x)/p(t,x)$ appearing throughout all the above considerations. We also note that an "infinitesimal transport" on $\mathbb{R}$ is automatically an optimal transport. Intuitively this corresponds to the geometric insight in the one-dimensional case that the transport lines of infinitesimal length cannot cross each other.

Let us go one step beyond Bachelier's thoughts and consider the relation of the above infinitesimal Wasserstein transport to time reversal (which Bachelier had not yet considered in his solitary exploration of Brownian motion). Visualizing again the grid (**??**) and the corresponding probabilities (**??**) and (**??**), a moment's reflection reveals that the transport from $p^{t+\Delta t}$ to p^t, i.e., in reverse direction, is accomplished by going from x_n to x_{n+1} with probability $\frac{1}{2}+\frac{p'(t,x)}{p(t,x)}\,\mathrm{d}x$ and from x_{n+1} to x_n with probability $\frac{1}{2}-\frac{p'(t,x)}{p(t,x)}\,\mathrm{d}x$, with the identifications $x=x_n=x_{n+1}$, and $\mathrm{d}x=\Delta x$. In other words, the above Brownian motion $(W(t))_{t\geqslant 0}$ considered in reverse direction $(W(T-s))_{0\leqslant s\leqslant T}$ is *not* a Brownian motion, as the transition probabilities are not $(1/2,1/2)$ any more. Rather, one has to correct these probabilities by a term which — once again — involves our familiar score function $\nabla p(t,x)/p(t,x)$ (compare (**??**) above). At this stage, it should come as no surprise, that the passage to reverse time is closely related to the Wasserstein transport induced by $p(t,x)$.

Let us play this infinitesimal reasoning one more time, in order to visualize the Fontbona-Jourdain result (Theorem **??**). Arguing in the reverse direction of time, we may ask the following question: how do we have to choose the transition probabilities to go from x at time $t+\Delta t$ to either $x+\mathrm{d}x$ or $x-\mathrm{d}x$ at time t, so that the density process $p(t,x)$ becomes a martingale in reverse time under these transition probabilities? As the difference between the probabilities $p(t,x+\mathrm{d}x)$ and $p(t,x-\mathrm{d}x)$ equals $2\,p'(t,x)\,\mathrm{d}x$ (up to terms of smaller order than $\mathrm{d}x$) we conclude that the transition probabilities have to be changed from $(1/2,1/2)$ to

$$\left(\frac{1}{2}-\frac{p'(t,x)}{p(t,x)}\,\mathrm{d}x\,,\;\frac{1}{2}+\frac{p'(t,x)}{p(t,x)}\,\mathrm{d}x\right) \tag{A.9}$$

in order to counterbalance this difference of probabilities (again up to terms of smaller order than $\mathrm{d}x$). In other words, we have found again precisely the same transition probabilities which we had encountered in the context of the reversed Brownian process $(W(T-s))_{0\leqslant s\leqslant T}$. This provides some intuition for the Fontbona-Jourdain assertion that $(p(T-s,W(T-s))_{0\leqslant s\leqslant T}$ is a martingale in the reverse direction of time.

We finish the section by returning to Bachelier's thesis. The *rapporteur* of Bachelier's dissertation was no lesser a figure than Henri Poincaré. Apparently he was aware of the enormous potential of the section *"Rayonnement de la probabilité"* in Bachelier's thesis, when he added to his very positive report the handwritten phrase: *"On peut regretter que M. Bachelier n'ait pas développé davantage cette partie de sa thèse."* That is: One might regret that Mr. Bachelier did not develop further this part of his thesis.

B The proofs of Lemmas ?? and ??

Proof of Lemma **??**. Let the real constants $c,R\geqslant 0$ be as in condition **??** of Assumptions **??**, and denote

$$m_R := \max_{|x|\leqslant R}\left|1-\left\langle x\,,\,2\,\nabla\Psi(x)\right\rangle_{\mathbb{R}^n}\right|<\infty, \qquad \tau_k := \inf\left\{t\geqslant 0\colon\ |X(t)|>k\right\} \tag{B.1}$$

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

$$\mathrm{d}|X(t)|^2 = \Big(1 - \Big\langle X(t)\,,\, 2\,\nabla\Psi\big(X(t)\big)\Big\rangle_{\mathbb{R}^n}\Big)\,\mathrm{d}t + \big\langle 2\,X(t)\,,\,\mathrm{d}W(t)\big\rangle_{\mathbb{R}^n} \tag{B.2}$$

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

$$\varphi_k(t) = \varphi(0) + \mathbb{E}_{\mathbb{P}}\bigg[\int_0^{t\wedge\tau_k}\Big(1 - \Big\langle X(u)\,,\,2\,\nabla\Psi\big(X(u)\big)\Big\rangle_{\mathbb{R}^n}\Big)\,\mathbb{1}_{\{|X(u)|\leqslant R\}}\,\mathrm{d}u\bigg] \tag{B.3}$$

$$+\,\mathbb{E}_{\mathbb{P}}\bigg[\int_0^{t\wedge\tau_k}\Big(1 - \Big\langle X(u)\,,\,2\,\nabla\Psi\big(X(u)\big)\Big\rangle_{\mathbb{R}^n}\Big)\,\mathbb{1}_{\{|X(u)|>R\}}\,\mathrm{d}u\bigg] \tag{B.4}$$

$$\leqslant \varphi(0) + m_R\,\mathbb{E}_{\mathbb{P}}[t\wedge\tau_k] + \mathbb{E}_{\mathbb{P}}\bigg[\int_0^{t\wedge\tau_k}\big(1 + 2\,c\,|X(u)|^2\,\big)\,\mathrm{d}u\bigg] \tag{B.5}$$

$$\leqslant \varphi(0) + (1+m_R)\,t + 2\,c\int_0^t \varphi_k(u)\,\mathrm{d}u. \tag{B.6}$$

