

PATHWISE OTTO CALCULUS*

Ioannis Karatzas[†]Walter Schachermayer[‡]Bertram Tschiderer[§]

7th February 2019

ABSTRACT. We revisit the [JKO98] variational characterization of diffusion as entropic gradient flux, and provide for it a probabilistic interpretation based on stochastic calculus. It was shown by Jordan, Kinderlehrer, and Otto in [JKO98] that, for diffusions of Langevin-Smoluchowski type, the Fokker-Planck probability density flow minimizes the rate of relative entropy dissipation, as measured by the distance traveled in terms of the quadratic Wasserstein metric. We obtain novel, stochastic-process versions of these features, valid along almost every trajectory of the diffusive motion in both the forward and, most transparently, the backward, directions of time, using a very direct perturbation analysis; the original results follow then simply by taking expectations. As a bonus, we derive the Cordero-Erausquin version of the so-called HWI inequality relating relative entropy, Fisher information and Wasserstein distance.

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

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

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

A preliminary version of this work, under the title “Applying Itô calculus to Otto calculus”, was posted on arXiv on 21 November 2018 (arXiv:1811.08686v1).

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

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

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

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

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

CONTENTS

1. INTRODUCTION	2
1.1. Preview	5
2. THE STOCHASTIC APPROACH	5
3. THE THEOREMS	7
3.1. Main results	10
3.2. Important consequences	11
3.3. Ramifications	16
4. DETAILS AND PROOFS	20
4.1. Some preliminaries	20
4.2. The case $Q(\mathbb{R}^n) < \infty$	21
4.3. The proof of Theorem 3.6	23
4.4. The proof of Theorem 3.8	24
4.5. Some useful lemmas	27
4.6. Completing the proof of Proposition 3.14	32
4.7. The dynamics in the forward direction of time	33
5. THE WASSERSTEIN TRANSPORT	34
APPENDICES	38
A. BACHELIER’S WORK RELATING BROWNIAN MOTION TO THE HEAT EQUATION	38
B. THE PROOFS OF LEMMAS 2.1 AND 3.3	39
C. RELATIVE ENTROPY WITH RESPECT TO A σ -FINITE MEASURE Q	40
D. A MEASURE-THEORETIC RESULT	41
E. THE PROOF OF THE FONTBONA-JOURDAIN THEOREM	42
F. THE PROOF OF LEMMA 3.19	43
G. TIME REVERSAL OF DIFFUSIONS	44
G.1. Introduction	44
G.2. The setting	45
G.3. Time reversal and the backwards filtration	46
G.4. Some remarkable Brownian motions	47
G.5. The diffusion property under time reversal	48
G.6. The backwards Itô integral	50

1. INTRODUCTION

We give a trajectorial interpretation of a seminal result by Jordan, Kinderlehrer, and Otto [JKO98], and provide a proof based on stochastic calculus. The basic theme of our approach could be described as “applying Itô calculus to Otto calculus”. More precisely, we follow a stochastic analysis approach to Otto’s characterization of diffusions of Langevin-Smoluchowski type as entropic gradient fluxes in Wasserstein space, and provide stronger, trajectorial versions of these results. For consistency and

better readability we adopt the setting and notation of [JKO98], and even copy some paragraphs of this paper almost verbatim in the remainder of this introductory section.

Following the lines of [JKO98] we thus consider a Fokker-Planck equation of the form

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

with initial condition

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

Here, ρ is a real-valued function defined for $(t, x) \in [0, \infty) \times \mathbb{R}^n$, the function $\Psi: \mathbb{R}^n \rightarrow [0, \infty)$ is smooth and plays the role of a potential, $\beta > 0$ is a real constant, and ρ^0 is a probability density on \mathbb{R}^n . The solution $\rho(t, x)$ of (1.1) with initial condition (1.2) stays non-negative and conserves its mass, which means that the spatial integral

$$\int_{\mathbb{R}^n} \rho(t, x) dx \quad (1.3)$$

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

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

$$dX(t) = -\nabla \Psi(X(t)) dt + \sqrt{2\beta^{-1}} dW(t), \quad X(0) = X^0. \quad (1.4)$$

In the equation above, $(W(t))_{t \geq 0}$ is an n -dimensional Brownian motion started from 0, and the \mathbb{R}^n -valued random variable X^0 is independent of the process $(W(t))_{t \geq 0}$. The probability distribution of X^0 has density ρ^0 and, unless specified otherwise, the reference measure will always be Lebesgue measure on \mathbb{R}^n . Then $\rho(t, \cdot)$, the solution of (1.1) with initial condition (1.2), gives at any given time $t \geq 0$ the probability density function of the random variable $X(t)$ from (1.4).

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

$$Z(\beta) = \int_{\mathbb{R}^n} e^{-\beta\Psi(x)} dx \quad (1.5)$$

is finite and there exists a unique stationary solution of the Fokker-Planck equation (1.1); namely, the probability density ρ_s of the Gibbs distribution given by [Gar09, JK96, Ris96]

$$\rho_s(x) = (Z(\beta))^{-1} e^{-\beta\Psi(x)} \quad (1.6)$$

for $x \in \mathbb{R}^n$. When it exists, the probability measure on \mathbb{R}^n with density function ρ_s is called *Gibbs distribution*, and is the unique invariant measure for the Markov process $(X(t))_{t \geq 0}$ defined by the stochastic differential equation (1.4); see, e.g., [KS91, Exercise 5.6.18, p. 361].

In [JK96] it is shown that the stationary probability density ρ_s satisfies the following variational principle: it minimizes the *free energy functional*

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

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

$$E(\rho) := \int_{\mathbb{R}^n} \Psi \rho dx \quad (1.8)$$

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

$$S(\rho) := \int_{\mathbb{R}^n} \rho \log \rho dx. \quad (1.9)$$

In accordance with [JKO98] we consider the following regularity assumptions.

Assumptions 1.1 (REGULARITY ASSUMPTIONS OF [JKO98, THEOREM 5.1]).

- (i) The potential $\Psi: \mathbb{R}^n \rightarrow [0, \infty)$ is of class $\mathcal{C}^\infty(\mathbb{R}^n; [0, \infty))$ and satisfies, for some real constant $C > 0$, the bound

$$|\nabla\Psi| \leq C(\Psi + 1). \quad (1.10)$$

- (ii) The distribution of $X(0)$ in (1.4) 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.,

$$\int_{\mathbb{R}^n} \rho^0(x) |x|^2 dx < \infty \quad \text{and} \quad F(\rho^0) \in \mathbb{R}. \quad (1.11)$$

We shall impose also the following, additional assumptions.

