Applying Itô calculus to Otto calculus*

Ioannis Karatzas† Walter Schachermayer‡ Bertram Tschiderer§

21st November 2018

Abstract. We revisit the [**?**] variational characterization of diffusion as entropic gradient flow, and provide for it a probabilistic interpretation based on stochastic calculus. It was shown by Jordan, Kinderlehrer, and Otto in [**?**] that, for diffusions of Langevin type, the Fokker-Planck probability density flow minimizes the rate of entropy dissipation as measured by the distance traveled in terms of the 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 the backward directions of time, using a very direct perturbation analysis; the original results follow then simply by taking expectations. As a bonus, we derive a slightly improved version of the so-called HWI inequality relating relative entropy, Fisher information and Wasserstein distance.

1 Introduction

We give a trajectorial interpretation of a seminal result by Jordan, Kinderlehrer, and Otto [**?**], and provide a proof based on stochastic calculus. The basic theme of our approach is outlined epigrammatically in the title; more precisely, we follow a stochastic approach to Otto's characterization of diffusions of Langevin-Schmoluchowski type as entropic gradient flows in Wasserstein space. For consistency and better readability we adopt the setting and notation of [**?**], and even copy some paragraphs of this paper almost verbatim.

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

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

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

*We thank Mathias Beiglböck, David Kinderlehrer, Jan Maas, Chris Rogers, and Oleg Szehr for their advice and comments during the preparation of this paper. Special thanks go to Jan Maas for his guidance and expertise, which helped us navigate through several difficult narrows successfully. 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.

†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).

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-Schmoluchowski type [**?**, **?**, **?**, **?**]

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

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 distribution of X^0 has probability 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 $t > 0$ 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 [**?**, **?**, **?**]

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

for $x \in \mathbb{R}^n$. When it exists, the probability measure on $\mathbb{R}^n$ with density 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 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 enegery is given by the negative of the Gibbs-Boltzmann entropy functional

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

In accordance with [**?**] we consider the following regularity assumptions.

Assumptions 1.1 (Regularity assumptions of [**?**, Theorem 5.1]).

(i) The potential $\Psi\colon \mathbb{R}^n \to [0,\infty)$ is smooth and satisfies, for some $C \in (0,\infty)$, the bound

$$|\nabla\Psi| \leqslant C(\Psi + 1). \tag{1.10}$$

(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.,

$$\int_{\mathbb{R}^n} \rho^0(x)\,|x|^2\,\mathrm{d}x < \infty \qquad \text{and} \qquad F(\rho^0) < \infty. \tag{1.11}$$

These assumptions are not strong enough to ensure that the constant $Z(\beta)$ in (**??**) is finite, thereby allowing for cases in which the stationary density ρ_s does not exist. In fact, in [**?**] the authors point out explicitly that, even when the stationary density ρ_s is not defined, the free energy (**??**) of a density $\rho(t,x)$ satisfying the Fokker-Planck equation (**??**) with initial condition (**??**) may be defined, provided that $F(\rho^0)$ is finite.

In the present paper, however, we also impose the more restrictive assumption that the stationary density ρ_s actually defines a probability measure, i.e., $Z(\beta) < \infty$. We do believe that our methods can be adapted to cover also the case $Z(\beta) = \infty$, but this will need additional work.

For these reasons we place ourselves in the following setting.

Assumptions 1.2 (Regularity assumptions of the present paper). In addition to conditions **??** and **??** of Assumptions **??**, we also impose that:

(iii) The constant $Z(\beta)$ in (**??**) is finite, so that the invariant probability measure with density ρ_s exists. In addition, we suppose that Ψ is sufficiently well-behaved to guarantee that the solution of (**??**) with initial condition (**??**) is smooth in the space variable x, Lipschitz in the time variable t on each interval $[\varepsilon, T]$, and strictly positive, for each $\varepsilon, t, T > 0$. For example, by requiring that all derivatives of Ψ grow at most exponentially, as $|x|$ converges to infinity, one may adapt the arguments from [**?**] showing that this is indeed the case.

2 The stochastic approach

Thus far, we have been mostly quoting from [**?**]. We take 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

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

We shall study the stochastic differential equation (**??**) under two different initial distributions. We let $P(0)$ be a probability measure with density $p^0 := \rho^0$, and denote by $Q(0)$ the invariant probability measure on $\mathbb{R}^n$ with stationary density $q(0) := \rho_s$ as in (**??**).

While we make an effort to follow the setting and notation of [**?**] as closely as possible, our notation differs slightly from [**?**]. To conform with our more probabilistic approach, we shall use the letters $p(0)$ and $q(0)$ rather than ρ^0 and ρ_s.

The initial probability measures $P(0)$ and $Q(0)$ on $\mathbb{R}^n$, defined by the densities $p(0)$ and $q(0)$, induce probability measures $\mathbb{P}$ and $\mathbb{Q}$ on the path space $\Omega = \mathcal{C}(\mathbb{R}_+;\mathbb{R}^n)$ of $\mathbb{R}^n$-valued continuous functions on $\mathbb{R}_+ = [0,\infty)$, so that the canonical coordinate process $(X(t)(\omega))_{t\geqslant 0} \equiv (\omega(t))_{t\geqslant 0}$ satisfies the stochastic differential equation (**??**) with initial distribution $P(0)$ under $\mathbb{P}$, and $Q(0)$ under $\mathbb{Q}$. We shall denote by $P(t)$ and $Q(t)$ the distributions of the random vector $X(t)$ under the probability measures $\mathbb{P}$ and $\mathbb{Q}$, respectively, at each time $t \geqslant 0$; and by $p(t) \equiv p(t,\,\cdot\,)$, $q(t) \equiv q(t,\,\cdot\,)$ the respective probability density functions. Of course, $Q(t)$ does not depend on time and equals the invariant distribution $Q \equiv Q(0)$ with stationary density $q \equiv q(t)$ for all times $t \geqslant 0$.

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

$$\left.\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right|_{\sigma(X(t))} = \frac{\mathrm{d}P(t)}{\mathrm{d}Q}\big(X(t)\big) = \ell\big(t, X(t)\big), \qquad \text{where} \quad \ell(t,x) := \frac{p(t,x)}{q(x)} \tag{2.2}$$

for $t \geqslant 0$ and $x \in \mathbb{R}^n$.

The *relative entropy* of $P(t)$ with respect to Q is defined by

$$H\big(P(t)\,|\,Q\big) := \mathbb{E}_{\mathbb{P}}\big[\log \ell\big(t, X(t)\big)\big] = \int_{\mathbb{R}^n} \log\left(\frac{p(t,x)}{q(x)}\right) p(t,x)\,\mathrm{d}x, \qquad t \geqslant 0. \tag{2.3}$$

The evaluation of the free energy functional F in (**??**) for the probability density function $p(t,\,\cdot\,)$ can be interpreted as the relative entropy $H(P(t)\,|\,Q)$; the following well-known identity (**??**) spells this out. In light of condition **??** in Assumptions **??**, this identity implies $H(P(0)\,|\,Q) < \infty$, so the quantity in (**??**) is well-defined and finite for $t = 0$.

Lemma 2.1. *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*

$$2\,F\big(p(t,\,\cdot\,)\big) = H\big(P(t)\,|\,Q\big) - \log Z(2). \tag{2.4}$$

Proof. Indeed,

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\big[\,\log \ell(t, X(t))\big] &= \mathbb{E}_{\mathbb{P}}\Big[\,\log\Big(Z(2)\,\mathrm{e}^{2\Psi(X(t))}\,p(t, X(t))\Big)\Big] \\
&= \log Z(2) + \mathbb{E}_{\mathbb{P}}\big[2\Psi\big(X(t)\big)\big] + \mathbb{E}_{\mathbb{P}}\big[\,\log p\big(t, X(t)\big)\big] \\
&= \log Z(2) + 2\int_{\mathbb{R}^n} \Psi(x)\,p(t,x)\,\mathrm{d}x + \int_{\mathbb{R}^n} p(t,x)\,\log p(t,x)\,\mathrm{d}x,
\end{aligned}$$

which equals $2\,F(p(t,\,\cdot\,))$, up to the constant $\log Z(2)$. □

At this point we notice that the normalizing constant $Z(2)$ is irrelevant for the present problem of studying the *decay* of the free energy functional $F(p(t,\,\cdot\,))$. For notational convenience we therefore may and do assume throughout this paper that the constant $Z(2)$ in (**??**) is normalized to equal one.

3 The theorems

As already indicated in (**??**) and (**??**), the probability density function $p(t,\,\cdot\,)\colon \mathbb{R}^n \to (0,\infty)$ 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 stationary density $\rho_s(\,\cdot\,) = q(\,\cdot\,)$ 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 the light of Lemma **??**, the object of interest in [**?**] is to relate the decay of the relative entropy functional

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P\,|\,Q) \in \mathbb{R}_+ \tag{3.4}$$

along the curve $(P(t))_{t\geqslant 0}$, to the quadratic Wasserstein distance $W_2(\,\cdot\,,\,\cdot\,)$, defined in (**??**) in Section **??**. We resume the remarkable relation between these two quantities in the following two theorems.

Theorem 3.1. *Under the* Assumptions **??**, *for each $t_0 \geqslant 0$ we have*

$$\lim_{t\downarrow t_0} \frac{H\big(P(t)\,|\,Q\big) - H\big(P(t_0)\,|\,Q\big)}{W_2\big(P(t), P(t_0)\big)} = -\sqrt{I\big(P(t_0)\,|\,Q\big)} \tag{3.5}$$

as well as, for $t_0 > 0$,

$$\lim_{t\uparrow t_0} \frac{H\big(P(t)\,|\,Q\big) - H\big(P(t_0)\,|\,Q\big)}{W_2\big(P(t), P(t_0)\big)} = \sqrt{I\big(P(t_0)\,|\,Q\big)}. \tag{3.6}$$

The expression on the left-hand sides of (**??**) and (**??**) may be interpreted as the slope of the relative entropy functional $P \mapsto H(P \,|\, Q)$ at $P = P(t_0)$ along the curve $(P(t))_{t \geqslant 0}$, if we measure distances in $\mathscr{P}_2(\mathbb{R}^n)$ by the quadratic Wasserstein distance $W_2(\,\cdot\,,\,\cdot\,)$ of (**??**). The quantity appearing on the right-hand sides of (**??**) and (**??**) is the *relative Fisher information* (see, e.g., [**?**]), defined as

$$I(P(t_0) \,|\, Q) := \mathbb{E}_{\mathbb{P}}\Big[\big|\nabla \log \ell(t_0, X(t_0))\big|^2\Big] \tag{3.7}$$

and, written more explicitly in terms of the "score function" $\nabla \ell(t,\,\cdot\,)/\ell(t,\,\cdot\,)$, as

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

The remarkable insight of [**?**] states that the slope in (**??**) and (**??**) 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 $P(t_0)$.