The Gronwall inequality gives now

$$\varphi_k(t) \leqslant g(t) := \varphi(0) + (1+m_R)\,t + 2\,c\int_0^t \big(\varphi(0) + (1+m_R)\,u\big)\,\mathrm{e}^{2c(t-u)}\,\mathrm{d}u. \tag{B.7}$$

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

$$\varphi(t) = \mathbb{E}_{\mathbb{P}}\big[\,|X(t)|^2\,\big] \leqslant g(t) < \infty, \qquad t \geqslant 0. \tag{B.8}$$

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

Proof of Lemma **??**. The proof of Lemma **??** follows by analogy with the proof of Lemma **??** above. Indeed, one just has to add the perturbation β to the gradient $\nabla\Psi$ in the expressions (**??**) – (**??**), write $W^\beta(t)$ instead of $W(t)$ in (**??**), and replace all the $\mathbb{P}$-expectations by expectations with respect to the probability measure $\mathbb{P}^\beta$. Then the constant m_R in (**??**) is still finite and, because of its compact support, the perturbation β in the expression (**??**) 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 $t \geqslant t_0$. □

C Relative entropy with respect to a σ-finite measure Q

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

$$H(\mathcal{P}\,|\,\mathcal{Q}) := \int_{\mathbb{R}^n} \log\left(\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{Q}}\right)\mathrm{d}\mathcal{P} \in [0,\infty] \tag{C.1}$$

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 consider the σ-finite measure Q on the Borel sets of $\mathbb{R}^n$ with density $q(x) = \mathrm{e}^{-2\Psi(x)}$ for $x \in \mathbb{R}^n$, introduced in Section **??**. Following the approach of [**?**, Section 2], 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 (**??**) now takes values in $(-\infty,\infty]$.

To this end, we let P be a probability measure in $\mathscr{P}_2(\mathbb{R}^n)$ having density $p\colon \mathbb{R} \to [0,\infty)$ with respect to Lebesgue measure. The non-negativity of the potential Ψ implies that for the function

$f(x) := \mathrm{e}^{-|x|^2}$, $x \in \mathbb{R}$, we have $\mathbb{E}_{\mathrm{Q}}[f] \in (0,\infty)$. Following [**?**, Section 2], we let $\mathcal{Q}$ be the probability measure on the Borel sets of $\mathbb{R}^n$ having probability density function $f/\mathbb{E}_{\mathrm{Q}}[f]$ with respect to the measure Q, so that

$$\frac{\mathrm{d}P}{\mathrm{dQ}} = \frac{f}{\mathbb{E}_{\mathrm{Q}}[f]} \frac{\mathrm{d}P}{\mathrm{d}\mathcal{Q}}. \tag{C.2}$$

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

$$H(P \,|\, \mathrm{Q}) = H(P \,|\, \mathcal{Q}) - \int_{\mathbb{R}^n} p(x)\, |x|^2 \,\mathrm{d}x - \log\Big(\int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2 - 2\Psi(x)} \,\mathrm{d}x \Big), \tag{C.3}$$

which is justified by (**??**) 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 (**??**) takes values in the interval $(-\infty,\infty]$. Summing up, we can define well the relative entropy $H(P \,|\, \mathrm{Q})$ as in (**??**) provided that $P \in \mathscr{P}_2(\mathbb{R}^n)$, even when the σ-finite measure Q has infinite total mass.

Remark C.1. Wherever in this paper the relative entropy $H(P \,|\, \mathrm{Q})$ is considered for some σ-finite measure Q on the Borel sets of $\mathbb{R}^n$ with density $q(x) = \mathrm{e}^{-2\Psi(x)}$, $x \in \mathbb{R}^n$, the probability measure P will always be assumed to belong to $\mathscr{P}_2(\mathbb{R}^n)$. This is in accordance with Lemmas **??** and **??**, as well as Lemma **??**. 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 $(-\infty,\infty]$. ⋄

D A measure-theoretic result

In the proofs of Propositions **??** and **??** we have used a result on conditional expectations, which we will formulate and prove below. 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 state the following result, which is known as *Scheffé's lemma* [**?**, 5.10].

Lemma D.1 (Scheffé's lemma)**.** *For a sequence of integrable random variables $(X_n)_{n \in \mathbb{N}}$ which converges almost surely to another integrable random variable X, convergence with respect to the norm of $L^1(\mathbb{P})$ (i.e., $\lim_{n\to\infty} \mathbb{E}[|X_n - X|] = 0$) is equivalent to convergence of the $L^1(\mathbb{P})$-norms (i.e., $\lim_{n\to\infty} \mathbb{E}[|X_n|] = \mathbb{E}[|X|]$).*

Proposition D.2. *Let $(B(t))_{0 \leqslant t \leqslant T}$ and $(C(t))_{0 \leqslant t \leqslant T}$ be adapted continuous processes, which are non-negative and uniformly bounded, respectively. Define the process $(A(t))_{0 \leqslant t \leqslant T}$ as their primitive, i.e.,*

$$A(t) = \int_0^t \big(B(u) + C(u)\big) \,\mathrm{d}u, \qquad 0 \leqslant t \leqslant T \tag{D.1}$$

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

$$\lim_{t \to t_0} \mathbb{E}\left[\frac{A(t) - A(t_0)}{t - t_0}\right] = \lim_{t \to t_0} \mathbb{E}\left[\frac{1}{t - t_0} \int_{t_0}^t \big(B(u) + C(u)\big) \,\mathrm{d}u\right] = \mathbb{E}\big[B(t_0) + C(t_0)\big]. \tag{D.2}$$

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

$$\lim_{t \uparrow t_0} \frac{\mathbb{E}\big[A(t_0) - A(t) \mid \mathcal{F}(t)\big]}{t_0 - t} = \lim_{t \downarrow t_0} \frac{\mathbb{E}\big[A(t) - A(t_0) \mid \mathcal{F}(t_0)\big]}{t - t_0} = B(t_0) + C(t_0), \tag{D.3}$$

where the limits exist in the norm of $L^1(\mathbb{P})$.