Assumptions 1.2 (REGULARITY ASSUMPTIONS OF THE PRESENT PAPER). In addition to conditions

(i) and (ii) of Assumptions 1.1, we also impose that:

- (iii) The potential Ψ satisfies the drift condition

$$\langle x, \nabla\Psi(x) \rangle_{\mathbb{R}^n} \geq -c|x|^2 \quad (1.12)$$

for all $x \in \mathbb{R}^n$ with $|x| \geq R$, for some real constants $c \geq 0$, $R \geq 0$.

- (iv) The potential Ψ is sufficiently well-behaved to guarantee that the solution of (1.4) is well-defined for all $t \geq 0$, and that the solution $(t, x) \mapsto \rho(t, x)$ of (1.1) with initial condition (1.2) is continuous and strictly positive on $(0, \infty) \times \mathbb{R}^n$, and smooth in the space variable x for each $t > 0$. We also assume that the logarithmic derivative $(t, x) \mapsto \nabla \log \rho(t, x)$ is continuous on $(0, \infty) \times \mathbb{R}^n$. For example, by requiring that all derivatives of Ψ grow at most exponentially as $|x|$ tends to infinity, one may adapt the arguments from [Rog85] showing that this is indeed the case.

For the formulation of Theorem 3.4 we will need a vector field $\beta: \mathbb{R}^n \rightarrow \mathbb{R}^n$ which is the gradient of a potential $B: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying the following regularity assumption:

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

The Assumptions 1.2 are satisfied by typical convex potentials Ψ . They also accommodate examples such as double-well potentials of the form $\Psi(x) = (x^2 - \alpha^2)^2$ on the real line, for real constants $\alpha > 0$. Furthermore, they guarantee that the second-moment condition in (1.11) propagates in time, i.e.,

$$\int_{\mathbb{R}^n} \rho(t, x) |x|^2 dx < \infty, \quad t \geq 0; \quad (1.13)$$

see Lemma 2.1 below. Also note that these assumptions do not rule out the important case when the constant $Z(\beta)$ in (1.5) is infinite; thus, they allow for cases (such as $\Psi = 0$) in which the stationary probability density function ρ_s does not exist. In fact, in [JKO98] the authors point out explicitly that, even when the stationary probability density ρ_s is not defined, the free energy (1.7) of a density $\rho(t, x)$ satisfying the Fokker-Planck equation (1.1) with initial condition (1.2) can be defined, provided that $F(\rho^0)$ is finite.

1.1. PREVIEW

We set up in Section 2 our model for the Langevin-Smoluchowski diffusion and introduce its fundamental quantities, such as the current and the invariant distribution of particles, the resulting likelihood ratio process, as well as the associated concepts of free energy, relative entropy, and relative Fisher information.

Section 3 presents our basic results. These include Theorem 3.1, which computes in terms of the relative Fisher information the rate of relative entropy decay in the ambient Wasserstein space of probability density functions with finite second moment; as well as its “perturbed” counterpart, Theorem 3.4. We compute explicitly the difference between these perturbed and unperturbed rates, and show that it is always non-negative, in fact strictly positive unless the perturbation and the gradient of the log-likelihood ratio are collinear. This way, the Langevin-Smoluchowski diffusion emerges as the *steepest descent* (or “gradient flux”) of the relative entropy functional with respect to the Wasserstein metric.

We also show that both Theorems 3.1 and 3.4 follow as very simple consequences of their stronger, pathwise versions, Theorems 3.6 and 3.8, respectively. These latter are the main results of this work; they provide very detailed descriptions of the semimartingale dynamics for the relative entropy process, in both its “pure” and “perturbed” forms. Such descriptions are most transparent when time is *reversed*, so we choose to present them primarily in this context. Several important consequences and ramifications of Theorems 3.6, 3.8 are presented in Subsections 3.2 and 3.3, including a derivation of the Cordero-Erausquin form of the famous HWI inequality that relates relative entropy (H) to Wasserstein distance (W) and to relative Fisher information (I).

Most of the detailed arguments and proofs are collected in Section 4 and in the appendices. In particular, Appendix G presents a completely self-contained account of time reversal for Itô diffusion processes. The necessary background on optimal Wasserstein transport is recalled in Section 5.

2. THE STOCHASTIC APPROACH

Thus far, we have been mostly quoting from [JKO98]. We adopt now a more probabilistic point of view, and translate our setting into the language of stochastic processes and probability measures. For notational convenience, and without loss of generality, we fix the constant $\beta > 0$ to equal 2, so that the stochastic differential equation (1.4) becomes

$$dX(t) = -\nabla\Psi(X(t)) dt + dW(t), \quad t \geq 0. \quad (2.1)$$

Let $P(0)$ be a probability measure on the Borel sets of \mathbb{R}^n with density function $p^0(x) := \rho^0(x)$. We shall study the stochastic differential equation (2.1) with initial probability distribution $P(0)$.

While we do make an effort to follow the setting and notation of [JKO98] as closely as possible, our notation here differs slightly from [JKO98]. To conform with our more probabilistic approach, we shall use from now onward the familiar letters p^0 and $p(0, \cdot)$ rather than ρ^0 and $\rho(0, \cdot)$.

The initial probability measure $P(0)$ on \mathbb{R}^n with density function $p(0, \cdot)$, induces a probability measure \mathbb{P} on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$ of \mathbb{R}^n -valued continuous functions on $\mathbb{R}_+ = [0, \infty)$, under which the canonical coordinate process $(X(t)(\omega))_{t \geq 0} = (\omega(t))_{t \geq 0}$ satisfies the stochastic differential equation (2.1) with initial probability distribution $P(0)$. We shall denote by $P(t)$ the probability distribution of the random vector $X(t)$ under \mathbb{P} , and by $p(t) \equiv p(t, \cdot)$ the corresponding probability density function, at each time $t \geq 0$, which is the solution of (1.1).

We shall see in Appendix B that, in conjunction with the second-moment condition in (1.11), the drift condition (1.12) guarantees finite second moments of the probability density functions $p(t)$ at all times $t \geq 0$; equivalently, membership of the probability distribution $P(t)$ in the space $\mathcal{P}_2(\mathbb{R}^n)$ of

definition (5.1) in Section 5, for all $t \geq 0$. This property also holds when the potential Ψ is replaced by $\Psi + B$ as in condition (v) of Assumptions 1.2; see Lemma 3.3.