To formalize this assertion, we fix $t_0 \geqslant 0$ as well as a compactly supported, and possibly time-dependent, vector field $\beta\colon [t_0, \infty) \times \mathbb{R}^n \to \mathbb{R}^n$ of class $\mathcal{C}^{1,\infty}$, which will serve as a perturbation. Consider the thus perturbed Fokker-Planck equation

$$\partial_t p^\beta(t,x) = \operatorname{div}\Big(\big(\nabla \Psi(x) + \beta(t,x)\big)\, p^\beta(t,x)\Big) + \tfrac{1}{2}\Delta p^\beta(t,x), \qquad (t,x) \in (t_0,\infty) \times \mathbb{R}^n, \tag{3.9}$$

with initial condition

$$p^\beta(t_0, x) = p(t_0, x), \qquad x \in \mathbb{R}^n. \tag{3.10}$$

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

$$\mathrm{d}X(t) = -\Big(\nabla \Psi\big(X(t)\big) + \beta\big(t, X(t)\big)\Big)\,\mathrm{d}t + \mathrm{d}W(t), \qquad t \geqslant t_0, \tag{3.11}$$

with initial distribution $P(t_0)$. The distribution of $X(t)$ under $\mathbb{P}^\beta$ on $\mathbb{R}^n$ will be denoted by $P^\beta(t)$; once again, the corresponding probability density function $p^\beta(t) \equiv p^\beta(t,\,\cdot\,)$ is a solution of the equation (**??**) subject to the initial condition (**??**).

Theorem 3.2. *Under the* Assumptions **??**, *we fix $t_0 \geqslant 0$ and let $\beta\colon [t_0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ be a gradient vector field, i.e., of the form $\beta(t,\,\cdot\,) = \nabla B(t,\,\cdot\,)$ for some time-dependent potential $B(t,\,\cdot\,)$, for $t \geqslant t_0$. Assume that β is compactly supported and of class $\mathcal{C}^{1,\infty}$, introduce the elements $a = \nabla \log \ell(t_0, X(t_0))$ and $b = \beta(t_0, X(t_0))$ of the Hilbert space $L^2(\mathbb{P}; \mathbb{R}^n)$, and suppose that $\|a + 2b\|_{L^2(\mathbb{P};\mathbb{R}^n)} > 0$. Then*

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) \,|\, Q) - H(P^\beta(t_0) \,|\, Q)}{W_2\big(P^\beta(t), P^\beta(t_0)\big)} = \lim_{t \downarrow t_0} \frac{H(P(t) \,|\, Q) - H(P(t_0) \,|\, Q)}{W_2\big(P(t), P(t_0)\big)} \tag{3.12}$$

$$+ \|a\|_{L^2(\mathbb{P};\mathbb{R}^n)} - \left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P};\mathbb{R}^n)}} \right\rangle_{L^2(\mathbb{P};\mathbb{R}^n)}. \tag{3.13}$$

Remark 3.3. On the strength of the Cauchy-Schwarz inequality, the expression (**??**) is non-negative, and vanishes if and only if a and b are collinear. Consequently, if the vector field $\beta(t_0,\,\cdot\,)$ is not a scalar multiple of $\nabla \log \ell(t_0,\,\cdot\,)$, the slope on the left-hand side of (**??**) is strictly bigger than the corresponding (negative) slope in (**??**), i.e., the right-hand side 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 invariant measure in (**??**) is standard Gaussian, the relation

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\, H(P(t) \,|\, Q) = -\tfrac{1}{2}\, I(P(t) \,|\, Q) \tag{3.14}$$

has been known since [**?**] as *de Bruijn's identity*; we revisit this identity in (**??**) below in our more general context, along the lines of the seminal work [**?**]. 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 excellent 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 [**?**] on the relation of (**??**) to the central limit theorem.

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 flow. This relation was further investigated by Otto in the paper [**?**], where the theory now known as *"Otto calculus"* was developed.

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

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

$$\log \ell(t, X(t)) = \log\left(\frac{p(t, X(t))}{q(X(t))}\right), \qquad 0 \leqslant t \leqslant T \tag{3.15}$$

along the trajectory $(X(t))_{0\leqslant t\leqslant T}$ and calculating its dynamics (stochastic differential) under the probability measures $\mathbb{P}$ and $\mathbb{Q}$. A decisive tool in this endeavor is to pass to reverse time, and to use a remarkable insight due to Fontbona and Jourdain [**?**]. These authors consider the coordinate process $(X(t))_{0\leqslant t\leqslant T}$ on path space $\Omega = \mathcal{C}([0,T];\mathbb{R}^n)$ in the *reverse direction* of time, i.e., they work with the time-reversed process $(X(T-t))_{0\leqslant t\leqslant T}$; 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 mention the relevant filtrations: We denote by $(\mathcal{F}(t))_{t\geqslant 0}$ the usual filtration generated by the coordinate process $(X(t))_{t\geqslant 0}$, that is,

$$\mathcal{F}(t) := \sigma\big(X(u)\colon\ 0 \leqslant u \leqslant t\big), \qquad t \geqslant 0; \tag{3.16}$$

while by $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$ we denote the filtration generated by the time-reversed coordinate process $(X(T-t))_{0\leqslant t\leqslant T}$, namely,

$$\mathcal{G}(T-t) := \sigma\big(X(T-u)\colon\ 0 \leqslant u \leqslant t\big), \qquad 0 \leqslant t \leqslant T. \tag{3.17}$$

As already mentioned, the following two theorems are the main new results of this paper. They can be regarded as trajectorial versions of Theorems **??** and **??**. The message 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 relative Fisher information, is valid not only in expectation, but also along (almost) each trajectory, provided we run time in the reverse direction.[1]

[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 in the discretization.

Theorem 3.4. *Under the assumptions of* Theorem **??**, *we define the Fisher information process* $(F(T-t))_{0\leqslant t\leqslant T}$ *accumulated from the right, as*

$$\begin{aligned} F(T-t) &:= \int_0^t \frac{\big|\nabla \ell(T-u, X(T-u))\big|^2}{\ell(T-u, X(T-u))^2}\,\mathrm{d}u \\ &= \int_0^t \bigg|\frac{\nabla p(T-u, X(T-u))}{p(T-u, X(T-u))} + 2\,\Psi\big(X(T-u)\big)\bigg|^2\,\mathrm{d}u \end{aligned} \tag{3.18}$$

for $t \in [0,T]$. *Then the difference*

$$M(T-t) := \log \ell(T-t, X(T-t)) - \tfrac{1}{2}\,F(T-t)\,, \qquad 0 \leqslant t \leqslant T \tag{3.19}$$

is a $\mathbb{P}$*-martingale with respect to the filtration* $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$. *More explicitly, at any given time* $t \in [0,T]$, *this martingale can be represented as*

$$M(T-t) = M(T) + \int_0^t \frac{\nabla \ell(T-u, X(T-u))}{\ell(T-u, X(T-u))}\,\mathrm{d}\overline{W}^{\mathbb{P}}(T-u), \tag{3.20}$$

where $\big(\overline{W}^{\mathbb{P}}(T-t)\big)_{0\leqslant t\leqslant T}$ *is a* $\mathbb{P}$*-Brownian motion with respect to the filtration* $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$.

This result implies Theorem **??**, as we argue presently; one simply has to take expectations with respect to $\mathbb{P}$. Indeed, passing from reversed time to the original time direction, Theorem **??** entails, for $0 \leqslant t, t_0 \leqslant T$,

$$\mathbb{E}_{\mathbb{P}}\big[\log \ell(t, X(t))\big] - \mathbb{E}_{\mathbb{P}}\big[\log \ell(t_0, X(t_0))\big] = -\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\int_{t_0}^t \frac{\big|\nabla \ell(u, X(u))\big|^2}{\ell(u, X(u))^2}\,\mathrm{d}u\bigg]. \tag{3.21}$$

In particular, this shows that the relative entropy function $t \mapsto H(P(t)\,|\,Q)$ from (**??**), and thus also the free energy function $t \mapsto F(p(t,\,\cdot\,))$ from (**??**), is strictly decreasing provided $\ell(t,\,\cdot\,)$ is not constant. Furthermore, equation (**??**) yields in the limit the *generalized de Bruijn identity*

$$\lim_{t\to t_0} \frac{H(P(t)\,|\,Q) - H(P(t_0)\,|\,Q)}{t - t_0} = -\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\,\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2}\,\bigg], \tag{3.22}$$

as well as

$$\lim_{t\to t_0} \frac{\big|H(P(t)\,|\,Q) - H(P(t_0)\,|\,Q)\big|}{|t - t_0|} = \tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}}\bigg[\,\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2}\,\bigg]. \tag{3.23}$$

On the other hand, as is carefully worked out in [**?**], we know the limiting behavior of the Wasserstein distance (see Theorem **??** in Section **??** below for the details), namely

$$\lim_{t\to t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\bigg[\,\frac{\big|\nabla \ell(t_0, X(t_0))\big|^2}{\ell(t_0, X(t_0))^2}\,\bigg]\right)^{1/2}. \tag{3.24}$$

Dividing the one-sided limits corresponding to (**??**) by the one-sided limits corresponding to (**??**) and using the definition of the relative Fisher information (**??**), as well as (**??**), we obtain equations (**??**) and (**??**) of Theorem **??** (the latter for $t_0 > 0$).

Summing up, we have deduced Theorem **??** from Theorem **??**. □

Next, we state also a trajectorial version of Theorem **??**. As above, we consider the perturbation β and denote the perturbed likelihood ratio function by

$$\ell^\beta(t,x) := \frac{p^\beta(t,x)}{q(x)}\,, \qquad (t,x) \in [t_0,\infty) \times \mathbb{R}^n. \tag{3.25}$$

Theorem 3.5. *Under the assumptions of* Theorem **??**, *for each* $t_0 \geqslant 0$ *we have*

$$\begin{aligned}\lim_{t \downarrow t_0} \frac{\mathbb{E}_{\mathbb{P}^\beta}\Big[\log \ell^\beta(t, X(t)) \mid \mathcal{F}(t_0)\Big] - \mathbb{E}_{\mathbb{P}}\Big[\log \ell(t, X(t)) \mid \mathcal{F}(t_0)\Big]}{t - t_0} \\ = \operatorname{div} \beta(t_0, X(t_0)) - 2\Big\langle \beta(t_0, X(t_0)), \nabla\Psi\big(X(t_0)\big)\Big\rangle_{L^2(\mathbb{P};\mathbb{R}^n)},\end{aligned} \tag{3.26}$$

the limit holding true $\mathbb{P}$*-almost surely and in the norm of* $L^1(\mathbb{P})$. *Furthermore,*

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\left|\frac{\nabla \ell(t_0, X(t_0))}{\ell(t_0, X(t_0))} + 2\,\beta(t_0, X(t_0))\right|^2\right]\right)^{1/2}. \tag{3.27}$$

Remark 3.6. In the statement of Theorem **??** above, the limit (**??**) also exists $\mathbb{P}^\beta$-almost surely and in the norm of $L^1(\mathbb{P}^\beta)$. Furthermore, the expectation $\mathbb{E}_{\mathbb{P}}$ appearing in (**??**) can be replaced by $\mathbb{E}_{\mathbb{P}^\beta}$. The reason is simply that $X(t_0)$ has the same distribution under $\mathbb{P}$, as it does under $\mathbb{P}^\beta$. ◇

Again, Theorem **??** implies Theorem **??** by taking expectations. Indeed, we can calculate the limits of the four terms appearing in the numerators and denominators in (**??**) explicitly, after normalizing by the factor $t - t_0$. Recalling the abbreviations $a = \nabla \log \ell(t_0, X(t_0))$ and $b = \beta(t_0, X(t_0))$, we claim that

$$\lim_{t \downarrow t_0} \frac{H\big(P(t) \,|\, Q\big) - H\big(P(t_0) \,|\, Q\big)}{t - t_0} = -\tfrac{1}{2}\,\|a\|^2_{L^2(\mathbb{P};\mathbb{R}^n)}, \tag{3.28}$$

$$\lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0} = \tfrac{1}{2}\,\|a\|_{L^2(\mathbb{P};\mathbb{R}^n)}, \tag{3.29}$$

$$\lim_{t \downarrow t_0} \frac{H\big(P^\beta(t) \,|\, Q\big) - H\big(P^\beta(t_0) \,|\, Q\big)}{t - t_0} = -\big\langle a, \tfrac{a}{2} + b\big\rangle_{L^2(\mathbb{P};\mathbb{R}^n)}, \tag{3.30}$$

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \tfrac{1}{2}\,\|a + 2b\|_{L^2(\mathbb{P};\mathbb{R}^n)}. \tag{3.31}$$

Subtracting the quotient of (**??**) and (**??**) from the quotient of (**??**) and (**??**), we arrive at the expression

$$\|a\|_{L^2(\mathbb{P};\mathbb{R}^n)} - \left\langle a\,,\, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P};\mathbb{R}^n)}}\right\rangle_{L^2(\mathbb{P};\mathbb{R}^n)}, \tag{3.32}$$

which is just (**??**).