Proof. Using the uniform boundedness of the process $(C(t))_{0\leqslant t\leqslant T}$, it is easy to see that the existence of the limit in (**??**) only depends on the process $(B(t))_{0\leqslant t\leqslant T}$. Therefore we can assume without loss of generality that $C(t)\equiv 0$ for all $0\leqslant t\leqslant T$.

Fix a Lebesgue point $t_0\in[0,T]$ for which (**??**) does hold. As the process $(B(t))_{0\leqslant t\leqslant T}$ is continuous, the fundamental theorem of calculus ensures that the limit

$$\lim_{t\to t_0}\frac{A(t)-A(t_0)}{t-t_0}=\lim_{t\to t_0}\frac{1}{t-t_0}\int_{t_0}^{t}B(u)\,\mathrm{d}u=B(t_0) \tag{D.4}$$

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

$$\lim_{t\to t_0}\left\|\frac{A(t)-A(t_0)}{t-t_0}-B(t_0)\right\|_{L^1(\mathbb{P})}=0. \tag{D.5}$$

From (**??**) we can deduce now the $L^1(\mathbb{P})$-convergence of (**??**) as follows. Regarding the second limit in (**??**), for $t>t_0$, we find

$$\left\|\frac{\mathbb{E}\big[A(t)-A(t_0)\mid\mathcal{F}(t_0)\big]}{t-t_0}-B(t_0)\right\|_{L^1(\mathbb{P})}=\left\|\mathbb{E}\Big[\frac{A(t)-A(t_0)}{t-t_0}-B(t_0)\;\Big|\;\mathcal{F}(t_0)\Big]\right\|_{L^1(\mathbb{P})} \tag{D.6}$$

$$\leqslant\left\|\frac{A(t)-A(t_0)}{t-t_0}-B(t_0)\right\|_{L^1(\mathbb{P})}; \tag{D.7}$$

and according to (**??**), the expression in (**??**) converges to zero as $t\downarrow t_0$. Similarly, to handle the first limit in (**??**), we use for $t<t_0$ the estimate

$$\left\|\frac{\mathbb{E}\big[A(t_0)-A(t)\mid\mathcal{F}(t)\big]}{t_0-t}-B(t_0)\right\|_{L^1(\mathbb{P})}\leqslant\left\|\frac{A(t_0)-A(t)}{t_0-t}-B(t_0)\right\|_{L^1(\mathbb{P})} \tag{D.8}$$

$$+\left\|\mathbb{E}\big[B(t_0)\,|\,\mathcal{F}(t)\big]-B(t_0)\right\|_{L^1(\mathbb{P})}. \tag{D.9}$$

As $t\uparrow t_0$, the expression on the right-hand side of (**??**) converges to zero as before. The same is true also for the term in (**??**), on account of [**?**, Theorem 9.4.8] and the left-continuity of the filtration $(\mathcal{F}(t))_{t\geqslant 0}$. This completes the proof of Proposition **??**. □

E The proof of the Fontbona-Jourdain theorem

Proof of Theorem **??** [**?**]. For $0\leqslant s\leqslant T$, we define the random variable $N(T-s)$ as the conditional expectation of the random variable

$$\ell(0,X(0))=\frac{p(0,X(0))}{q(X(0))}\in L^1(\mathbb{Q}) \tag{E.1}$$

with respect to the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$, i.e.,

$$N(T-s):=\mathbb{E}_{\mathbb{Q}}\Big[\ell(0,X(0))\;\mid\;\mathcal{G}(T-s)\Big],\qquad 0\leqslant s\leqslant T. \tag{E.2}$$

Obviously the process $(N(T-s))_{0\leqslant s\leqslant T}$ is a martingale of the backwards filtration $(\mathcal{G}(T-s))_{0\leqslant s\leqslant T}$ under the probability measure $\mathbb{Q}$. Now we make the following elementary, but crucial, observation: as

the stochastic process $(X(t))_{0\leqslant t\leqslant T}$, which solves the stochastic differential equation (**??**), is a *Markov process*, the time-reversed process $(X(T-s))_{0\leqslant s\leqslant T}$ is a Markov process, too, under the probability measure $\mathbb{P}$ as well as under $\mathbb{Q}$. Hence

$$N(T-s) = \mathbb{E}_{\mathbb{Q}}\Big[\ell(0, X(0)) \mid X(T-s)\Big], \qquad 0 \leqslant s \leqslant T. \tag{E.3}$$

We have to show that this last conditional expectation equals $\ell(T-s, X(T-s))$. To this end, we fix $s \in [0,T]$ as well as a Borel set $A \subseteq \mathbb{R}^n$, and denote by $\pi(T-s; x, A)$ the transition probability of the event $\{X(T-s) \in A\}$, conditionally on $X(0) = x$. Note that this transition probability does not depend on whether we consider the process $(X(t))_{0\leqslant t\leqslant T}$ under $\mathbb{P}$ or under $\mathbb{Q}$. Then we find