Lemma 2.1. *Under the Assumptions 1.2, the Langevin-Smoluchowski diffusion (2.1) with initial distribution $P(0)$ admits a pathwise unique, strong solution, which satisfies $P(t) \in \mathcal{P}_2(\mathbb{R}^n)$ for all $t \geq 0$.*

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

$$\frac{dP(t)}{dQ}(X(t)) = \ell(t, X(t)), \quad \text{where} \quad \ell(t, x) := \frac{p(t, x)}{q(x)} = p(t, x) e^{2\Psi(x)} \quad (2.2)$$

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

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

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

$$H(P(t) | Q) := \mathbb{E}_P[\log \ell(t, X(t))] = \int_{\mathbb{R}^n} \log \left(\frac{p(t, x)}{q(x)} \right) p(t, x) dx, \quad t \geq 0, \quad (2.4)$$

$$I(P(t) | Q) := \mathbb{E}_P[|\nabla \log \ell(t, X(t))|^2], \quad t \geq 0. \quad (2.5)$$

Remark 2.2. Following the approach of [Léo14, Section 2], we show in Appendix C that the relative entropy $H(P | Q)$ is well-defined for every probability measure P in $\mathcal{P}_2(\mathbb{R}^n)$ and takes values in $(-\infty, \infty]$. \diamond

The following well-known identity (2.6) states that the relative entropy $H(P(t) | Q)$ is equal to the free energy functional $F(p(t, \cdot))$, up to a factor of 2, for all $t \geq 0$. In light of condition (ii) in Assumptions 1.1, this identity implies $H(P(0) | Q) \in \mathbb{R}$, so the quantity in (2.4) is finite for $t = 0$; thus, on account of (3.30) below, finite also for $t > 0$.

Lemma 2.3. *Under the Assumptions 1.2, and along the curve of probability measures $(P(t))_{t \geq 0}$, the free energy functional in (1.7) and the relative entropy in (2.4) are related for each $t \geq 0$ through the equation*

$$2F(p(t, \cdot)) = H(P(t) | Q). \quad (2.6)$$

Proof. Indeed,

$$\begin{aligned} \mathbb{E}_P[\log \ell(t, X(t))] &= \mathbb{E}_P\left[\log \left(e^{2\Psi(X(t))} p(t, X(t)) \right)\right] = \mathbb{E}_P[2\Psi(X(t))] + \mathbb{E}_P[\log p(t, X(t))] \\ &= 2 \int_{\mathbb{R}^n} \Psi(x) p(t, x) dx + \int_{\mathbb{R}^n} p(t, x) \log p(t, x) dx, \end{aligned}$$

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

The identity (2.6) shows that studying the decay of the free energy functional $F(p(t, \cdot))$, is equivalent to studying the decay of the relative entropy $H(P(t) | Q)$.

3. THE THEOREMS

As already indicated in (1.1) and (1.4), the probability density function $p: [0, \infty) \times \mathbb{R}^n \rightarrow [0, \infty)$ solves the Fokker-Planck or forward Kolmogorov [Kol31] equation [Fri75, Gar09, Ris96, Sch80]

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

with initial condition

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

By contrast, the function $q(\cdot)$ does not depend on the temporal variable, and solves the stationary version of the forward Kolmogorov equation (3.1), namely

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

In light of Lemma 2.3, the object of interest in [JKO98] is to relate the decay of the *relative entropy functional*

$$\mathcal{P}_2(\mathbb{R}^n) \ni P \mapsto H(P | \mathbb{Q}) \in (-\infty, \infty] \quad (3.4)$$

along the curve $(P(t))_{t \geq 0}$, to the quadratic Wasserstein distance W_2 defined in (5.3) of Section 5. We resume the remarkable relation between these two quantities in the following two theorems; these provide a way to quantify the relationship between displacement in the ambient space (the denominator of the expression in (3.7)) and fluctuations of the free energy, or equivalently of the relative entropy (the numerator in the expression (3.7)). The proofs will be given in Subsection 3.2 below.

Theorem 3.1. *Under the Assumptions 1.2, we have for Lebesgue-almost every $t_0 \geq 0$ the generalized de Bruijn identity*

$$\lim_{t \rightarrow t_0} \frac{H(P(t) | \mathbb{Q}) - H(P(t_0) | \mathbb{Q})}{t - t_0} = -\frac{1}{2} I(P(t_0) | \mathbb{Q}), \quad (3.5)$$

as well as the local behavior of the Wasserstein distance

$$\lim_{t \rightarrow t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \frac{1}{2} \sqrt{I(P(t_0) | \mathbb{Q})}, \quad (3.6)$$

so that

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

The ratio on the left-hand side of (3.7) can be interpreted as the slope of the relative entropy functional (3.4) at $P = P(t_0)$ along the curve $(P(t))_{t \geq 0}$, if we measure distances in $\mathcal{P}_2(\mathbb{R}^n)$ by the quadratic Wasserstein distance W_2 of (5.3). The quantity appearing on the right-hand side of (3.7) is the square root of the relative Fisher information in (2.5), written more explicitly in terms of the “score function” $\nabla \ell(t, \cdot) / \ell(t, \cdot)$ as

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

Remark 3.2. Under the Assumptions 1.2 it is perfectly possible for the relative Fisher information $I(P(t_0) | \mathbb{Q})$ to be infinite at $t_0 = 0$. For instance, think of $p(0, \cdot)$ as the indicator function of a subset of \mathbb{R}^n with Lebesgue measure equal to 1. In this case both sides of equation (3.7) are equal to $-\infty$, as will follow from Corollary 3.10. \diamond

The remarkable insight of [JKO98] states that the slope in (3.7) in the direction of the curve $(P(t))_{t \geq 0}$ is, in fact, the slope of steepest descent for the relative entropy functional (3.4) at the point $P = P(t_0)$. To formalize this assertion, we fix a time $t_0 \geq 0$ and let $\beta = \nabla B: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a gradient vector field of class $C^\infty(\mathbb{R}^n; \mathbb{R}^n)$ having compact support, as in condition (v) of Assumptions 1.2. This vector field β will serve as a perturbation, and we consider the thus perturbed Fokker-Planck equation

$$\partial_t p^\beta(t, x) = \operatorname{div} \left((\nabla \Psi(x) + \beta(x)) p^\beta(t, x) \right) + \frac{1}{2} \Delta p^\beta(t, x), \quad (t, x) \in (t_0, \infty) \times \mathbb{R}^n \quad (3.9)$$

with initial condition

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

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

$$dX(t) = - \left(\nabla \Psi(X(t)) + \beta(X(t)) \right) dt + dW^\beta(t), \quad t \geq t_0 \quad (3.11)$$

with initial probability distribution $P(t_0)$. Here, the process $(W^\beta(t))_{t \geq t_0}$ is a Brownian motion under \mathbb{P}^β . The probability distribution of $X(t)$ under \mathbb{P}^β on \mathbb{R}^n will be denoted by $P^\beta(t)$, for $t \geq t_0$; once again, the corresponding density function $p^\beta(t) \equiv p^\beta(t, \cdot)$ solves the equation (3.9) subject to the initial condition (3.10).