We still have to verify the claims (**??**) – (**??**). The limits (**??**) and (**??**) are well-known [**?**] and follow from (**??**), as will be explained in Section **??**. As regards (**??**), we have already computed this limit in (**??**). We still have to show (**??**). Taking expectations in (**??**) yields

$$\begin{aligned}\lim_{t \downarrow t_0} \frac{\mathbb{E}_{\mathbb{P}^\beta}\Big[\log \ell^\beta(t, X(t))\Big] - \mathbb{E}_{\mathbb{P}}\Big[\log \ell(t, X(t))\Big]}{t - t_0} \\ = \mathbb{E}_{\mathbb{P}}\Big[\operatorname{div} \beta(t_0, X(t_0)) - 2\Big\langle \beta(t_0, X(t_0)), \nabla\Psi\big(X(t_0)\big)\Big\rangle\Big].\end{aligned} \tag{3.33}$$

The numerator of the left-hand side of (**??**) equals

$$\mathbb{E}_{\mathbb{P}^\beta}\left[\log\left(\frac{\ell^\beta(t, X(t))}{\ell^\beta(t_0, X(t_0))}\right)\right] - \mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{\ell(t, X(t))}{\ell(t_0, X(t_0))}\right)\right], \tag{3.34}$$

as $\ell^\beta(t_0, X(t_0)) = \ell(t_0, X(t_0))$, and the expression (**??**) is equal to

$$H(P^\beta(t)\,|\,Q) - H(P^\beta(t_0)\,|\,Q) - \Big(H(P(t)\,|\,Q) - H(P(t_0)\,|\,Q)\Big), \tag{3.35}$$

where we know already the asymptotics of the second half of the expression (**??**); namely, (**??**) once again. The first half contains what we want to calculate, namely (**??**). The right-hand side of (**??**) equals

$$\int_{\mathbb{R}^n} \Big(\operatorname{div}\beta(t_0,x) - 2\,\big\langle \beta(t_0,x), \nabla\Psi(x)\big\rangle_{\mathbb{R}^n}\Big)\, p(t_0,x)\,\mathrm{d}x. \tag{3.36}$$

Using integration by parts and the fact that the perturbation $\beta(t_0, \cdot\,)$ is assumed to be smooth and have compact support, this expression becomes

$$-\int_{\mathbb{R}^n} \Big\langle \beta(t_0,x), \nabla \log p(t_0,x) + 2\,\nabla\Psi(x)\Big\rangle_{\mathbb{R}^n}\, p(t_0,x)\,\mathrm{d}x, \tag{3.37}$$

which is the same as

$$-\Big\langle \beta(t_0,X(t_0)), \nabla\log \ell(t_0,X(t_0))\Big\rangle_{L^2(\mathbb{P};\mathbb{R}^n)} = -\langle b, a\rangle_{L^2(\mathbb{P};\mathbb{R}^n)}. \tag{3.38}$$

Combining (**??**), (**??**), (**??**) and (**??**), we obtain (**??**).

Summing up, we have proved that Theorem **??** implies Theorem **??**. □

Remark 3.7. In Theorems **??** and **??** we have required $\beta\colon [t_0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$ to be a gradient field, i.e., of the form $\beta(t,\,\cdot\,) = \nabla B(t,\,\cdot\,)$ for some time-dependent potential $B(t,\,\cdot\,)\colon \mathbb{R}^n \to \mathbb{R}$.

This assumption is crucial for the rate of change of the Wasserstein distance in (**??**) to be valid, as is well known [**?**] and will be recalled in Section **??** below. On the other hand, for the limiting behavior of the relative entropy in (**??**), this assumption plays no role. If $\beta(t,\,\cdot\,)$ is a (smooth and compactly supported) vector field which is not necessarily induced by a potential $B(t,\,\cdot\,)$, the assertion (**??**) is still valid as will become clear from the proof of Theorem **??** below. ◇

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

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P\,|\,Q) - H(P(0)\,|\,Q) \tag{3.39}$$

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

$$\mathscr{P}_2(\mathbb{R}^n) \ni P \longmapsto \langle a, c\rangle_{L^2(\mathbb{P};\mathbb{R}^n)}, \tag{3.40}$$

where the random variable c corresponds to $-\frac{a}{2} - b$ in (**??**). Now we fix a general element $P \in \mathscr{P}_2(\mathbb{R}^n)$ and let $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ be the optimal transport map from $P(0)$ to P. Then (**??**) suggests that the "displacement interpolation" $(P_t)_{0\leqslant t\leqslant 1}$ between $P_0 = P(0)$ 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

$$\gamma(x) = -\tfrac{1}{2}\,\nabla \log \ell(0,x) - \beta(x). \tag{3.41}$$

We formalize these intuitive geometric insights in the subsequent lemma, and place ourselves in the following setting.

Assumptions 3.8. In addition to Assumptions **??**, we impose that:

(iv) 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 $\gamma = \nabla\Gamma$ for some convex function $\Gamma\colon \mathbb{R}^n \to \mathbb{R}$, uniquely defined on and supported by the support of P_0, and smooth in the interior of this set. The map γ induces the optimal quadratic Wasserstein transport from P_0 to P_1 via

$$T_t^\gamma(x) := x + t\gamma(x) \qquad \text{and} \qquad (T_t^\gamma)_\#(P_0) =: P_t \tag{3.42}$$

for $0 \leqslant t \leqslant 1$; to wit, the displacement interpolation between P_0 and P_1.

Remark 3.9. For the existence and uniqueness of the optimal transport map $\gamma\colon \mathbb{R}^n \to \mathbb{R}^n$ we refer to [**?**, Theorem 2.44], and for its smoothness to [**?**, Theorem 4.14] as well as [**?**, Remarks 4.15]. ◇

Remark 3.10. We warn at this point, that we have chosen the subscript notation for P_t in order not to confuse it 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 **??**. Similarly, the likelihood ratio function

$$\ell_t(x) = \frac{p_t(x)}{q(x)}, \qquad (t,x) \in [0,1] \times \mathbb{R}^n, \tag{3.43}$$

is different from $\ell(t,\,\cdot\,)$, as now $p_t(\,\cdot\,)$ is the density function of the probability measure P_t. We relegate the proof of Lemma **??** below to Appendix **??**. ◇

Lemma 3.11. *Under the* Assumptions **??**, *recall the probability measure Q on $\mathbb{R}^n$ with density $q = \rho_s$ as in* (**??**), *and let X_0 be a random variable with distribution $P_0 = P(0)$, defined on some probability space $(S, \mathcal{S}, \nu)$. Then we have*

$$\lim_{t\downarrow 0} \frac{H(P_t \,|\, Q) - H(P_0 \,|\, Q)}{t} = \big\langle \nabla \log \ell_0(X_0), \gamma(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)}. \tag{3.44}$$

Combining Lemma **??** with well-known arguments, in particular, a fundamental result on displacement convexity due to McCann [**?**, **?**], we obtain an improvement of the *HWI inequality* obtained by Otto and Villani [**?**] relating the fundamental quantities of relative entropy (H), Wasserstein distance (W) and Fisher information (I).

Theorem 3.12 (HWI inequality)**.** *Under the* Assumptions **??**, *we let $P_0 = P(0)$ and Q be the probability measure on $\mathbb{R}^n$ with density $q = \rho_s$ as in* (**??**). *We suppose in addition that the potential $\Psi\colon \mathbb{R}^n \to [0,\infty)$ satisfies a curvature lower bound*

$$\mathrm{Hess}(\Psi) \geqslant \kappa\, \mathrm{Id}, \tag{3.45}$$

for some $\kappa \in \mathbb{R}$. Let $P_1 \in \mathscr{P}_2(\mathbb{R}^n)$ be such that $H(P_1 \,|\, Q) < \infty$, then we have

$$H(P_0 \,|\, Q) - H(P_1 \,|\, Q) \leqslant -\big\langle \nabla \log \ell_0(X_0), \gamma(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1), \tag{3.46}$$

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

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

$$-\big\langle \nabla \log \ell_0(X_0), \gamma(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)} \leqslant \|\nabla \log \ell_0(X_0)\|_{L^2(\nu;\mathbb{R}^n)}\, \|\gamma(X_0)\|_{L^2(\nu;\mathbb{R}^n)}, \tag{3.47}$$

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

$$I(P_0 \,|\, Q) = \mathbb{E}_\nu\Big[|\nabla \log \ell_0(X_0)|^2 \Big] = \|\nabla \log \ell_0(X_0)\|^2_{L^2(\nu;\mathbb{R}^n)}, \tag{3.48}$$

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

$$\|\gamma(X_0)\|_{L^2(\nu;\mathbb{R}^n)} = W_2(P_0, P_1). \tag{3.49}$$

Consequently, we get the inequality

$$-\big\langle \nabla \log \ell_0(X_0), \gamma(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)} \leqslant \sqrt{I(P_0 \,|\, Q)}\; W_2(P_0, P_1). \tag{3.50}$$

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

$$H(P_0 \,|\, Q) - H(P_1 \,|\, Q) \leqslant W_2(P_0, P_1)\, \sqrt{I(P_0 \,|\, Q)} - \tfrac{\kappa}{2}\, W_2^2(P_0, P_1). \tag{3.51}$$

When there is a non-trivial angle between $\nabla \log \ell_0(X_0)$ and $\gamma(X_0)$ in $L^2(\nu; \mathbb{R}^n)$, the inequality (**??**) gives a sharper bound than (**??**). We refer to the original paper [**?**], as well as to [**?**, Chapter 5], and the recent paper [**?**] for a detailed discussion of the HWI inequality (**??**), which contains as special cases Talagrand's inequality [**?**], as well as the logarithmic Sobolev inequality [**?**].