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{p(0, X(0))}{q(X(0))}\ \mathbb{1}_A(X(T-s))\right] = \int_{\mathbb{R}^n} \frac{p(0,x)}{q(x)}\, \pi(T-s; x, A)\, q(x)\, \mathrm{d}x = P(T-s)[A]. \tag{E.4}$$

Note also that

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{p\big(T-s, X(T-s)\big)}{q\big(X(T-s)\big)}\ \mathbb{1}_A\big(X(T-s)\big)\right] = P(T-s)[A]. \tag{E.5}$$

Because the Borel set $A \subseteq \mathbb{R}^n$ is arbitrary, we deduce from (**??**) and (**??**) that

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{p(0, X(0))}{q(X(0))}\ \Big|\ X(T-s)\right] = \frac{p\big(T-s, X(T-s)\big)}{q\big(X(T-s)\big)} = \ell\big(T-s, X(T-s)\big). \tag{E.6}$$

This completes the proof of Theorem **??**. □

F The proof of Lemma ??

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

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

which is well-defined P_t-almost everywhere, 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.,

$$T_t^\gamma(x) = x + \int_0^t v_\theta(T_\theta^\gamma(x))\, \mathrm{d}\theta, \tag{F.2}$$

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

$$\partial_t p_t(x) + \operatorname{div}\big(v_t(x)\, p_t(x)\big) = 0, \qquad (t,x) \in (0,1) \times \mathbb{R}^n, \tag{F.3}$$

which can be written equivalently as

$$-\,\partial_t p_t(x) = \operatorname{div}\big(v_t(x)\big)\, p_t(x) + \big\langle v_t(x)\,, \nabla p_t(x)\big\rangle_{\mathbb{R}^n}, \qquad (t,x) \in (0,1) \times \mathbb{R}^n. \tag{F.4}$$

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

$$X_t = X_0 + \int_0^t v_\theta(X_\theta)\, \mathrm{d}\theta, \qquad 0 \leqslant t \leqslant 1 \tag{F.5}$$

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

$$\mathrm{d}p_t(X_t) = \partial_t p_t(X_t)\, \mathrm{d}t + \big\langle \nabla p_t(X_t)\,,\, \mathrm{d}X_t\big\rangle_{\mathbb{R}^n} = -p_t(X_t) \operatorname{div}\big(v_t(X_t)\big)\, \mathrm{d}t \tag{F.6}$$

on account of (**??**), (**??**), thus also

$$\mathrm{d}\log p_t(X_t) = -\operatorname{div}\big(v_t(X_t)\big)\,\mathrm{d}t, \qquad 0 \leqslant t \leqslant 1. \tag{F.7}$$

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

$$\mathrm{d}\log q(X_t) = -\big\langle 2\,\nabla\Psi(X_t)\,,\,\mathrm{d}X_t\big\rangle_{\mathbb{R}^n} = -\big\langle 2\,\nabla\Psi(X_t)\,,\,v_t(X_t)\big\rangle_{\mathbb{R}^n}\,\mathrm{d}t. \tag{F.8}$$

For the likelihood ratio function $\ell_t(\,\cdot\,)$ of (**??**) we get from (**??**) and (**??**) that

$$\mathrm{d}\log \ell_t(X_t) = \big\langle 2\,\nabla\Psi(X_t)\,,\,v_t(X_t)\big\rangle_{\mathbb{R}^n}\,\mathrm{d}t \,-\, \operatorname{div}\big(v_t(X_t)\big)\,\mathrm{d}t, \qquad 0 \leqslant t \leqslant 1. \tag{F.9}$$

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

$$H(P_t \,|\, \mathrm{Q}) - H(P_0 \,|\, \mathrm{Q}) = \mathbb{E}_\nu\big[\log \ell_t(X_t)\big] - \mathbb{E}_\nu\big[\log \ell_0(X_0)\big] \tag{F.10}$$

is equal to

$$\mathbb{E}_\nu\bigg[\int_0^t \Big(\big\langle 2\,\nabla\Psi(X_\theta)\,,\,v_\theta(X_\theta)\big\rangle_{\mathbb{R}^n} - \operatorname{div}\big(v_\theta(X_\theta)\big)\Big)\,\mathrm{d}\theta\bigg] \tag{F.11}$$

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

$$\lim_{t\downarrow 0} \frac{H(P_t \,|\, \mathrm{Q}) - H(P_0 \,|\, \mathrm{Q})}{t} = \mathbb{E}_\nu\Big[\big\langle 2\,\nabla\Psi(X_0)\,,\,v_0(X_0)\big\rangle_{\mathbb{R}^n} - \operatorname{div}\big(v_0(X_0)\big)\Big]. \tag{F.12}$$

Integrating by parts, we see that

$$\mathbb{E}_\nu\big[\operatorname{div}\big(v_0(X_0)\big)\big] = \int_{\mathbb{R}^n} \operatorname{div}\big(v_0(x)\big)\,p_0(x)\,\mathrm{d}x = -\int_{\mathbb{R}^n} \big\langle v_0(x)\,,\nabla p_0(x)\big\rangle_{\mathbb{R}^n}\,\mathrm{d}x \tag{F.13}$$

$$= -\big\langle \nabla \log p_0(X_0)\,,\,v_0(X_0)\big\rangle_{L^2(\nu)}. \tag{F.14}$$

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

$$\lim_{t\downarrow 0} \frac{H(P_t \,|\, \mathrm{Q}) - H(P_0 \,|\, \mathrm{Q})}{t} = \big\langle \nabla \log \ell_0(X_0)\,,\,v_0(X_0)\big\rangle_{L^2(\nu)}. \tag{F.15}$$

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

G Time reversal of diffusions

We review in the present section the theory of time reversal for diffusion processes developed by Föllmer [**?**, **?**], Haussmann and Pardoux [**?**], and Pardoux [**?**]. This section can be read independently of the rest of the paper; it does not present novel results.