In the following analogue of Lemma 2.1, we state that the perturbed probability density functions $p^\beta(t)$ also admit finite second moments at all times $t \geq t_0$. For the proof we refer again to Appendix B.

Lemma 3.3. *Under the Assumptions 1.2, let $t_0 \geq 0$. The perturbed diffusion (3.11) with initial distribution $P^\beta(t_0) = P(t_0)$ admits a pathwise unique, strong solution, which satisfies $P^\beta(t) \in \mathcal{P}_2(\mathbb{R}^n)$ for all $t \geq t_0$.*

After these preparations we can state the result formalizing the gradient flux, or *steepest descent*, property of the flow $(P(t))_{t \geq 0}$ generated by the Langevin-Smoluchowski diffusion (2.1), with respect to the Wasserstein metric.

Theorem 3.4. *Under the Assumptions 1.2, the following assertions hold for Lebesgue-almost every $t_0 \geq 0$: The \mathbb{R}^n -valued random vectors*

$$a := \nabla \log \ell(t_0, X(t_0)) = \nabla \log p(t_0, X(t_0)) + 2 \nabla \Psi(X(t_0)), \quad b := \beta(X(t_0)) \quad (3.12)$$

are elements of the Hilbert space $L^2(\mathbb{P})$, and the perturbed version of the generalized de Bruijn identity (3.5) reads

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) | \mathbb{Q}) - H(P^\beta(t_0) | \mathbb{Q})}{t - t_0} = -\frac{1}{2} I(P(t_0) | \mathbb{Q}) - \langle a, b \rangle_{L^2(\mathbb{P})} = -\langle a, \frac{a}{2} + b \rangle_{L^2(\mathbb{P})}. \quad (3.13)$$

Furthermore, the local behavior of the Wasserstein distance (3.6) in this perturbed context is given by

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \frac{1}{2} \|a + 2b\|_{L^2(\mathbb{P})}. \quad (3.14)$$

Combining (3.13) with (3.14), and assuming $\|a + 2b\|_{L^2(\mathbb{P})} > 0$, we have

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) | \mathbb{Q}) - H(P^\beta(t_0) | \mathbb{Q})}{W_2(P^\beta(t), P^\beta(t_0))} = - \left\langle a, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}, \quad (3.15)$$

and therefore

$$\lim_{t \downarrow t_0} \left(\frac{H(P^\beta(t) | Q) - H(P^\beta(t_0) | Q)}{W_2(P^\beta(t), P^\beta(t_0))} - \frac{H(P(t) | Q) - H(P(t_0) | Q)}{W_2(P(t), P(t_0))} \right) \quad (3.16)$$

$$= \|a\|_{L^2(\mathbb{P})} - \left\langle a, \frac{a + 2b}{\|a + 2b\|_{L^2(\mathbb{P})}} \right\rangle_{L^2(\mathbb{P})}. \quad (3.17)$$

Remark 3.5. On the strength of the Cauchy-Schwarz inequality, the expression in (3.17) is non-negative, and vanishes if and only if a and b are collinear. Consequently, when the vector field β is not a scalar multiple of $\nabla \log \ell(t_0, \cdot)$, the difference of the two slopes in (3.16) is strictly positive. In other words, the slope quantified by the first term of the difference (3.16), is then strictly bigger than the (negative) slope expressed by the second term of (3.16). \diamond

These two theorems are essentially well known. They build upon a vast amount of previous work. In the quadratic case $\Psi(x) = |x|^2/4$, i.e., when the process $(X(t))_{t \geq 0}$ in (2.1) is Ornstein-Uhlenbeck with invariant measure in (1.6) standard Gaussian, the relation

$$\frac{d}{dt} H(P(t) | Q) = -\frac{1}{2} I(P(t) | Q) \quad (3.18)$$

has been known since [Sta59] as *de Bruijn's identity*. This relationship between the two fundamental information measures, due to Shannon and Fisher, respectively, is a dominant theme in many aspects of information theory and probability. We refer to the book [CT06] by Cover and Thomas for an account of the results by Barron, Blachman, Brown, Linnik, Rényi, Shannon, Stam and many others in this vein, as well as to the book [Vil03] by Villani. See also the paper by Carlen and Soffer [CS91] on the relation of (3.18) to the central limit theorem.

In (3.5), the de Bruijn identity (3.18) is established for more general Q that satisfy Assumptions 1.2; see also the seminal work [BÉ85] by Bakry and Émery.

The paper [JKO98] broke new ground in this respect, as it considered a general potential Ψ and established the relation to the quadratic Wasserstein distance, culminating with the characterization of $(p(t, \cdot))_{t \geq 0}$ as a gradient flux. This relation was further investigated by Otto in the paper [Ott01], where the theory now known as “*Otto calculus*” was developed.

The statements of our Theorems 3.1, 3.4 complement the existing results in some detail, e.g., the precise form (3.17), measuring the difference of the two slopes appearing in (3.16). The main novelty of our approach, however, will only become apparent with the formulation of the pathwise Theorems 3.6, 3.8 below.

We shall thus investigate Theorems 3.1 and 3.4 in a trajectorial fashion, by considering the *relative entropy process*

$$\log \ell(t, X(t)) = \log \left(\frac{p(t, X(t))}{q(X(t))} \right) = \log p(t, X(t)) + 2\Psi(X(t)), \quad 0 \leq t \leq T, \quad (3.19)$$

along each trajectory of the canonical coordinate process $(X(t))_{0 \leq t \leq T}$, and calculating its dynamics (stochastic differential) under the probability measure \mathbb{P} . The expectation with respect to \mathbb{P} of this quantity is, of course, the relative entropy in (2.4).