Proof of Theorem **??**. As elaborated in [**?**, Section 9.4] we may assume without loss of generality that P_0 and P_1 satisfy the assumptions of Lemma **??**. 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

$$f(t) := H(P_t \,|\, Q), \qquad 0 \leqslant t \leqslant 1. \tag{3.52}$$

The displacement convexity results of McCann [**?**], see also [**?**, Section 5.2], imply

$$f''(t) \geqslant \kappa\, W_2^2(P_0, P_1), \qquad 0 \leqslant t \leqslant 1. \tag{3.53}$$

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

$$f'(0+) = \lim_{t \downarrow 0} \frac{f(t) - f(0)}{t} = \big\langle \nabla \log \ell_0(X_0), \gamma(X_0) \big\rangle_{L^2(\nu; \mathbb{R}^n)}. \tag{3.54}$$

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

Remark 3.13. It is worth noting at this point that, in the hands of [**?**], the strong non-degeneracy condition (**??**) leads — via quite intricate and detailed analysis — to the exponential temporal dissipation of the Fisher information. For an exposition of the Bakry-Émery theory we refer to [**?**]. ◇

4 Details and proofs

In this section we provide the proofs of Theorems **??** and **??**. In fact, all we have to do is to apply Itô's formula to calculate the dynamics, i.e., the stochastic differentials of the relative entropy process

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

as well as those of the perturbed relative entropy process

$$\log \ell^\beta\big(t, X(t)\big) = \log \left(\frac{p^\beta\big(t, X(t)\big)}{q\big(X(t)\big)} \right), \qquad t \geqslant 0, \tag{4.2}$$

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

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

$$\log \ell\big(T-t, X(T-t)\big) = \log \left(\frac{p\big(T-t, X(T-t)\big)}{q\big(X(T-t)\big)} \right), \qquad 0 \leqslant t \leqslant T, \tag{4.3}$$

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

$$\begin{aligned} \mathrm{d}\log \ell\big(T-t, X(T-t)\big) = \frac{\nabla \ell\big(T-t, X(T-t)\big)}{\ell\big(T-t, X(T-t)\big)}\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-t) \\ + \frac{1}{2}\frac{\big|\nabla \ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\, \mathrm{d}t, \end{aligned} \tag{4.4}$$

for $0 \leqslant t \leqslant T$, with respect to the filtration $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$. As we have seen, the stochastic differential (**??**) of the process (**??**) yields a very direct and illuminating "trajectorial" sharpening of Theorem **??**. When deducing Theorem **??** from Theorem **??** we did not have to argue with partial integration. Taking expectations of the dynamics of (**??**) one can directly observe the trade-off between the decay of entropy and the traveled Wasserstein distance along each trajectory. We mention already here that partial integration appears to be unavoidable when working with the processes (**??**) and (**??**) in the forward direction.

The eye-opener (at least for the present authors) leading to (**??**) is the subsequent remarkable insight due to Fontbona and Jourdain [**?**]. It provided the present authors with much of the original motivation, to start this line of research. This theorem holds true 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 **??**.

Theorem 4.1 ([**?**])**.** *Under the* Assumptions **??**, *for any given $T > 0$, the time-reversed likelihood ratio process*

$$\ell\big(T-t, X(T-t)\big) = \frac{p\big(T-t, X(T-t)\big)}{q\big(X(T-t)\big)}\,, \qquad 0 \leqslant t \leqslant T, \tag{4.5}$$

is a $\mathbb{Q}$-martingale with respect to the reverse filtration $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$.

For the convenience of the reader we recall in Appendix **??** the surprisingly straightforward proof of Theorem **??**.

Our aim 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 process $(X(T-t))_{0\leqslant t\leqslant T}$ 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 of diffusion processes in Appendix **??**. The idea of time reversal goes back to the thoughts 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 we defined the probability measure $\mathbb{P}$ on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$ such that the canonical coordinate process $(X(t)(\omega))_{t\geqslant 0} \equiv (\omega(t))_{t\geqslant 0}$ satisfies the stochastic differential equation (**??**) with initial distribution $P(0)$ under $\mathbb{P}$. In other words, the process

$$W(t) := X(t) - X(0) + \int_0^t \nabla\Psi\big(X(u)\big)\, \mathrm{d}u, \qquad t \geqslant 0, \tag{4.6}$$

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

Proposition 4.2. *The process $\big(\overline{W}^{\mathbb{P}}(T-t)\big)_{0\leqslant t\leqslant T}$ defined by*

$$\overline{W}^{\mathbb{P}}(T-t) := W(T-t) - W(T) - \int_0^t \frac{\nabla p\big(T-u, X(T-u)\big)}{p\big(T-u, X(T-u)\big)}\, \mathrm{d}u, \qquad 0 \leqslant t \leqslant T, \tag{4.7}$$

is a Brownian motion under $\mathbb{P}$*, adapted to the filtration* $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$*. Moreover, the time-reversed process* $(X(T-t))_{0\leqslant t\leqslant T}$ *satisfies the stochastic differential equation*

$$\mathrm{d}X(T-t) = \nabla \log \ell\big(T-t, X(T-t)\big)\,\mathrm{d}t - \nabla\Psi\big(X(T-t)\big)\,\mathrm{d}t + \mathrm{d}\overline{W}^{\mathbb{P}}(T-t), \tag{4.8}$$

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

Since Theorem **??** states that the time-reversed likelihood ratio process (**??**) is a $\mathbb{Q}$-martingale with respect to the filtration $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$, we will first need the analogue of Proposition **??** in terms of the probability measure $\mathbb{Q}$, which is induced by the invariant distribution Q.

Proposition 4.3. *The process* $\big(\overline{W}^{\mathbb{Q}}(T-t)\big)_{0\leqslant t\leqslant T}$ *defined by*

$$\overline{W}^{\mathbb{Q}}(T-t) := W(T-t) - W(T) + 2\int_0^t \Psi\big(X(T-u)\big)\,\mathrm{d}u, \qquad 0 \leqslant t \leqslant T, \tag{4.9}$$

is a Brownian motion under $\mathbb{Q}$*, adapted to the filtration* $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$*. Furthermore, the time-reversed process* $(X(T-t))_{0\leqslant t\leqslant T}$ *satisfies the stochastic differential equation*

$$\mathrm{d}X(T-t) = -\nabla\Psi\big(X(T-t)\big)\,\mathrm{d}t + \mathrm{d}\overline{W}^{\mathbb{Q}}(T-t), \tag{4.10}$$

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

We provide proofs and references for these well-known results in Theorems **??** and **??** of Appendix **??**. In the following lemma we determine the drift term in order to change from the Brownian motion $\big(\overline{W}^{\mathbb{Q}}(T-t)\big)_{0\leqslant t\leqslant T}$ to the Brownian motion $\big(\overline{W}^{\mathbb{P}}(T-t)\big)_{0\leqslant t\leqslant T}$ and vice versa.

Lemma 4.4. *For* $0 \leqslant t \leqslant T$*, we have*

$$\mathrm{d}\overline{W}^{\mathbb{Q}}(T-t) = \frac{\nabla\ell\big(T-t, X(T-t)\big)}{\ell\big(T-t, X(T-t)\big)}\,\mathrm{d}t + \mathrm{d}\overline{W}^{\mathbb{P}}(T-t). \tag{4.11}$$

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

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

Corollary 4.5. *Under* Assumptions **??**, *the time-reversed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\mathrm{d}\ell\big(T-t, X(T-t)\big) = \nabla\ell\big(T-t, X(T-t)\big)\,\mathrm{d}\overline{W}^{\mathbb{Q}}(T-t), \tag{4.12}$$

respectively

$$\begin{aligned}\mathrm{d}\log\ell\big(T-t, X(T-t)\big) = {} & \frac{\nabla\ell\big(T-t, X(T-t)\big)}{\ell\big(T-t, X(T-t)\big)}\,\mathrm{d}\overline{W}^{\mathbb{Q}}(T-t) \\ & - \frac{1}{2}\frac{\big|\nabla\ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\,\mathrm{d}t,\end{aligned} \tag{4.13}$$

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

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

We note that the fact that the drift term in (**??**) vanishes can also be obtained from mechanically applying Itô's formula to the process (**??**), and using (**??**) as well as the backwards Kolmogorov equation (**??**) for the likelihood ratio function $\ell(t,x)$. 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 (**??**) to obtain (**??**). Assertion (**??**) follows from applying Itô's formula to the logarithm of the process (**??**) and using (**??**). □

Now we have all the ingredients to show Theorem **??**.

Proof of Theorem **??**. Plugging formula (**??**) into the stochastic equation (**??**) we see that the time-reversed relative entropy process (**??**) satisfies the stochastic differential equation

$$\begin{aligned} \mathrm{d}\log \ell\big(T-t, X(T-t)\big) = \frac{\nabla \ell\big(T-t, X(T-t)\big)}{\ell\big(T-t, X(T-t)\big)}\, \mathrm{d}\overline{W}^{\mathbb{P}}(T-t) \\ + \frac{1}{2}\frac{\big|\nabla \ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\, \mathrm{d}t, \end{aligned} \tag{4.14}$$

for $0 \leqslant t \leqslant T$, with respect to the filtration $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$. Hence, for $\varepsilon > 0$, the process $(M(T-t))_{0\leqslant t\leqslant T-\varepsilon}$ in (**??**) is a true martingale. Indeed, by condition **??** of Assumptions **??**, the coefficients in (**??**) remain uniformly bounded as long as $0 \leqslant t \leqslant T-\varepsilon$. To show that, in fact, $(M(T-t))_{0\leqslant t\leqslant T}$ is a true martingale, we have to rely on the finite free energy condition (**??**), which in the light of Lemma **??** asserts that the relative entropy $H(P(0)\,|\,Q)$ is finite. This implies that

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T \frac{1}{2}\frac{\big|\nabla \ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\, \mathrm{d}t\right] < \infty. \tag{4.15}$$

Indeed,

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T \frac{1}{2}\frac{\big|\nabla \ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\, \mathrm{d}t\right] = \lim_{\varepsilon\downarrow 0} \mathbb{E}_{\mathbb{P}}\left[\int_0^{T-\varepsilon} \frac{1}{2}\frac{\big|\nabla \ell\big(T-t, X(T-t)\big)\big|^2}{\ell\big(T-t, X(T-t)\big)^2}\, \mathrm{d}t\right] \tag{4.16}$$

$$= \lim_{\varepsilon\downarrow 0} H\big(P(\varepsilon)\,|\,Q\big) - H\big(P(T)\,|\,Q\big) < \infty, \tag{4.17}$$

where the equality (**??**) follows after taking expectations with respect to the probability measure $\mathbb{P}$ in (**??**) at time $t = T-\varepsilon$, and using that $(M(T-t))_{0\leqslant t\leqslant T-\varepsilon}$ is a true martingale. From (**??**) we deduce that the stochastic integral in (**??**) defines an $L^2(\mathbb{P})$-bounded martingale for $0 \leqslant t \leqslant T$.

Summing up, we conclude that $(M(T-t))_{0\leqslant t\leqslant T}$ is a martingale satisfying (**??**), which finishes the proof of Theorem **??**. □

Our next goal is to calculate the limit (**??**) from Theorem **??**. To this end, we do not rely on [**?**] and time reversal any longer, but rather pass to explicit calculations. We first compute the differentials of the likelihood ratio process

$$\ell\big(t, X(t)\big) = \frac{p\big(t, X(t)\big)}{q\big(X(t)\big)}, \qquad t \geqslant 0, \tag{4.18}$$

and its logarithm under the measure $\mathbb{P}$ in the forward direction of time.

We start by recalling the Fokker-Planck equation (**??**), which we write in the form

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

As $p(t,x)$ can be represented in the form

$$p(t,x) = \ell(t,x)\, q(x) = \ell(t,x)\, \mathrm{e}^{-2\Psi(x)}, \tag{4.20}$$

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

$$\partial_t \ell(t,x) = \tfrac{1}{2}\Delta \ell(t,x) - \big\langle \nabla \ell(t,x), \nabla \Psi(x)\big\rangle_{\mathbb{R}^n}. \tag{4.21}$$

We note that equation (**??**) also follows from the proof of Corollary **??**. With its help, we can compute the forward dynamics of the likelihood ratio process (**??**) in the following manner.