G.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., [**?**, 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., [**?**, p. 330]), and neither is the semimartingale property (e.g., [**?**]). 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

$$\widehat{X}(s) := X(T - s), \qquad 0 \leqslant s \leqslant T, \tag{G.1}$$

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 [**?**, **?**, **?**], Schrödinger [**?**, **?**] and Kolmogorov [**?**]; they were dealt with systematically by Nelson [**?**] (see also Carlen [**?**]) 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 [**?**], and Pardoux [**?**]. Very interesting related results in a non-Markovian context, but with dispersion structure given by the identity matrix, have been obtained by Föllmer [**?**, **?**]. Let us refer also to the papers [**?**, **?**] dealing with time reversal of Markov processes. In what follows, this theory is presented in the spirit of the expository paper by Meyer [**?**].

G.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

$$X_i(t) = \xi_i + \int_0^t b_i(\theta, X(\theta))\, \mathrm{d}\theta + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta))\, \mathrm{d}W_\nu(\theta), \qquad 0 \leqslant t \leqslant T, \tag{G.2}$$

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 (**??**)), which implies that we have also

$$\mathcal{F}(t) = \sigma\big(X(\theta), W(\theta) \colon\ 0 \leqslant \theta \leqslant t\big) = \sigma\big(X(0), W(t) - W(\theta) \colon\ 0 \leqslant \theta \leqslant t\big) \tag{G.3}$$

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 $\mathrm{s}_{i\nu}(\,\cdot\,,\,\cdot\,)$, $i = 1, \ldots, n$, $\nu = 1, \ldots, d$. We shall denote the $(i,j)^{\text{th}}$ entry of the covariance matrix $\mathrm{a}(t,x) := \mathrm{s}(t,x)\, \mathrm{s}'(t,x)$ by

$$\mathrm{a}_{ij}(t,x) := \sum_{\nu=1}^{d} \mathrm{s}_{i\nu}(t,x)\, \mathrm{s}_{j\nu}(t,x), \qquad 1 \leqslant i, j \leqslant n.$$

These 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

$$\mathbb{P}\big[X(t) \in A\big] = \int_A p(t,x)\, \mathrm{d}x, \qquad A \in \mathscr{B}(\mathbb{R}^n),$$

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 [**?**],

[**?**] for this result, as well as [**?**] for a very simple argument in the one-dimensional case ($n = d = 1$), and to the case of Langevin type equation (**??**) for arbitrary $n \in \mathbb{N}$. We refer to [**?**], [**?**] or [**?**] for the basics of the theory of stochastic equations of the form (**??**).

The probability density function $p(t, \cdot)\colon \mathbb{R}^n \to (0, \infty)$ solves the forward Kolmogorov [**?**] equation [**?**, p. 149]

$$\partial_t p(t,x) = \frac{1}{2} \sum_{i,j=1}^{n} D_{ij}^2\big(\mathrm{a}_{ij}(t,x)\, p(t,x)\big) - \sum_{i=1}^{n} D_i\big(b_i(t,x)\, p(t,x)\big), \qquad (t,x) \in (0,T] \times \mathbb{R}^n. \tag{G.4}$$

If the drift and dispersion characteristics do not depend on time, and an invariant probability measure exists for the diffusion process of (**??**), the density function $p(\,\cdot\,)$ of this measure solves the stationary version of this forward Kolmogorov equation, to wit

$$\frac{1}{2} \sum_{i,j=1}^{n} D_{ij}^2\big(\mathrm{a}_{ij}(x)\, p(x)\big) = \sum_{i=1}^{n} D_i\big(b_i(x)\, p(x)\big), \qquad x \in \mathbb{R}^n. \tag{G.5}$$

G.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

$$\widehat{\mathcal{F}}(t) := \sigma\big(X(\theta), W(\theta) - W(t)\colon\ t \leqslant \theta \leqslant T\big), \qquad 0 \leqslant t \leqslant T. \tag{G.6}$$

It is not hard to see that the σ-algebra in (**??**) is expressed equivalently as

$$\begin{aligned} \widehat{\mathcal{F}}(t) &= \sigma\big(X(t), W(\theta) - W(t)\colon\ t \leqslant \theta \leqslant T\big) = \sigma\big(X(t), W(\theta) - W(T)\colon\ t \leqslant \theta \leqslant T\big) \\ &= \sigma\big(X(T), W(\theta) - W(t)\colon\ t \leqslant \theta \leqslant T\big) = \sigma\big(X(T)\big) \vee \mathcal{H}(t). \end{aligned} \tag{G.7}$$