A decisive tool in the analysis of the relative entropy process (3.19) is to reverse time, and use a remarkable insight due to Fontbona and Jourdain [FJ16]. These authors consider the coordinate

process $(X(t))_{0 \leq t \leq T}$ on the path space $\Omega = \mathcal{C}([0, T]; \mathbb{R}^n)$ in the *reverse direction* of time, i.e., they work with the time-reversed process $(X(T - s))_{0 \leq s \leq T}$; it is then notationally convenient to consider a finite time interval $[0, T]$, rather than \mathbb{R}_+ . Of course, this does not restrict the generality of the arguments.

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

$$\mathcal{F}(t) = \sigma(X(u) : 0 \leq u \leq t), \quad t \geq 0; \quad (3.20)$$

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

$$\mathcal{G}(T - s) = \sigma(X(T - u) : 0 \leq u \leq s), \quad 0 \leq s \leq T, \quad (3.21)$$

modulo \mathbb{P} -augmentation. For the necessary measure-theoretic operations that ensure the continuity (from both left and right) of filtrations associated with continuous processes, consult Section 2.7 in [KS91]; in particular, Problems 7.1 – 7.6 and Proposition 7.7.

3.1. MAIN RESULTS

The following two Theorems 3.6 and 3.8 are the main new results of this paper. They can be regarded as trajectorial versions of Theorems 3.1 and 3.4, whose proofs will follow from Theorems 3.6 and 3.8 simply by taking expectations. Similar “trajectorial” approaches have already been applied successfully to the theory of optimal stopping in [DK94], to Doob’s martingale inequalities in [ABP⁺13], and to the Burkholder-Davis-Gundy inequality in [BS15].

The significance of Theorem 3.6 right below, is that the trade-off between the decay of relative entropy and the “Wasserstein transportation cost”, both of which are characterized in terms of the cumulative relative Fisher information process, is valid not only in expectation, but also along (almost) each trajectory, provided we run time in the reverse direction.¹

Theorem 3.6. *Under the Assumptions 1.2, we let $T > 0$ and define the cumulative relative Fisher information process, accumulated from the right, as*

$$F(T - s) := \int_0^s \frac{1}{2} \frac{|\nabla \ell(T - u, X(T - u))|^2}{\ell(T - u, X(T - u))^2} du \quad (3.22)$$

$$= \int_0^s \frac{1}{2} \left| \frac{\nabla p(T - u, X(T - u))}{p(T - u, X(T - u))} + 2\Psi(X(T - u)) \right|^2 du$$

for $0 \leq s \leq T$. Then $\mathbb{E}_{\mathbb{P}}[F(0)] < \infty$, and the process

$$M(T - s) := \left(\log \ell(T - s, X(T - s)) - \log \ell(T, X(T)) \right) - F(T - s) \quad (3.23)$$

for $0 \leq s \leq T$, is a martingale of the backwards filtration $(\mathcal{G}(T - s))_{0 \leq s \leq T}$ under the probability measure \mathbb{P} . More explicitly, the martingale of (3.23) can be represented as

$$M(T - s) = \int_0^s \left\langle \frac{\nabla \ell(T - u, X(T - u))}{\ell(T - u, X(T - u))}, d\overline{W}^{\mathbb{P}}(T - u) \right\rangle_{\mathbb{R}^n}, \quad 0 \leq s \leq T, \quad (3.24)$$

¹As David Kinderlehrer kindly pointed out to the second named author, the implicit Euler scheme used in [JKO98] also reflects the idea of going back in time at each step of the discretization, as does the backward Kolmogorov equation.

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

Remark 3.7. The representation (3.24), in conjunction with the finiteness of $\mathbb{E}_{\mathbb{P}}[F(0)]$, shows that the martingale of (3.23) is bounded in $L^2(\mathbb{P})$. \diamond

Next, we state the trajectorial version of Theorem 3.4 — or equivalently, the “perturbed” analogue of Theorem 3.6. As we did in Theorem 3.4, in particular in the preceding equations (3.9) – (3.11), we consider the perturbation $\beta: \mathbb{R}^n \rightarrow \mathbb{R}^n$ and denote the *perturbed likelihood ratio function* by

$$\ell^\beta(t, x) := \frac{p^\beta(t, x)}{q(x)} = p^\beta(t, x) e^{2\Psi(x)}, \quad (t, x) \in [t_0, \infty) \times \mathbb{R}^n. \quad (3.25)$$

Its stochastic analogue is the *perturbed relative entropy process*

$$\log \ell^\beta(t, X(t)) = \log \left(\frac{p^\beta(t, X(t))}{q(X(t))} \right) = \log p^\beta(t, X(t)) + 2\Psi(X(t)), \quad t_0 \leq t \leq T. \quad (3.26)$$

Theorem 3.8. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. We define the perturbed cumulative relative Fisher information process, accumulated from the right, as*

$$F^\beta(T-s) := \int_0^s \left(\frac{1}{2} \frac{|\nabla \ell^\beta(T-u, X(T-u))|^2}{\ell^\beta(T-u, X(T-u))^2} + \left(\langle \beta, 2\nabla \Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \right)(X(T-u)) \right) du \quad (3.27)$$

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

$$M^\beta(T-s) := \left(\log \ell^\beta(T-s, X(T-s)) - \log \ell^\beta(T, X(T)) \right) - F^\beta(T-s) \quad (3.28)$$

for $0 \leq s \leq T - t_0$, is a martingale of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$ under the probability measure \mathbb{P}^β . More explicitly, the martingale of (3.28) can be represented as

$$M^\beta(T-s) = \int_0^s \left\langle \frac{\nabla \ell^\beta(T-u, X(T-u))}{\ell^\beta(T-u, X(T-u))}, d\overline{W}^{\mathbb{P}^\beta}(T-u) \right\rangle_{\mathbb{R}^n}, \quad 0 \leq s \leq T - t_0, \quad (3.29)$$

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

Remark 3.9. The representation (3.29), in conjunction with the finiteness of $\mathbb{E}_{\mathbb{P}^\beta}[F^\beta(0)]$, shows that the martingale of (3.28) is bounded in $L^2(\mathbb{P}^\beta)$. \diamond

3.2. IMPORTANT CONSEQUENCES

We state now several important consequences of these two basic results, Theorems 3.6 and 3.8. In particular, we indicate how the corresponding assertions in the earlier Theorems 3.1, 3.4 follow directly from these results by taking expectations.