Lemma 4.6. *The likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\mathrm{d}\ell(t,X(t)) = \Delta\ell(t,X(t))\,\mathrm{d}t - 2\Big\langle\nabla\ell(t,X(t)),\nabla\Psi(X(t))\Big\rangle_{\mathbb{R}^n}\mathrm{d}t + \nabla\ell(t,X(t))\,\mathrm{d}W(t), \tag{4.22}$$

respectively

$$\begin{aligned}\mathrm{d}\log\ell(t,X(t)) = \frac{\Delta\ell(t,X(t))}{\ell(t,X(t))}\,\mathrm{d}t - \frac{2\Big\langle\nabla\ell(t,X(t)),\nabla\Psi(X(t))\Big\rangle_{\mathbb{R}^n}}{\ell(t,X(t))}\,\mathrm{d}t \\ - \frac{1}{2}\frac{\big|\nabla\ell(t,X(t))\big|^2}{\ell(t,X(t))^2}\,\mathrm{d}t + \frac{\nabla\ell(t,X(t))}{\ell(t,X(t))}\,\mathrm{d}W(t),\end{aligned} \tag{4.23}$$

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

Proof. The canonical coordinate process $(X(t))_{t\geqslant 0}$ satisfies the stochastic equation (**??**). Applying Itô's formula, using (**??**) and (**??**), we obtain (**??**). One more application of Itô's formula leads to (**??**). □

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

$$\ell^\beta(t,X(t)) = \frac{p^\beta(t,X(t))}{q(X(t))}\,, \qquad t\geqslant t_0, \tag{4.24}$$

and its logarithm, again in the forward direction.

Similarly as before, we write the perturbed Fokker-Planck equation (**??**) as

$$\begin{aligned}\partial_t p^\beta(t,x) = \tfrac{1}{2}\Delta p^\beta(t,x) + \big\langle\nabla p^\beta(t,x),\nabla\Psi(x)+\beta(t,x)\big\rangle_{\mathbb{R}^n} \\ + p^\beta(t,x)\,\big(\Delta\Psi(x) + \operatorname{div}\beta(t,x)\big), \qquad t>t_0.\end{aligned} \tag{4.25}$$

Using the relation

$$p^\beta(t,x) = \ell^\beta(t,x)\,q(x) = \ell^\beta(t,x)\,\mathrm{e}^{-2\Psi(x)}, \tag{4.26}$$

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

$$\begin{aligned}\partial_t\ell^\beta(t,x) = \tfrac{1}{2}\Delta\ell^\beta(t,x) + \big\langle\nabla\ell^\beta(t,x),\beta(t,x)-\nabla\Psi(x)\big\rangle_{\mathbb{R}^n} \\ + \ell^\beta(t,x)\,\Big(\operatorname{div}\beta(t,x) - 2\big\langle\beta(t,x),\nabla\Psi(x)\big\rangle_{\mathbb{R}^n}\Big),\end{aligned} \tag{4.27}$$

the analogue of the backwards Kolmogorov equation (**??**) in this "perturbed" context. This helps us obtain the forward dynamics of the perturbed likelihood ratio process (**??**), as follows.

Lemma 4.7. *The perturbed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

$$\begin{aligned}\frac{\mathrm{d}\ell^\beta(t,X(t))}{\ell^\beta(t,X(t))} = \frac{\Delta\ell^\beta(t,X(t))}{\ell^\beta(t,X(t))}\,\mathrm{d}t - \frac{2\Big\langle\nabla\ell^\beta(t,X(t)),\nabla\Psi(X(t))\Big\rangle_{\mathbb{R}^n}}{\ell^\beta(t,X(t))}\,\mathrm{d}t \\ + \operatorname{div}\beta(t,X(t))\,\mathrm{d}t - 2\Big\langle\beta(t,X(t)),\nabla\Psi(X(t))\Big\rangle_{\mathbb{R}^n}\mathrm{d}t + \frac{\nabla\ell^\beta(t,X(t))}{\ell^\beta(t,X(t))}\,\mathrm{d}W(t),\end{aligned} \tag{4.28}$$

and

$$\begin{aligned}\mathrm{d}\log\ell^{\beta}(t,X(t)) = \frac{\Delta\ell^{\beta}(t,X(t))}{\ell^{\beta}(t,X(t))}\,\mathrm{d}t - \frac{2\left\langle\nabla\ell^{\beta}(t,X(t)),\nabla\Psi(X(t))\right\rangle_{\mathbb{R}^n}}{\ell^{\beta}(t,X(t))}\,\mathrm{d}t \\ + \operatorname{div}\beta(t,X(t))\,\mathrm{d}t - 2\left\langle\beta(t,X(t)),\nabla\Psi(X(t))\right\rangle_{\mathbb{R}^n}\mathrm{d}t \\ - \frac{1}{2}\frac{\left|\nabla\ell^{\beta}(t,X(t))\right|^2}{\ell^{\beta}(t,X(t))^2}\,\mathrm{d}t + \frac{\nabla\ell^{\beta}(t,X(t))}{\ell^{\beta}(t,X(t))}\,\mathrm{d}W(t),\end{aligned} \tag{4.29}$$

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

Proof. The canonical coordinate process $(X(t))_{t\geqslant 0}$ satisfies the stochastic equation (**??**). Together with (**??**) and Itô's formula, this yields the stochastic equations (**??**) and (**??**). □

Proof of Theorem **??**. Relying on (**??**), we compute the limit

$$\begin{aligned}\lim_{t\downarrow t_0}\frac{\mathbb{E}_{\mathbb{P}}\left[\log\ell(t,X(t))\mid\mathcal{F}(t_0)\right]}{t-t_0} &= \log\ell(t_0,X(t_0)) + \frac{\Delta\ell(t_0,X(t_0))}{\ell(t_0,X(t_0))} \\ &\quad - \frac{2\left\langle\nabla\ell(t_0,X(t_0)),\nabla\Psi(X(t_0))\right\rangle_{\mathbb{R}^n}}{\ell(t_0,X(t_0))} - \frac{1}{2}\frac{\left|\nabla\ell(t_0,X(t_0))\right|^2}{\ell(t_0,X(t_0))^2},\end{aligned} \tag{4.30}$$

where we used the fact that the conditional expectation of the stochastic integral in (**??**) with respect to $\mathcal{F}(t_0)$ vanishes. Similarly, by means of (**??**), we obtain

$$\begin{aligned}\lim_{t\downarrow t_0}\frac{\mathbb{E}_{\mathbb{P}^{\beta}}\left[\log\ell^{\beta}(t,X(t))\mid\mathcal{F}(t_0)\right]}{t-t_0} &= \log\ell^{\beta}(t_0,X(t_0)) + \frac{\Delta\ell^{\beta}(t_0,X(t_0))}{\ell^{\beta}(t_0,X(t_0))} \\ &\quad - \frac{2\left\langle\nabla\ell^{\beta}(t_0,X(t_0)),\nabla\Psi(X(t_0))\right\rangle_{\mathbb{R}^n}}{\ell^{\beta}(t_0,X(t_0))} - \frac{1}{2}\frac{\left|\nabla\ell^{\beta}(t_0,X(t_0))\right|^2}{\ell^{\beta}(t_0,X(t_0))^2} \\ &\quad + \operatorname{div}\beta(t_0,X(t_0)) - 2\left\langle\beta(t_0,X(t_0)),\nabla\Psi(X(t_0))\right\rangle_{\mathbb{R}^n}.\end{aligned} \tag{4.31}$$

Finally, subtracting (**??**) from (**??**) and noting that $\ell^{\beta}(t_0,X(t_0)) = \ell(t_0,X(t_0))$, we obtain as difference

$$\operatorname{div}\beta(t_0,X(t_0)) - 2\left\langle\beta(t_0,X(t_0)),\nabla\Psi(X(t_0))\right\rangle_{\mathbb{R}^n}, \tag{4.32}$$

which is indeed the right-hand side of (**??**).

It remains to compute the limit (**??**). This is a well-known result and will be shown in Theorem **??**. □

For the sake of completeness, in the remainder of this section we compute also the stochastic differentials of the time-reversed perturbed likelihood ratio process

$$\ell^{\beta}(T-t,X(T-t)) = \frac{p^{\beta}(T-t,X(T-t))}{q(X(T-t))}\,,\qquad 0\leqslant t\leqslant T-t_0, \tag{4.33}$$

and its logarithm. We only do that to make clear that in the perturbed situation the time reversal does *not* work as nicely as in Theorem **??**.

By analogy with previous developments (see Theorems **??** and **??**), the following result is well known to hold under the Assumptions **??** and our assumptions on β.

Proposition 4.8. *The process* $\big(\overline{W}^{\mathbb{P}^\beta}(T-t)\big)_{0\leqslant t\leqslant T-t_0}$ *defined by*

$$\overline{W}^{\mathbb{P}^\beta}(T-t) := W(T-t) - W(T) - \int_0^t \frac{\nabla p^\beta\big(T-u, X(T-u)\big)}{p^\beta\big(T-u, X(T-u)\big)}\,\mathrm{d}u, \qquad 0 \leqslant t \leqslant T-t_0, \tag{4.34}$$

is a Brownian motion with respect to the measure $\mathbb{P}^\beta$ *and the filtration* $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T-t_0}$. *Furthermore, the semimartingale decomposition for the time-reversed process* $(X(T-t))_{0\leqslant t\leqslant T-t_0}$ *is given by*

$$\begin{aligned}\mathrm{d}X(T-t) = \nabla \log \ell^\beta\big(T-t, X(T-t)\big)\,\mathrm{d}t - \nabla\Psi\big(X(T-t)\big)\,\mathrm{d}t \\ + \beta\big(T-t, X(T-t)\big)\,\mathrm{d}t + \mathrm{d}\overline{W}^{\mathbb{P}^\beta}(T-t),\end{aligned} \tag{4.35}$$

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

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

Lemma 4.9. *The time-reversed perturbed likelihood ratio process* (**??**) *and its logarithm satisfy the stochastic differential equations*

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

and

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

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

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

5 The Wasserstein transport

For the convenience of the reader we review in this section some well-known results on Wasserstein transport to show the limits (**??**) and (**??**) in order to complete the proofs of Theorems **??** and **??**.

We recall the definitions of the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$ and of the quadratic Wasserstein distance $W_2(\,\cdot\,,\,\cdot\,)$. 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 subsets 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 with finite second moment, i.e.,

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

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

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

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

In this 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. To this end, we define the time-dependent velocity field

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

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

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

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 the expressions in (**??**) and (**??**) are simply the same. In particular, (**??**) implies that we have $\|v(t)\|_{L^1(\mathbb{R}^n, p(t))} \in L^1(0,T)$, and we can apply [**?**, Lemma 8.1.2] in order to choose a continuous representative. In other words, there exists a narrowly continuous curve $(\tilde{p}(t))_{0 \leqslant t \leqslant T}$ in $\mathscr{P}(\mathbb{R}^n)$ such that $p(t) = \tilde{p}(t)$ for $\mathscr{L}^1$-a.e. $t \in (0,T)$. For convenience, we denote the continuous representative $(\tilde{p}(t))_{0 \leqslant t \leqslant T}$ again by $(p(t))_{0 \leqslant t \leqslant T}$. The narrowly continuous curve $(p(t))_{0 \leqslant t \leqslant T}$ in $\mathscr{P}(\mathbb{R}^n)$ with $p(0) \in \mathscr{P}_2(\mathbb{R}^n)$ satisfies the *continuity equation* (**??**), and condition (**??**). Hence we can use approximation by regular curves [**?**, Lemma 8.1.9] and the representation formula for the continuity equation [**?**, Proposition 8.1.8] in order to see that the assumption $p(0) \in \mathscr{P}_2(\mathbb{R}^n)$ already implies that the curve $(p(t))_{0 \leqslant t \leqslant T}$ is in $\mathscr{P}_2(\mathbb{R}^n)$. Therefore, we are indeed allowed to view $(p(t))_{0 \leqslant t \leqslant T}$ as a curve in the quadratic Wasserstein space $\mathscr{P}_2(\mathbb{R}^n)$.