Here, the σ-algebra generated by the Brownian increments after time t, namely,

$$\mathcal{H}(t) := \sigma\big(W(\theta) - W(t)\colon\ t \leqslant \theta \leqslant T\big), \qquad 0 \leqslant t \leqslant T, \tag{G.8}$$

is independent of the random vector $X(t)$. The time-reversed processes $\widehat{X}$ as in (**??**), as well as

$$\widetilde{W}(s) := W(T-s) - W(T), \qquad 0 \leqslant s \leqslant T, \tag{G.9}$$

are both adapted to the *backwards filtration* $\widehat{\mathbb{F}} := \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T}$, where

$$\begin{aligned} \widehat{\mathcal{F}}(T-s) &= \sigma\big(X(T-u), W(T-u) - W(T-s)\colon\ 0 \leqslant u \leqslant s\big) \\ &= \sigma\big(\widehat{X}(u), \widetilde{W}(u) - \widetilde{W}(s)\colon\ 0 \leqslant u \leqslant s\big) \end{aligned} \tag{G.10}$$

from (**??**). Note that, by complete analogy with (**??**), we have also

$$\widehat{\mathcal{F}}(T-s) = \sigma\big(X(T), W(T-u) - W(T-s)\colon\ 0 \leqslant u \leqslant s\big) = \sigma\big(\widehat{X}(0)\big) \vee \mathcal{H}(T-s) \tag{G.11}$$

on account of (**??**), where

$$\mathcal{H}(T-s) = \sigma\big(W(T-u) - W(T-s)\colon\ 0 \leqslant u \leqslant s\big) = \sigma\big(\widetilde{W}(u) - \widetilde{W}(s)\colon\ 0 \leqslant u \leqslant s\big). \tag{G.12}$$

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 $[0, s]$; see the expressions right above. Furthermore, the σ-algebra $\widehat{\mathcal{F}}(T-s)$ measures all the random variables $\widehat{X}(u)$, $u \in [0, s]$.

Remark G.1. In fact, the time-reversed process $\widetilde{W}$ *is a Brownian motion of the backwards filtration* $\mathbb{H} := (\mathcal{H}(T-s))_{0 \leqslant s \leqslant T} \subseteq \widehat{\mathbb{F}}$ as in (**??**), 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 [**?**, Theorem 5.1]). In the next subsection we shall see that the process $\widetilde{W}$ *is only a semimartingale of the larger backwards filtration* $\widehat{\mathbb{F}} = \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T}$, and identify its semimartingale decomposition. ◇

G.4 Some remarkable Brownian motions

Following the exposition and ideas in [**?**], we start with a couple of observations. First, for every $t \in [0,T]$ and every integrable, $\widehat{\mathcal{F}}(t)$-measurable random variable $\mathcal{K}$, we have

$$\mathbb{E}\big[\mathcal{K} \,|\, \mathcal{F}(t)\big] = \mathbb{E}\big[\mathcal{K} \,|\, X(t)\big], \quad \text{almost surely.} \tag{G.13}$$

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

$$G(X(t)) - g(\theta, X(\theta)) = g(t, X(t)) - g(\theta, X(\theta)) = \sum_{i=1}^n \sum_{\nu=1}^d \int_\theta^t D_i g(v, X(v))\, \mathrm{s}_{i\nu}(v, X(v))\, \mathrm{d}W_\nu(v).$$

For every index $\nu = 1, \dots, d$ this gives, after integrating by parts,

$$\begin{aligned}
&\mathbb{E}\big[(W_\nu(t) - W_\nu(\theta)) \cdot G(X(t))\big] = \mathbb{E}\Big[(W_\nu(t) - W_\nu(\theta)) \cdot \Big(g(t,X(t)) - g(\theta, X(\theta))\Big)\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^n \int_\theta^t D_i g(v, X(v))\, \mathrm{s}_{i\nu}(v, X(v))\, \mathrm{d}v\Big] = \sum_{i=1}^n \int_\theta^t \int_{\mathbb{R}^n} (D_i g \cdot \mathrm{s}_{i\nu})(v,x)\, p(v,x)\, \mathrm{d}x\, \mathrm{d}v \\
&= -\sum_{i=1}^n \int_\theta^t \int_{\mathbb{R}^n} g(v,x)\, D_i\big(p(v,x)\, \mathrm{s}_{i\nu}(v,x)\big)\, \mathrm{d}x\, \mathrm{d}v = -\int_\theta^t \int_{\mathbb{R}^n} g(v,x)\ \mathrm{div}\,\big(p(v,x)\, \overline{\mathrm{s}}_\nu(v,x)\big)\, \mathrm{d}x\, \mathrm{d}v \\
&= -\int_\theta^t \mathbb{E}\Big[g(v, X(v)) \cdot \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(v, X(v))\Big]\, \mathrm{d}v = -\mathbb{E}\Big[G(X(t)) \cdot \int_\theta^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v\Big].
\end{aligned}$$

Here $\overline{\mathrm{s}}_\nu(v, \cdot\,)$ is the ν^{th} 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

$$\mathbb{E}\Big[G(X(t)) \cdot \Big(W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v\Big)\Big] = 0 \tag{G.14}$$

for every $G \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and thus by extension for every bounded, measurable $G\colon \mathbb{R}^n \to \mathbb{R}$.

Theorem G.2. *The vector process $B = (B_1, \dots, B_d)'$ defined as*

$$B_\nu(s) := \widetilde{W}_\nu(s) - \int_0^s \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(T-u, \widehat{X}(u))\, \mathrm{d}u \tag{G.15}$$

$$= W_\nu(T-s) - W_\nu(T) - \int_{T-s}^T \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v, \qquad 0 \leqslant s \leqslant T, \tag{G.16}$$

for $\nu = 1, \dots, d$, is a Brownian motion of the backwards filtration $\widehat{\mathbb{F}} = \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T}$.

Remark G.3. 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,

$$\big\{B(T-\theta) - B(T-t)\colon\ 0 \leqslant \theta \leqslant t\big\} \quad \text{ is independent of } \quad \widehat{\mathcal{F}}(t) \supseteq \sigma\big(X(v)\colon\ t \leqslant v \leqslant T\big).$$

Note also from (**??**) that

$$B_\nu(T-\theta) - B_\nu(T-t) = W_\nu(\theta) - W_\nu(t) - \int_\theta^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}(v, X(v))\, \mathrm{d}v, \qquad 0 \leqslant \theta \leqslant t. \qquad \diamond$$

Reversing time once again, we obtain the following corollary of Theorem **??**.