Corollary 3.10. *Under the Assumptions 1.2, we have for all $t, t_0 \geq 0$ the relative entropy identity*

$$H(P(t) | Q) - H(P(t_0) | Q) = \mathbb{E}_{\mathbb{P}} \left[\log \left(\frac{\ell(t, X(t))}{\ell(t_0, X(t_0))} \right) \right] = \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(-\frac{1}{2} \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} \right) du \right]. \quad (3.30)$$

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

$$\lim_{t \rightarrow t_0} \frac{H(P(t) | Q) - H(P(t_0) | Q)}{t - t_0} = -\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right], \quad (3.31)$$

as well as the local behavior of the Wasserstein distance

$$\lim_{t \rightarrow t_0} \frac{W_2(P(t), P(t_0))}{|t - t_0|} = \frac{1}{2} \left(\mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right] \right)^{1/2}. \quad (3.32)$$

Proof of Corollary 3.10 from Theorem 3.6: The identity (3.30) follows by taking expectations with respect to the probability measure \mathbb{P} , and invoking the martingale property of the process in (3.23) for $T \geq \max\{t_0, t\}$. In particular, (3.30) shows that the relative entropy function $t \mapsto H(P(t) | Q)$ from (2.2), thus also the free energy function $t \mapsto F(p(t, \cdot))$ from (2.6), are strictly decreasing provided $\ell(t, \cdot)$ is not constant.

According to the Lebesgue differentiation theorem, the monotone function $t \mapsto H(P(t) | Q)$ is differentiable for Lebesgue-almost every $t_0 \geq 0$, in which case (3.30) leads to the identity (3.31).

The limiting behavior of the Wasserstein distance (3.32), for Lebesgue-almost every $t_0 \geq 0$, is well known and carefully worked out in [AGS08]; see Theorem 5.1 in Section 5 below for the details. \square

Proof of Theorem 3.1 from Theorem 3.6: Recalling the definition of the relative Fisher information (2.5) as well as (3.8), we realize that the limits (3.5) and (3.6) in Theorem 3.1 correspond to the limits (3.31) and (3.32) in the just proved Corollary 3.10. If $t_0 \geq 0$ is chosen such that both limits (3.5) and (3.6) exist, we can divide them by each other, in order to obtain equation (3.7) of Theorem 3.1 for Lebesgue-almost every $t_0 \geq 0$. \square

In a similar way as we derived the above Corollary 3.10 from Theorem 3.6, we now deduce the following Corollary 3.11 from Theorem 3.8. Its first identity (3.33) shows, in particular, that the relative entropy $H(P^\beta(t) | Q)$ is real-valued for all $t \geq t_0$.

Corollary 3.11. *Under the Assumptions 1.2, we have for all $t \geq t_0 \geq 0$ the relative entropy identity*

$$\begin{aligned} H(P^\beta(t) | Q) - H(P^\beta(t_0) | Q) &= \mathbb{E}_{\mathbb{P}^\beta} \left[\log \left(\frac{\ell^\beta(t, X(t))}{\ell^\beta(t_0, X(t_0))} \right) \right] \\ &= \mathbb{E}_{\mathbb{P}^\beta} \left[\int_{t_0}^t \left(-\frac{1}{2} \frac{|\nabla \ell^\beta(u, X(u))|^2}{\ell^\beta(u, X(u))^2} + \left(\operatorname{div} \beta - \langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n} \right) (X(u)) \right) du \right]. \end{aligned} \quad (3.33)$$

Furthermore, we have for Lebesgue-almost every $t_0 \geq 0$ the limiting identities

$$\begin{aligned} \lim_{t \downarrow t_0} \frac{H(P^\beta(t) | Q) - H(P^\beta(t_0) | Q)}{t - t_0} \\ = \mathbb{E}_{\mathbb{P}} \left[-\frac{1}{2} \frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} + \left(\operatorname{div} \beta - \langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n} \right) (X(t_0)) \right], \end{aligned} \quad (3.34)$$

as well as

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \frac{1}{2} \left(\mathbb{E}_{\mathbb{P}} \left[\left| \frac{\nabla \ell(t_0, X(t_0))}{\ell(t_0, X(t_0))} + 2 \beta(X(t_0)) \right|^2 \right] \right)^{1/2}. \quad (3.35)$$

Proof of Corollary 3.11 from Theorem 3.8: Taking expectations under the probability measure \mathbb{P}^β and using the martingale property of the process in (3.28) for $T \geq t$, leads to the identity (3.33).

In order to derive the limiting identity (3.34) from the equation (3.33), some care is needed to show that (3.34) is valid for every time $t_0 \geq 0$ which is not an exceptional point excluded by Theorem 3.1 or equivalently by Corollary 3.10. More precisely, if $t_0 \geq 0$ is such that the one-sided limit from the right, corresponding to (3.31), can be derived from (3.30) in Corollary 3.10, we have to show that for the same point t_0 the perturbed equation (3.33) leads to the identity (3.34); see in this context also Remark 4.14 below.

We shall verify in Lemma 4.12 of Subsection 4.5 below the following estimates on the ratio between the probability density function $p(t, x)$ and its perturbed version $p^\beta(t, x)$: For every $t_0 \geq 0$ and $T > t_0$ there is a constant $C > 0$ such that

$$\left| \frac{\ell^\beta(t, x)}{\ell(t, x)} - 1 \right| = \left| \frac{p^\beta(t, x)}{p(t, x)} - 1 \right| \leq C(t - t_0), \quad (t, x) \in [t_0, T] \times \mathbb{R}^n \quad (3.36)$$

as well as

$$\mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left| \nabla \log \left(\frac{\ell^\beta(u, X(u))}{\ell(u, X(u))} \right) \right|^2 du \right] \leq C(t - t_0)^2, \quad t_0 \leq t \leq T. \quad (3.37)$$

We turn now to the derivation of (3.34) from (3.33). First, as the perturbation β is smooth and compactly supported, we have clearly

$$\lim_{t \downarrow t_0} \frac{1}{t - t_0} \mathbb{E}_{\mathbb{P}^\beta} \left[\int_{t_0}^t (\operatorname{div} \beta - \langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n})(X(u)) du \right] = \mathbb{E}_{\mathbb{P}^\beta} \left[(\operatorname{div} \beta - \langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n})(X(t_0)) \right]. \quad (3.38)$$