As $(p(t))_{0 \leqslant t \leqslant T}$ is a narrowly continuous curve in $\mathscr{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (**??**) and $\|v(t)\|_{L^2(\mathbb{R}^n, p(t))} \in L^1(0,T)$, according to (**??**), we can invoke the second implication

of [**?**, Theorem 8.3.1]. The cited theorem relates absolutely continuous curves and the continuity equation. In particular, it tells us that the curve $(p(t))_{0\leqslant t\leqslant T}$ is absolutely continuous [**?**, Definition 1.1.1]. As a consequence, its metric derivative [**?**, Theorem 1.1.2]

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

exists for $\mathscr{L}^1$-a.e. $t \in (0,T)$. Furthermore, [**?**, Theorem 8.3.1] provides the estimate

$$|p'|(t) \leqslant \|v(t)\|_{L^2(\mathbb{R}^n,p(t))} \tag{5.8}$$

for $\mathscr{L}^1$-a.e. $t \in (0,T)$. On the other hand, the time-dependent velocity field $v(t) \equiv v(t,\,\cdot\,)$ of (**??**) is a *gradient*, and therefore 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.,

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

Since $(p(t))_{0\leqslant t\leqslant T}$ is an absolutely continuous curve in $\mathscr{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (**??**) for the time-dependent velocity field $v(t) \equiv v(t,\,\cdot\,)$, which is tangent to the curve, we can apply [**?**, Proposition 8.4.5]. This result characterizes tangent vectors to absolutely continuous curves, and entails for $\mathscr{L}^1$-a.e. $t \in (0,T)$ the inequality

$$\|v(t)\|_{L^2(\mathbb{R}^n,p(t))} \leqslant |p'|(t). \tag{5.10}$$

Combining (**??**) and (**??**), we obtain for $\mathscr{L}^1$-a.e. $t \in (0,T)$ the equality

$$|p'|(t) = \|v(t)\|_{L^2(\mathbb{R}^n,p(t))}. \tag{5.11}$$

Using the metric derivative (**??**) of the curve $(p(t))_{0\leqslant t\leqslant T}$, we can compute the arc length L of the curve with respect to the quadratic Wasserstein distance $W_2(\,\cdot\,,\,\cdot\,)$ by

$$L = \int_0^T |p'|(t)\,\mathrm{d}t. \tag{5.12}$$

This arc length L is nothing other than the quadratic Wasserstein distance between $p(0)$ and $p(T)$ along the curve $(p(t))_{0\leqslant t\leqslant T}$.

Let $t_1, t_2 \geqslant 0$. Motivated by (**??**), we define the *Wasserstein transportation cost* of moving $p(t_1)$ to $p(t_2)$ along the curve $(p(t))_{t\geqslant 0}$ as

$$\mathcal{T}_{\mathsf{c}}(p(t_1),p(t_2)) := \int_{t_1}^{t_2} |p'|(t)\,\mathrm{d}t \tag{5.13}$$

so that, in particular $\mathcal{T}_{\mathsf{c}}(p(0),p(T)) = L$ is the quantity of (**??**). According to (**??**), this transportation cost can be computed as

$$\mathcal{T}_{\mathsf{c}}(p(t_1),p(t_2)) = \int_{t_1}^{t_2} \Big(\int_{\mathbb{R}^n} |v(t,x)|^2\, p(t,x)\,\mathrm{d}x \Big)^{1/2} \mathrm{d}t. \tag{5.14}$$

Furthermore, we note that

$$W_2(p(t_1),p(t_2)) \leqslant \mathcal{T}_{\mathsf{c}}(p(t_1),p(t_2)) \tag{5.15}$$

for $0 \leqslant t_1 \leqslant t_2$, see [**?**, p. 186].

We rephrase these well-known results as follows.

Theorem 5.1. *Let $(p(t))_{t\geqslant 0}$ be a solution of the Fokker-Planck equation* (**??**) *with initial condition $p(0) \in \mathscr{P}_2(\mathbb{R}^n)$ satisfying* Assumptions **??**. *For each $t_1 \geqslant 0$ we have*

$$\lim_{t\to t_1} \frac{W_2(p(t),p(t_1))}{|t-t_1|} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\frac{|\nabla \ell(t_1,X(t_1))|^2}{\ell(t_1,X(t_1))^2} \right] \right)^{1/2}, \tag{5.16}$$

where for $t_1 = 0$ one has to interpret (**??**) *as a limit from the right. Furthermore, for $t_1, t_2 \geqslant 0$, the Wasserstein transportation cost of moving $p(t_1)$ to $p(t_2)$ along the curve $(p(t))_{t\geqslant 0}$ amounts to*

$$\mathcal{T}_{\mathsf{c}}(p(t_1),p(t_2)) = \tfrac{1}{2}\int_{t_1}^{t_2} \left(\int_{\mathbb{R}^n} \frac{|\nabla \ell(t,x)|^2}{\ell(t,x)^2}\, p(t,x)\,\mathrm{d}x \right)^{1/2} \mathrm{d}t. \tag{5.17}$$

Proof. The identity (**??**) is just another way of phrasing the equality (**??**). The Wasserstein transportation cost (**??**) was derived in (**??**). □

Remark 5.2. We note that, in the case $t_1 = 0$, it may very well happen that the Fisher information $I(P(0)\,|\,Q)$ diverges although Assumptions **??** guarantee that $H(P(0)\,|\,Q) < \infty$. In this case (**??**) has to be interpreted as $\infty = \infty$. ◇

Now we consider the solution $(p^\beta(t))_{t\geqslant t_0}$ of the perturbed Fokker-Planck equation (**??**) with initial condition (**??**), and define the time-dependent perturbed velocity field

$$[t_0,T]\times\mathbb{R}^n \ni (t,x) \longmapsto v^\beta(t,x) := -\left(\frac{1}{2}\frac{\nabla p^\beta(t,x)}{p^\beta(t,x)} + \nabla\Psi(x) + \beta(t,x) \right) \in \mathbb{R}^n. \tag{5.18}$$

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

$$\partial_t p^\beta(t,x) + \operatorname{div}\left(v^\beta(t,x)\, p^\beta(t,x) \right) = 0, \qquad (t,x) \in (t_0,T]\times\mathbb{R}^n. \tag{5.19}$$

To follow the same reasoning as above, we need that $v(t,\,\cdot\,)$ is a gradient, and hence we see why we have required $\beta\colon [t_0,\infty)\times\mathbb{R}^n \to \mathbb{R}^n$ to be a gradient field, i.e., of the form $\beta(t,\,\cdot\,) = \nabla B(t,\,\cdot\,)$ for some time-dependent potential $B(t,\,\cdot\,)\colon \mathbb{R}^n \to \mathbb{R}$. Now, by the same token as above, and using the regularity assumption that the time-dependent *gradient* vector field $(\beta(t,\,\cdot\,))_{t\geqslant t_0}$ is compactly supported and of class $\mathcal{C}^{1,\infty}$, we obtain the following result.

Theorem 5.3. *Let $(p^\beta(t))_{t\geqslant t_0}$ be a solution of the perturbed Fokker-Planck equation* (**??**) *with initial condition $p^\beta(t_0)$ as in* (**??**). *Then*

$$\lim_{t\downarrow t_0} \frac{W_2(p^\beta(t),p^\beta(t_0))}{t-t_0} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}}\left[\left| \frac{\nabla \ell(t_0,X(t_0))}{\ell(t_0,X(t_0))} + 2\,\beta(t_0,X(t_0)) \right|^2 \right] \right)^{1/2}, \tag{5.20}$$

and for each $t_1 > t_0$, we have

$$\lim_{t\to t_1} \frac{W_2(p^\beta(t),p^\beta(t_1))}{|t-t_1|} = \tfrac{1}{2}\left(\mathbb{E}_{\mathbb{P}^\beta}\left[\left| \frac{\nabla \ell^\beta(t_1,X(t_1))}{\ell^\beta(t_1,X(t_1))} + 2\,\beta(t_1,X(t_1)) \right|^2 \right] \right)^{1/2}. \tag{5.21}$$

Moreover, for $t_1, t_2 \geqslant t_0$, the Wasserstein transportation cost of moving $p^\beta(t_1)$ to $p^\beta(t_2)$ along the perturbed curve $(p^\beta(t))_{t\geqslant t_0}$ amounts to

$$\mathcal{T}_{\mathsf{c}}(p^\beta(t_1),p^\beta(t_2)) = \tfrac{1}{2}\int_{t_1}^{t_2} \left(\int_{\mathbb{R}^n} \left| \frac{\nabla \ell^\beta(t,x)}{\ell^\beta(t,x)} + 2\,\beta(t,x) \right|^2 p(t,x)\,\mathrm{d}x \right)^{1/2} \mathrm{d}t. \tag{5.22}$$

Remark 5.4. Since $X(t_0)$ has the same distribution under $\mathbb{P}$, as it does under $\mathbb{P}^\beta$, the expectation $\mathbb{E}_{\mathbb{P}}$ appearing in (**??**) can be replaced by $\mathbb{E}_{\mathbb{P}^\beta}$. ◇

Appendices

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

In this section, which is only of historical interest, we want to point out that Bachelier already had some thoughts on *"horizontal transport of probability measures"* in his thesis *"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 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 lonely 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 $(W(T-t))_{0\leqslant t\leqslant T}$ 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 $\nabla p(t,x)/p(t,x)$. At this stage, it should come as no surprise, that the passage to reverse time is closely related to the Wasserstein transport induced by $p(t,x)$.

We finish the section by returning to Bachelier's thesis. The *rapporteur* of Bachelier's thesis was no less a figure than Poincaré. Apparently he saw the enormous potential of these ideas 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 Monsieur Bachelier did not develop further this part of his thesis.

B Proof of the Fontbona-Jourdain result

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

$$\ell(0,X(0)) = \frac{p(0,X(0))}{q(X(0))} \in L^1(\mathcal{C}[0,T],\mathcal{G}(0),\mathbb{Q}) \tag{B.1}$$

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

$$M(T-t) := \mathbb{E}_{\mathbb{Q}}\Big[\ell(0,X(0)) \;\big|\; \mathcal{G}(T-t)\Big], \qquad 0\leqslant t\leqslant T. \tag{B.2}$$

Obviously the stochastic process $(M(T-t))_{0\leqslant t\leqslant T}$ is a $\mathbb{Q}$-martingale with respect to the reverse filtration $(\mathcal{G}(T-t))_{0\leqslant t\leqslant T}$. 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 $(X(T-t))_{0\leqslant t\leqslant T}$ is a Markov process, too, under $\mathbb{P}$ as well as under $\mathbb{Q}$. Hence

$$M(T-t) = \mathbb{E}_{\mathbb{Q}}\Big[\ell(0,X(0)) \;\big|\; X(T-t)\Big], \qquad 0\leqslant t\leqslant T. \tag{B.3}$$

We have to show that this last conditional expectation equals $\ell(T-t,X(T-t))$. To this end, we fix $t\in[0,T]$ as well as a Borel set $A\subseteq\mathbb{R}^n$, and denote by $\pi(T-t;x,A)$ the transition probability of the event $\{X(T-t)\in A\}$, 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

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

Note also that

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

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

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

□

C Proof of Lemma ??