Corollary G.4. *The $\mathbb{F}$-adapted vector process $V = (V_1, \dots, V_d)'$ with components*

$$V_\nu(t) := B_\nu(T-t) - B_\nu(T) = W_\nu(t) + \int_0^t \frac{\operatorname{div}(p\,\overline{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v, \qquad 0 \leqslant t \leqslant T, \tag{G.17}$$

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

$$\mathcal{F}^V(t) := \sigma(V(\theta)\colon\ 0 \leqslant \theta \leqslant t) \tag{G.18}$$

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 **??**. It suffices to show that each component process B_ν is a martingale of the backwards filtration $\widehat{\mathbb{F}}$; 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 $\mathbb{E}\big[\big(B_\nu(T-\theta) - B_\nu(T-t)\big)\cdot\mathcal{K}\big] = 0$ for $0 \leqslant \theta \leqslant t \leqslant T$ and every bounded, $\widehat{\mathcal{F}}(t)$-measurable $\mathcal{K}$; equivalently,

$$\mathbb{E}\bigg[\mathbb{E}[\mathcal{K}\,|\,\mathcal{F}(t)]\cdot\bigg(W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\operatorname{div}(p\,\overline{\mathrm{s}}_\nu)}{p}(v, X(v))\,\mathrm{d}v\bigg)\bigg] = 0,$$

as the expression inside the curved braces is $\mathcal{F}(t)$-measurable. But recalling (**??**) 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 (**??**). □

G.5 The diffusion property under time reversal

Let us return now to the question, whether the time-reversed process $\widehat{X}$ of (**??**), (**??**) is a diffusion. We start by expressing X_i of (**??**) in terms of a backwards Itô integral (see Subsection **??**) as

$$\begin{aligned} X_i(t) - \xi_i - \int_0^t b_i(\theta, X(\theta))\,\mathrm{d}\theta &= \sum_{\nu=1}^d \int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta))\,\mathrm{d}W_\nu(\theta) \\ &= \sum_{\nu=1}^d \bigg(\int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta) - \langle \mathrm{s}_{i\nu}(\,\cdot\,, X), W_\nu\rangle(t)\bigg). \end{aligned}$$

From (**??**), we have by Itô's formula that the process

$$\mathrm{s}_{i\nu}(\,\cdot\,, X) - \mathrm{s}_{i\nu}(0, \xi) - \sum_{j=1}^n \sum_{\kappa=1}^d \int_0^\cdot D_j \mathrm{s}_{i\nu}(\theta, X(\theta)) \cdot \mathrm{s}_{j\kappa}(\theta, X(\theta))\,\mathrm{d}W_\kappa(\theta)$$

is of finite first variation, therefore $\langle \mathrm{s}_{i\nu}(\,\cdot\,, X), W_\nu\rangle(t) = \sum_{j=1}^n \int_0^t \mathrm{s}_{j\nu}(\theta, X(\theta))\, D_j \mathrm{s}_{i\nu}(\theta, X(\theta))\,\mathrm{d}\theta$. We conclude

$$X_i(t) = \xi_i - \int_0^t \bigg(\sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i\bigg)(\theta, X(\theta))\,\mathrm{d}\theta + \sum_{\nu=1}^d \int_0^t \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta).$$

Evaluating also at $t = T$, then subtracting, we obtain

$$X_i(t) = X_i(T) + \int_t^T \Big(\sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i \Big)(\theta, X(\theta))\, \mathrm{d}\theta - \sum_{\nu=1}^d \int_t^T \mathrm{s}_{i\nu}(\theta, X(\theta)) \bullet \mathrm{d}W_\nu(\theta),$$

as well as

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \Big(\sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i \Big)(T-u, \widehat{X}(u))\, \mathrm{d}u + \sum_{\nu=1}^d \int_0^s \mathrm{s}_{i\nu}(T-u, \widehat{X}(u))\, \mathrm{d}\widetilde{W}_\nu(u)$$

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 (**??**).

But now let us recall (**??**), on the strength of which the above expression takes the form

$$\begin{aligned} \widehat{X}_i(s) = \widehat{X}_i(0) &+ \sum_{\nu=1}^d \int_0^s \mathrm{s}_{i\nu}(T-u, \widehat{X}(u))\, \mathrm{d}B_\nu(u) \\ &+ \int_0^s \Big(\sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} + \sum_{\nu=1}^d \mathrm{s}_{i\nu}\, \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p} - b_i \Big)(T-u, \widehat{X}(u))\, \mathrm{d}u, \qquad 0 \leqslant s \leqslant T. \end{aligned}$$

But in conjunction with Theorem **??**, this means that the time-reversed process $\widehat{X}$ of (**??**), (**??**) is a semimartingale of the backwards filtration $\widehat{\mathbb{F}} = \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T}$, with decomposition