Since the random variable $X(t_0)$ has the same distribution under \mathbb{P} , as it does under \mathbb{P}^β , it is immaterial whether we express the expectation appearing on the right-hand side of (3.38) with respect to the probability measure \mathbb{P} or \mathbb{P}^β . Hence this expression in (3.38) equals the corresponding term in the second line of (3.34), as required. Regarding the remaining term, we apply (3.36) and (3.37) to obtain

$$\lim_{t \downarrow t_0} \frac{1}{t - t_0} \mathbb{E}_{\mathbb{P}^\beta} \left[\int_{t_0}^t \left(-\frac{1}{2} \frac{|\nabla \ell^\beta(u, X(u))|^2}{\ell^\beta(u, X(u))^2} \right) du \right] = \lim_{t \downarrow t_0} \frac{1}{t - t_0} \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(-\frac{1}{2} \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} \right) du \right], \quad (3.39)$$

if one of the limits exists. Hence, the one-sided limiting assertion from the right, corresponding to (3.31), and the limiting assertion in (3.34), fail on precisely the same set of exceptional points $t_0 \geq 0$. Therefore, both limits in (3.39) exist if $t_0 \geq 0$ is not contained in this exceptional set of zero Lebesgue measure, and their common value is

$$-\frac{1}{2} I(P(t_0) | Q) = -\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right]; \quad (3.40)$$

in conjunction with (3.38), this proves the limiting identity (3.34) for every such t_0 .

As regards the final assertion (3.35) we note that, by analogy with (3.32), the limiting behavior of the Wasserstein distance (3.35), for Lebesgue-almost every $t_0 \geq 0$, is well known [AGS08]; for the details we refer to Theorem 5.2 in Section 5 below. Once again, concerning the relation between the limits in (3.32) and (3.35), we find the same picture as in the case of the Fisher information. The one-sided limiting assertion from the right, corresponding to (3.32), and the limiting assertion in (3.35), fail on exactly the same set of exceptional points $t_0 \geq 0$. We shall come back to this point in Remark 4.15 below.

Summing up, for precisely those points $t_0 \geq 0$ such that the two one-sided limiting assertions from the right, corresponding to (3.31) and (3.32), are valid, the limiting assertions (3.34) and (3.35), are also valid. Hence, except for a set of zero Lebesgue measure, we have shown the validity of (3.34) and (3.35), thus finishing the proof of Corollary 3.11. \square

Proof of Theorem 3.4 from Theorems 3.6, 3.8: Let $t_0 \geq 0$ be such that the one-sided limiting assertions from the right, corresponding to (3.31), (3.32) in Corollary 3.10 of Theorem 3.6, and thus also the limiting assertions (3.34), (3.35) in Corollary 3.11 of Theorem 3.8, are valid. Recalling the abbreviations $a = \nabla \log \ell(t_0, X(t_0))$ and $b = \beta(X(t_0))$ in (3.12), we summarize now the identities just mentioned as

$$\lim_{t \downarrow t_0} \frac{H(P(t) | Q) - H(P(t_0) | Q)}{t - t_0} = -\frac{1}{2} \|a\|_{L^2(\mathbb{P})}^2, \quad (3.41)$$

$$\lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0} = \frac{1}{2} \|a\|_{L^2(\mathbb{P})}, \quad (3.42)$$

$$\lim_{t \downarrow t_0} \frac{H(P^\beta(t) | Q) - H(P^\beta(t_0) | Q)}{t - t_0} = -\langle a, \frac{a}{2} + b \rangle_{L^2(\mathbb{P})}, \quad (3.43)$$

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \frac{1}{2} \|a + 2b\|_{L^2(\mathbb{P})}. \quad (3.44)$$

Indeed, the equations (3.41), (3.42), and (3.44) correspond precisely to (3.31), (3.32), and (3.35), respectively. As for (3.43), we note that, according to equation (3.34) of Corollary 3.11, the limit in (3.43) equals

$$-\frac{1}{2} \|a\|_{L^2(\mathbb{P})}^2 + \mathbb{E}_{\mathbb{P}} \left[\left(\operatorname{div} \beta - 2 \langle \beta, \nabla \Psi \rangle_{\mathbb{R}^n} \right) (X(t_0)) \right]. \quad (3.45)$$

Therefore, in view of the right-hand side of (3.43), we have to show the identity

$$\mathbb{E}_{\mathbb{P}} \left[\left(\operatorname{div} \beta - \langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n} \right) (X(t_0)) \right] = -\langle a, b \rangle_{L^2(\mathbb{P})}. \quad (3.46)$$

In order to do this, we write the left-hand side of (3.46) as

$$\int_{\mathbb{R}^n} \left(\operatorname{div} \beta(x) - \langle \beta(x), 2 \nabla \Psi(x) \rangle_{\mathbb{R}^n} \right) p(t_0, x) \, dx. \quad (3.47)$$

Using — for the first time, and only in order to show the identity (3.46) — integration by parts, and the fact that the perturbation β is assumed to be smooth and compactly supported, we see that the expression (3.47) becomes

$$-\int_{\mathbb{R}^n} \left\langle \beta(x), \nabla \log p(t_0, x) + 2 \nabla \Psi(x) \right\rangle_{\mathbb{R}^n} p(t_0, x) \, dx, \quad (3.48)$$

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

The limiting identities (3.41) – (3.44) now clearly imply the assertions of Theorem 3.4. \square

The following two results, Propositions 3.12 and 3.14 are “trajectorial” versions of Corollaries 3.10 and 3.11, respectively. They compute the rate of temporal change of relative entropy for the equation (2.1) and for its perturbed version (3.11), respectively, in the more precise trajectorial manner of Theorems 3.6, 3.8.

Proposition 3.12. *Under the Assumptions 1.2, for Lebesgue-almost every $t_0 \geq 0$ the relative entropy process (3.19) satisfies, with $T > t_0$, the following trajectorial relations:*

$$\lim_{s \uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}} \left[\log \ell(t_0, X(t_0)) \mid \mathcal{G}(T-s) \right] - \log \ell(T-s, X(T-s))}{T-t_0-s} \quad (3.49)$$

$$= \lim_{s \downarrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}} \left[\log \ell(T-s, X(T-s)) \mid \mathcal{G}(t_0) \right] - \log \ell(t_0, X(t_0))}{s-(T-t_0)} \quad (3.50)$$

$$= \frac{1}{2} \frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} = \frac{1}{2} \left| \frac{\nabla p(t_0, X(t_0))}{p(t_0, X(t_0))} + 2 \Psi(X(t_0)) \right|^2, \quad (3.51)$$

where the limits (3.49) and (3.50) exist in the norm of $L^1(\mathbb{P})$.