In order to show (**??**), we define the time-dependent velocity field

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

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.,

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,T_t^\gamma(x) = v_t(T_t^\gamma(x)). \tag{C.2}$$

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

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

which can be written equivalently as

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

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

$$X_t = X_0 + \int_0^t v_s(X_s)\,\mathrm{d}s, \qquad 0\leqslant t\leqslant 1, \tag{C.5}$$

or equivalently $P_t = (T_t^\gamma)_{\#}(P_0)$, $0\leqslant t\leqslant 1$, defines random variables X_t with distributions P_t for $t\in[0,1]$. We have now

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

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

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

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

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

For the likelihood ratio function

$$\ell_t(x) = \frac{p_t(x)}{q(x)}, \qquad (t,x) \in [0,1] \times \mathbb{R}^n \tag{C.9}$$

we get from (**??**) and (**??**) that

$$\mathrm{d} \log \ell_t(X_t) = 2 \left\langle \nabla \Psi(X_t), v_t(X_t) \right\rangle_{\mathbb{R}^n} \mathrm{d}t \, - \, \mathrm{div}\left(v_t(X_t)\right) \mathrm{d}t, \qquad 0 \leqslant t \leqslant 1. \tag{C.10}$$

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

$$H(P_t \,|\, Q) - H(P_0 \,|\, Q) = \mathbb{E}_\nu\big[\log \ell_t(X_t)\big] - \mathbb{E}_\nu\big[\log \ell_0(X_0)\big] \tag{C.11}$$

is equal to

$$\mathbb{E}_\nu\bigg[\int_0^t \Big(2 \left\langle \nabla \Psi(X_s), v_s(X_s) \right\rangle_{\mathbb{R}^n} - \mathrm{div}\left(v_s(X_s)\right) \Big) \mathrm{d}s \bigg] \tag{C.12}$$

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

$$\lim_{t \downarrow 0} \frac{H(P_t \,|\, Q) - H(P_0 \,|\, Q)}{t} = \mathbb{E}_\nu\Big[2 \left\langle \nabla \Psi(X_0), v_0(X_0) \right\rangle_{\mathbb{R}^n} - \mathrm{div}\left(v_0(X_0)\right) \Big]. \tag{C.13}$$

Integrating by parts, we see that

$$\mathbb{E}_\nu\big[\, \mathrm{div}\left(v_0(X_0)\right)\big] = \int_{\mathbb{R}^n} \mathrm{div}\left(v_0(x)\right) p_0(x) \, \mathrm{d}x \tag{C.14}$$

$$= - \int_{\mathbb{R}^n} \left\langle v_0(x), \nabla p_0(x) \right\rangle \mathrm{d}x \tag{C.15}$$

$$= - \big\langle \nabla \log p_0(X_0), v_0(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)}. \tag{C.16}$$

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

$$\lim_{t \downarrow 0} \frac{H(P_t \,|\, Q) - H(P_0 \,|\, Q)}{t} = \big\langle \nabla \log \ell_0(X_0), v_0(X_0) \big\rangle_{L^2(\nu;\mathbb{R}^n)}. \tag{C.17}$$

Since $v_0(X_0) = \gamma(X_0)$, this leads to (**??**). □

D Time reversal of diffusions

We review in the present section the theory of time reversal of diffusion processes developed by Föllmer [**?**, **?**], Haussmann and Pardoux [**?**], and Pardoux [**?**]. This section can be read independently of the rest of the paper.

D.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 time reversal, "we may know too much". Thus, the following questions arise rather naturally:

Given a diffusion process (in particular, a strong Markov process with continuous paths and a semimartingale) $X = (X(t))_{0 \leqslant t \leqslant T}$ with certain specific drift and dispersion characteristics, under what conditions might the time-reversed process

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

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 by workers in the theory of filtering, interpolation of 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 [**?**, **?**]. Here, this theory is presented in the spirit of the expository paper by Meyer [**?**].

D.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

$$\mathcal{F}(t) = \sigma\big(\xi, W(s) \colon\ 0 \leqslant s \leqslant t\big), \qquad 0 \leqslant t \leqslant T,$$

modulo $\mathbb{P}$-augmentation. Next, we assume that the system of stochastic equations

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

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

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

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

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

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]

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

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

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

D.3 Time reversal and the backwards filtration

Consider now the filtration $(\widehat{\mathcal{F}}(T-t))_{0\leqslant t\leqslant T}$ given by

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

It is not hard to see that this filtration is expressed equivalently as

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

Here, the σ-algebra of Brownian increments after time t, namely

$$\widehat{\mathcal{G}}(T-t) := \sigma\big(W(s) - W(t)\colon\ t \leqslant s \leqslant T\big), \qquad 0 \leqslant t \leqslant T, \tag{D.8}$$

is independent of the random vector $X(t)$. In particular, $\widehat{\mathcal{F}}(T-t)$ is generated by the terminal value $X(T)$ and by the increments of W on $[t,T]$.

The time-reversed processes $\widehat{X}$ as in (**??**), as well as

$$\widetilde{W}(t) := W(T-t) - W(T), \qquad 0 \leqslant t \leqslant T, \tag{D.9}$$

are both adapted to the *backwards filtration* $\widehat{\mathbb{F}} := (\widehat{\mathcal{F}}(t))_{0\leqslant t\leqslant T}$, where

$$\widehat{\mathcal{F}}(t) = \sigma\big(X(T-u), W(T-u) - W(T-t)\colon\ 0 \leqslant u \leqslant t\big) = \sigma\big(\widehat{X}(u), \widetilde{W}(u) - \widetilde{W}(t)\colon\ 0 \leqslant u \leqslant t\big)$$

from (**??**). Note that, by complete analogy with (**??**), we have also

$$\widehat{\mathcal{F}}(t) = \sigma\big(X(T), W(T-u) - W(T-t)\colon\ 0 \leqslant u \leqslant t\big) = \sigma\big(\widehat{X}(0)\big) \vee \widehat{\mathcal{G}}(t) \tag{D.10}$$

on account of (**??**), where

$$\widehat{\mathcal{G}}(t) = \sigma\big(W(T-u) - W(T-t)\colon\ 0 \leqslant u \leqslant t\big) = \sigma\big(\widetilde{W}(u) - \widetilde{W}(t)\colon\ 0 \leqslant u \leqslant t\big). \tag{D.11}$$

In words: the σ-algebra $\widehat{\mathcal{F}}(t)$ 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 $[0,t]$; see the expressions right above. Furthermore, the σ-algebra $\widehat{\mathcal{F}}(t)$ measures all the random variables $\widehat{X}(u)$, $u \in [0,t]$.

Remark D.1. In fact, the process $\widetilde{W}$ *is a Brownian motion of the filtration* $\widehat{\mathbb{G}} := (\widehat{\mathcal{G}}(t))_{0\leqslant t\leqslant T}$ as in (**??**), generated by the increments of W after time $T-t$, $0 \leqslant t \leqslant T$.

This is because it is a martingale with respect to 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 backwards filtration* $\widehat{\mathbb{F}}$ and identify its semimartingale decomposition. ⋄

D.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, $\widehat{\mathcal{F}}(T-t)$-measurable random variable $\mathcal{K}$, we have

$$\mathbb{E}\big[\mathcal{K} \,|\, \mathcal{F}(t)\big] = \mathbb{E}\big[\mathcal{K} \,|\, X(t)\big], \quad \text{almost surely.} \tag{D.12}$$

Secondly, we fix a function $G \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ and a time-point $t \in (0,T]$, and define

$$g(s,x) := \mathbb{E}\big[G\big(X(t)\big) \,|\, X(s) = x\big], \qquad (s,x) \in [0,t] \times \mathbb{R}^n.$$

Invoking the Markov property of X, we deduce that the process

$$g\big(s, X(s)\big) = \mathbb{E}\big[G\big(X(t)\big) \,|\, X(s)\big] = \mathbb{E}\big[G\big(X(t)\big) \,|\, \mathcal{F}(s)\big], \qquad 0 \leqslant s \leqslant t$$

is an $\mathbb{F}$-martingale, and obtain

$$G\big(X(t)\big) - g\big(s, X(s)\big) = g\big(t, X(t)\big) - g\big(s, X(s)\big) = \sum_{i=1}^n \sum_{\nu=1}^d \int_s^t D_i g\big(u, X(u)\big)\, \mathrm{s}_{i\nu}\big(u, X(u)\big)\, \mathrm{d}W_\nu(u).$$

For every index $\nu = 1, \dots, d$ this gives, after integrating by parts,

$$\mathbb{E}\big[\big(W_\nu(t) - W_\nu(s)\big) \cdot G\big(X(t)\big)\big] = \mathbb{E}\Big[\big(W_\nu(t) - W_\nu(s)\big) \cdot \Big(g\big(t, X(t)\big) - g\big(s, X(s)\big)\Big)\Big]$$

$$= \mathbb{E}\Big[\sum_{i=1}^n \int_s^t D_i g\big(u, X(u)\big)\, \mathrm{s}_{i\nu}\big(u, X(u)\big)\, \mathrm{d}u\Big] = \sum_{i=1}^n \int_s^t \int_{\mathbb{R}^n} (D_i g \cdot \mathrm{s}_{i\nu})(u,x)\, p(u,x)\, \mathrm{d}x\, \mathrm{d}u$$

$$= -\sum_{i=1}^n \int_s^t \int_{\mathbb{R}^n} g(u,x)\, D_i\big(p(u,x)\, \mathrm{s}_{i\nu}(u,x)\big)\, \mathrm{d}x\, \mathrm{d}u = -\int_s^t \int_{\mathbb{R}^n} g(u,x)\ \mathrm{div}\,\big(p(u,x)\, \overline{\mathrm{s}}_\nu(u,x)\big)\, \mathrm{d}x\, \mathrm{d}u$$

$$= -\int_s^t \mathbb{E}\Big[g\big(u, X(u)\big) \cdot \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}\big(u, X(u)\big)\Big]\, \mathrm{d}u = -\mathbb{E}\Big[G\big(X(t)\big) \cdot \int_s^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}\big(u, X(u)\big)\, \mathrm{d}u\Big].$$

Here $\overline{\mathrm{s}}_\nu(u, \cdot\,)$ is the ν^{th} column vector of the dispersion matrix, and we have set

$$\mathrm{div}\,\big(p(u,x)\, \overline{\mathrm{s}}_\nu(u,x)\big) := \sum_{i=1}^n D_i\big(p(u,x)\, \mathrm{s}_{i\nu}(u,x)\big), \qquad \nu = 1, \dots, d.$$

Comparing the first and last expressions in the above string of equalities, we see that with $0 \leqslant s \leqslant t$ we have

$$\mathbb{E}\Big[G\big(X(t)\big) \cdot \Big(W_\nu(t) - W_\nu(s) + \int_s^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}\big(u, X(u)\big)\, \mathrm{d}u\Big)\Big] = 0 \tag{D.13}$$

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 D.2. *The vector process $B = (B_1, \dots, B_d)'$ defined as*

$$B_\nu(t) := \widetilde{W}_\nu(t) - \int_0^t \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}\big(T-u, \widehat{X}(u)\big)\, \mathrm{d}u \tag{D.14}$$

$$= W_\nu(T-t) - W_\nu(T) - \int_{T-t}^T \frac{\mathrm{div}(p\, \overline{\mathrm{s}}_\nu)}{p}\big(v, X(v)\big)\, \mathrm{d}v, \qquad 0 \leqslant t \leqslant T, \tag{D.15}$$

for $\nu = 1, \dots, d$, is Brownian motion with respect to the backwards filtration $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}(t))_{0 \leqslant t \leqslant T}$.