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \widehat{b}_i(T-u, \widehat{X}(u))\, \mathrm{d}u + \sum_{\nu=1}^d \int_0^s \mathrm{s}_{i\nu}(T-u, \widehat{X}(u))\, \mathrm{d}B_\nu(u) \tag{G.19}$$

for $0 \leqslant s \leqslant T$, where, for each $i = 1, \ldots, n$, the function $\widehat{b}_i(\,\cdot\,,\,\cdot\,)$ is specified by

$$\begin{aligned} \widehat{b}_i(t,x) + b_i(t,x) &= \sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}(t,x)\, D_j \mathrm{s}_{i\nu}(t,x) + \sum_{\nu=1}^d \mathrm{s}_{i\nu}(t,x)\, \frac{\mathrm{div}\,\big(p(t,x)\, \overline{\mathrm{s}}_\nu(t,x)\big)}{p(t,x)} \\ &= \sum_{j=1}^n \sum_{\nu=1}^d \mathrm{s}_{j\nu}(t,x)\, D_j \mathrm{s}_{i\nu}(t,x) + \sum_{\nu=1}^d \frac{\mathrm{s}_{i\nu}(t,x)}{p(t,x)} \Big(\sum_{j=1}^n D_j \big(p(t,x)\, \mathrm{s}_{j\nu}(t,x)\big) \Big) \\ &= \sum_{j=1}^n \big(D_j \mathrm{a}_{ij}(t,x) + \mathrm{a}_{ij}(t,x) \cdot D_j \log p(t,x) \big). \end{aligned}$$

Theorem G.5. *Under the assumptions of this section, the time-reversed process $\widehat{X}$ of* (**??**), (**??**) *is a diffusion in the backwards filtration $\widehat{\mathbb{F}} = \big(\widehat{\mathcal{F}}(T-s)\big)_{0 \leqslant s \leqslant T}$, with characteristics as in* (**??**), *namely, dispersions $\mathrm{s}_{i\nu}(T-s,x)$ and drifts $\widehat{b}_i(T-s,x)$ given by the* generalized Nelson equation

$$\widehat{b}_i(t,x) + b_i(t,x) = \sum_{j=1}^n \Big(D_j \mathrm{a}_{ij}(t,x) + \mathrm{a}_{ij}(t,x) \cdot D_j \log p(t,x) \Big), \qquad i = 1, \ldots, n. \tag{G.20}$$

Equivalently, and with $\mathrm{div}\,\big(\mathrm{a}(t,x)\big) := \big(\sum_{j=1}^n D_j \mathrm{a}_{ij}(t,x) \big)_{1 \leqslant i \leqslant n}$, we write

$$\widehat{b}(t,x) + b(t,x) = \mathrm{div}\big(\mathrm{a}(t,x)\big) + \mathrm{a}(t,x) \cdot \nabla \log p(t,x). \tag{G.21}$$

Remark G.6. This result can be extended to the case where the sums of the distributional derivatives $\sum_{j=1}^n D_j \big(\mathrm{a}_{ij}(t,x)\, p(t,x)\big)$, $i = 1, \ldots, n$, are only assumed to be locally integrable functions of $x \in \mathbb{R}^n$; see [**?**, **?**]. ◇

Remark G.7 (Some filtration comparisons). For an invertible dispersion matrix $\mathrm{s}(\,\cdot\,,\,\cdot\,)$, it follows from (**??**) that the Brownian motion B is adapted to the filtration generated by $\widehat{X}$; that is,

$$\mathcal{F}^B(s) \subseteq \mathcal{F}^{\widehat{X}}(s) := \sigma\big(\widehat{X}(u)\colon\ 0 \leqslant u \leqslant s\big) = \sigma\big(X(T-u)\colon\ 0 \leqslant u \leqslant s\big), \qquad 0 \leqslant s \leqslant T. \tag{G.22}$$

Now recall (**??**); in its light, the filtration comparison in (**??**) implies $\mathcal{F}^{\widetilde{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s)$, thus $\mathcal{H}(T-s) \subseteq \mathcal{F}^{\widetilde{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s)$ from (**??**), for $0 \leqslant s \leqslant T$, and from (**??**) also

$$\widehat{\mathcal{F}}(T-s) \subseteq \mathcal{F}^{\widehat{X}}(s), \qquad 0 \leqslant s \leqslant T. \tag{G.23}$$

But we have also the reverse inclusion $\mathcal{F}^{\widehat{X}}(s) \subseteq \widehat{\mathcal{F}}(T-s)$ on account of (**??**) and (**??**); 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 $\mathcal{G}(T-s) \equiv \mathcal{F}^{\widehat{X}}(s)$, $0 \leqslant s \leqslant T$, in (**??**). ◇

G.6 The backwards Itô integral

For two continuous semimartingales $X = X(0)+M+B$ and $Y = Y(0)+N+C$, with B, C continuous adapted processes of finite variation and M, N continuous local martingales, let us recall the definition of the Fisk-Stratonovich integral in [**?**, Definition 3.3.13, p. 156], as well as its properties in [**?**, Problem 3.3.14] and [**?**, Problem 3.3.15].

By analogy with this definition, we introduce the *backwards Itô integral*

$$\int_0^{\cdot} Y(t) \bullet \mathrm{d}X(t) := \int_0^{\cdot} Y(t)\,\mathrm{d}M(t) + \int_0^{\cdot} Y(t)\,\mathrm{d}B(t) + \langle M, N\rangle, \tag{G.24}$$

where the first (respectively, the second) integral on the right-hand side is to be interpreted in the Itô (respectively, the Lebesgue-Stieltjes) sense.

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

$$\sum_{j=0}^{m-1} Y(t_{j+1})\big(X(t_{j+1}) - X(t_j)\big) \tag{G.25}$$

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

$$f(X) = f\big(X(0)\big) + \sum_{i=1}^{n} \int_0^{\cdot} D_i f\big(X(t)\big) \bullet \mathrm{d}X_i(t) - \frac{1}{2} \sum_{i,j=1}^{n} \int_0^{\cdot} D^2_{ij} f\big(X(t)\big)\,\mathrm{d}\langle M_i, M_j\rangle(t), \tag{G.26}$$

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.