Remark D.3. The Brownian motion process B is thus independent of $\widehat{\mathcal{F}}(0)$, and therefore also of the $\widehat{\mathcal{F}}(0)$-measurable random variable $X(T)$. A bit more generally,

$$\{B(T-s)-B(T-t)\colon\ 0\leqslant s\leqslant t\}\quad\text{ is independent of }\quad\widehat{\mathcal{F}}(T-t)\supseteq\sigma\big(X(u)\colon\ t\leqslant u\leqslant T\big).$$

Note also from (**??**) that

$$B_\nu(T-s)-B_\nu(T-t)=W_\nu(s)-W_\nu(t)-\int_s^t\frac{\operatorname{div}(p\,\overline{\mathrm{s}}_\nu)}{p}\big(v,X(v)\big)\,\mathrm{d}v,\qquad 0\leqslant s\leqslant t.\qquad\diamond$$

Reversing time once again, we obtain the following corollary of Theorem **??**.

Corollary D.4. *The $\mathbb{F}$-adapted vector process $V=(V_1,\dots,V_d)'$ with components*

$$V_\nu(t):=B_\nu(T-t)-B_\nu(T)=W_\nu(t)+\int_0^t\frac{\operatorname{div}(p\,\overline{\mathrm{s}}_\nu)}{p}\big(u,X(u)\big)\,\mathrm{d}u,\qquad 0\leqslant t\leqslant T,\tag{D.16}$$

for $\nu=1,\dots,d$, is yet another Brownian motion (with respect to its own filtration $\mathbb{F}^V\subseteq\mathbb{F}$). This process is independent of the random variable $X(T)$; and a bit more generally, for every $t\in(0,T]$, the σ-algebra

$$\mathcal{F}^V(t):=\sigma\big(V(u)\colon\ 0\leqslant u\leqslant t\big)\tag{D.17}$$

generated by present-and-past values of V, is independent of $\sigma(X(u)\colon\ t\leqslant u\leqslant T)$, the σ-algebra generated by present-and-future values of X.

Proof of Theorem **??**. It suffices to show that each B_ν is a martingale of $\widehat{\mathbb{F}}$; because then, in view of the continuity of paths and the easily checked property $\langle B_\nu,B_\ell\rangle(t)=t\,\delta_{\nu\ell}$, 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 $\ell\neq\nu$, appealing to Lévy's theorem once again.

Now we have to show $\mathbb{E}\big[\big(B_\nu(T-s)-B_\nu(T-t)\big)\cdot\mathcal{K}\big]=0$ for $0\leqslant s\leqslant t\leqslant T$ and every bounded, $\widehat{\mathcal{F}}(T-t)$-measurable $\mathcal{K}$; equivalently,

$$\mathbb{E}\bigg[\mathbb{E}\big[\mathcal{K}\,|\,\mathcal{F}(t)\big]\cdot\Big(W_\nu(t)-W_\nu(s)+\int_s^t\frac{\operatorname{div}(p\,\overline{\mathrm{s}}_\nu)}{p}\big(u,X(u)\big)\,\mathrm{d}u\Big)\bigg]=0,$$

as the expression inside the curved braces is $\mathcal{F}(t)$-measurable. But recalling (**??**) we have

$$\mathbb{E}\big[\mathcal{K}\,|\,\mathcal{F}(t)\big]=\mathbb{E}\big[\mathcal{K}\,|\,X(t)\big]=G\big(X(t)\big)$$

for some bounded, measurable $G\colon\mathbb{R}^n\to\mathbb{R}$, and the desired result follows from (**??**). □

D.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

$$\begin{aligned}X_i(t)-\xi_i-\int_0^t b_i\big(s,X(s)\big)\,\mathrm{d}s&=\sum_{\nu=1}^d\int_0^t\mathrm{s}_{i\nu}\big(s,X(s)\big)\,\mathrm{d}W_\nu(s)\\&=\sum_{\nu=1}^d\Big(\int_0^t\mathrm{s}_{i\nu}\big(s,X(s)\big)\bullet\mathrm{d}W_\nu(s)-\big\langle\mathrm{s}_{i\nu}(\,\cdot\,,X),W_\nu\big\rangle(t)\Big).\end{aligned}$$

From (**??**), we have by Itô's formula that the process

$$\mathrm{s}_{i\nu}(\,\cdot\,,X)-\mathrm{s}_{i\nu}(0,\xi)-\sum_{j=1}^n\sum_{\nu=1}^d\int_0^{\cdot}D_j\mathrm{s}_{i\nu}\big(t,X(t)\big)\cdot\mathrm{s}_{j\nu}\big(t,X(t)\big)\,\mathrm{d}W_\nu(t)$$

is of finite first variation, therefore

$$\langle \mathrm{s}_{i\nu}(\,\cdot\,, X), W_\nu\rangle(t) = \sum_{j=1}^{n} \int_0^t \mathrm{s}_{j\nu}(s, X(s))\, D_j \mathrm{s}_{i\nu}(s, X(s))\, \mathrm{d}s.$$

We conclude

$$X_i(t) = \xi_i - \int_0^t \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i \Big)(s, X(s))\, \mathrm{d}s + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(s, X(s)) \bullet \mathrm{d}W_\nu(s).$$

Evaluating also at $t = T$, then subtracting, we obtain

$$X_i(t) = X_i(T) + \int_t^T \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i \Big)(s, X(s))\, \mathrm{d}s - \sum_{\nu=1}^{d} \int_t^T \mathrm{s}_{i\nu}(s, X(s)) \bullet \mathrm{d}W_\nu(s),$$

as well as

$$\widehat{X}_i(t) = \widehat{X}_i(0) + \int_0^t \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} - b_i \Big)(T - s, \widehat{X}(s))\, \mathrm{d}s + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(T - s, \widehat{X}(s))\, \mathrm{d}\widetilde{W}_\nu(s)$$

by reversing time. 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

$$\begin{aligned} \widehat{X}_i(t) = \widehat{X}_i(0) &+ \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(T - s, \widehat{X}(s))\, \mathrm{d}B_\nu(s) \\ &+ \int_0^t \Big(\sum_{j=1}^{n} \sum_{\nu=1}^{d} \mathrm{s}_{j\nu}\, D_j \mathrm{s}_{i\nu} + \sum_{\nu=1}^{d} \mathrm{s}_{i\nu}\, \frac{\mathrm{div}(p\,\overline{\mathrm{s}}_\nu)}{p} - b_i \Big)(T - s, \widehat{X}(s))\, \mathrm{d}s, \qquad 0 \leqslant t \leqslant T. \end{aligned}$$

But in conjunction with Theorem **??**, this means that the time-reversed process $\widehat{X}$ of (**??**), (**??**) is a semimartingale of the backwards filtration $\widehat{\mathbb{F}}$, with decomposition

$$\widehat{X}_i(t) = \widehat{X}_i(0) + \int_0^t \widehat{b}_i(T - s, \widehat{X}(s))\, \mathrm{d}s + \sum_{\nu=1}^{d} \int_0^t \mathrm{s}_{i\nu}(T - s, \widehat{X}(s))\, \mathrm{d}B_\nu(s) \tag{D.18}$$

for $0 \leqslant t \leqslant T$, where, for each $i = 1, \dots, n$, 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 D.5. *Under the assumptions of this section, the time-reversed process $\widehat{X}$ of* (**??**), (**??**) *is a diffusion in the backwards filtration $\widehat{\mathbb{F}}$, with characteristics as in* (**??**)*, namely, dispersions $\mathrm{s}_{i\nu}(T - t, x)$ and drifts $\widehat{b}_i(T - t, x)$ given by the* **generalized Nelson equation**

$$\widehat{b}_i(t,x) + b_i(t,x) = \sum_{j=1}^{n} \Big(D_j \mathrm{a}_{ij}(t,x) + \mathrm{a}_{ij}(t,x) \cdot D_j \log p(t,x) \Big), \qquad i = 1, \dots, n. \tag{D.19}$$

Equivalently, and with $\operatorname{div}\big(\mathrm{a}(t,x)\big) := \big(\sum_{j=1}^n D_j \mathrm{a}_{ij}(t,x)\big)_{1\leqslant i\leqslant n}$, we write

$$\widehat{b}(t,x) + b(t,x) = \operatorname{div}\big(\mathrm{a}(t,x)\big) + \mathrm{a}(t,x)\cdot\nabla \log p(t,x). \tag{D.20}$$

Remark D.6. This result can be extended to the case where the sums of the distributional derivatives $\sum_{j=1}^n D_j\big(\mathrm{a}_{ij}(t,x)\,p(t,x)\big)$, $i = 1,\dots,n$, are only assumed to be locally integrable functions of $x \in \mathbb{R}^n$; see [**?**, **?**]. ◇

D.5.1 Some filtration comparisons

For an invertible matrix $\mathrm{s}(\,\cdot\,,\,\cdot\,)$, it follows from (**??**) that the Brownian motion B is adapted to the filtration generated by $\widehat{X}$; that is,

$$\mathcal{F}^B(t) \subseteq \mathcal{F}^{\widehat{X}}(t), \qquad 0 \leqslant t \leqslant T. \tag{D.21}$$

Now look at (**??**); in its light, the filtration comparison in (**??**) implies $\mathcal{F}^{\widetilde{W}}(t) \subseteq \mathcal{F}^{\widehat{X}}(t)$, $0 \leqslant t \leqslant T$, thus

$$\widehat{\mathcal{G}}(t) \subseteq \mathcal{F}^{\widetilde{W}}(t) \subseteq \mathcal{F}^{\widehat{X}}(t), \qquad 0 \leqslant t \leqslant T,$$

from (**??**), and from (**??**) also

$$\widehat{\mathcal{F}}(t) \subseteq \mathcal{F}^{\widehat{X}}(t), \qquad 0 \leqslant t \leqslant T. \tag{D.22}$$

These considerations inform our choice of backwards filtration in (**??**).

D.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*

$$\int_0^{\cdot} Y(t) \bullet \mathrm{d}X(t) := \int_0^{\cdot} Y(t)\,\mathrm{d}M(t) + \int_0^{\cdot} Y(t)\,\mathrm{d}B(t) + \langle M, N\rangle, \tag{D.23}$$

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 $\Pi = \{t_0, t_1, \dots, t_m\}$ is a partition of the interval $[0,T]$ with $0 = t_0 < t_1 < \ldots < t_m = T$, then the sums

$$\sum_{j=0}^{m-1} Y(t_{j+1})\big(X(t_{j+1}) - X(t_j)\big) \tag{D.24}$$

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

$$f(X) = f\big(X(0)\big) + \sum_{i=1}^n \int_0^{\cdot} D_i f\big(X(t)\big) \bullet \mathrm{d}X_i(t) - \frac{1}{2}\sum_{i,j=1}^n \int_0^{\cdot} D_{ij}^2 f\big(X(t)\big)\,\mathrm{d}\langle M_i, M_j\rangle(t), \tag{D.25}$$

where now $X = (X_1, \dots, X_n)'$ is a vector of continuous semimartingales $X_1, \dots, X_n$ of the form $X_i = X_i(0) + M_i + B_i$ as above, for $i = 1, \dots, n$. Note the change of sign, from (+) to (−) in the last, stochastic correction term.