Remark 3.13. The limiting assertions (3.49) – (3.51) of Proposition 3.12 are the conditional trajectorial versions of the generalized de Bruijn identity (3.31). \diamond

Proof of Proposition 3.12 from Theorem 3.6: Let $t_0 \geq 0$ be such that the one-sided limiting assertion from the right, corresponding to (3.31) in Corollary 3.10 of Theorem 3.6, is valid, and select $T > t_0$. The martingale property of the process in (3.23) allows us to write the numerator in (3.49) as

$$\mathbb{E}_{\mathbb{P}} \left[F(t_0) - F(T-s) \mid \mathcal{G}(T-s) \right], \quad 0 \leq s \leq T-t_0, \quad (3.52)$$

in the notation of (3.22). Similarly, the numerator in (3.50) equals $\mathbb{E}_{\mathbb{P}} [F(T-s) - F(t_0) \mid \mathcal{G}(t_0)]$, $T-t_0 \leq s \leq T$. By analogy with the derivation of (3.31) from (3.30), where we calculated real-valued expectations, we rely on the Lebesgue differentiation theorem to obtain the corresponding results (3.49) – (3.51) for the conditional expectations. Using the left-continuity of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$, we can invoke a measure-theoretic result spelled out in Proposition D.2 of Appendix D, which establishes the claims (3.49) – (3.51) pertaining to conditional expectations. \square

Proposition 3.14. *Under the Assumptions 1.2, for Lebesgue-almost every $t_0 \geq 0$ the relative entropy process (3.19) satisfies, with $T > t_0$, the following trajectorial relations:*

$$\begin{aligned} & \lim_{s \uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}^\beta} \left[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s) \right] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ &= \frac{1}{2} \frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} - \operatorname{div} \beta(X(t_0)) + \left\langle \beta(X(t_0)), 2 \nabla \Psi(X(t_0)) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (3.53)$$

as well as

$$\begin{aligned} & \lim_{s \uparrow T-t_0} \frac{\mathbb{E}_{\mathbb{P}} \left[\log \ell^\beta(t_0, X(t_0)) \mid \mathcal{G}(T-s) \right] - \log \ell^\beta(T-s, X(T-s))}{T-t_0-s} \\ &= \frac{1}{2} \frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} - \operatorname{div} \beta(X(t_0)) - \left\langle \beta(X(t_0)), \nabla \log p(t_0, X(t_0)) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (3.54)$$

and

$$\begin{aligned} & \lim_{s \uparrow T-t_0} \frac{\log \ell^\beta(T-s, X(T-s)) - \log \ell(T-s, X(T-s))}{T-t_0-s} \\ &= \operatorname{div} \beta(X(t_0)) + \left\langle \beta(X(t_0)), \nabla \log p(t_0, X(t_0)) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (3.55)$$

where the limits in (3.53) – (3.55) exist in the norms of both $L^1(\mathbb{P})$ and $L^1(\mathbb{P}^\beta)$.

Remark 3.15. It is perhaps noteworthy that the three limiting expressions in (3.53), (3.54) and (3.55) are quite different from each other.

The first limiting assertion (3.53) of Proposition 3.14 is the conditional trajectorial version of the perturbed de Bruijn identity (3.34).

We also note that in fact the third limiting assertion (3.55) is valid for all $t_0 > 0$. \diamond

Proof of the assertion (3.53) in Proposition 3.14, from Theorem 3.8: Let $t_0 \geq 0$ be such that the one-sided limiting assertion from the right, corresponding to (3.31) in Corollary 3.10 of Theorem 3.6, is valid, and select $T > t_0$. In Corollary 3.11 of Theorem 3.8 we have seen that the limits from the right in (3.31) and (3.34) have the same exceptional sets, hence also (3.34) holds. Now, for such t_0 , we show the limiting assertion (3.53) in the same way as the assertion (3.49) in the proof of Proposition 3.12 above. Indeed, this time we invoke the \mathbb{P}^β -martingale property of the process in (3.28), and write the numerator in the first line of (3.53) as $\mathbb{E}_{\mathbb{P}^\beta}[F^\beta(t_0) - F^\beta(T - s) \mid \mathcal{G}(T - s)]$, $0 \leq s \leq T - t_0$, in the notation of (3.27). Applying Proposition D.2 of Appendix D in this situation proves the limiting identity (3.53) with respect to the norm of $L^1(\mathbb{P}^\beta)$. As we shall see in Lemma 4.11 of Subsection 4.5 below, the probability measures \mathbb{P} and \mathbb{P}^β are equivalent, and the mutual Radon-Nikodým derivatives $\frac{d\mathbb{P}^\beta}{d\mathbb{P}}$ and $\frac{d\mathbb{P}}{d\mathbb{P}^\beta}$ are bounded on the σ -algebra $\mathcal{F}(T) = \mathcal{G}(0)$ (recall, in this vein, the claims of (3.36)). Hence, convergence in $L^1(\mathbb{P})$ is equivalent to convergence in $L^1(\mathbb{P}^\beta)$. This establishes the $L^1(\mathbb{P})$ -convergence of (3.53), which completes the proof of the limiting assertion (3.53).

The proofs of the limiting assertions (3.54) and (3.55) are postponed to Subsection 4.6. \square

3.3. RAMIFICATIONS

Theorem 3.4 and, in particular, equation (3.43) above, show — at least on a formal level — that the functional

$$\mathcal{P}_2(\mathbb{R}^n) \ni P \longmapsto H(P \mid Q) - H(P(0) \mid Q) \tag{3.56}$$

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

$$\mathcal{P}_2(\mathbb{R}^n) \ni P \longmapsto \langle a, c \rangle_{L^2(\mathbb{P})}, \quad \text{where } c = -\frac{a}{2} - b \tag{3.57}$$

as in (3.43) with $t_0 = 0$ and $a = \nabla \log \ell(0, X(0))$, $b = \beta(X(0))$. As it turns out, formula (3.57) is closely related to the *HWI inequality* of Otto and Villani [OV00] in its sharpened form by Cordero-Erausquin [CE02]; we explain presently how.

Consider the starting time $t_0 = 0$ and the curve $(P^\beta(t))_{t \geq 0}$ as in Theorems 3.4 and 3.8, for a fixed perturbation β as above, and suppose that this t_0 is not an exceptional point in the preceding limiting assertions. Let us study the tangent $(P_t)_{0 \leq t \leq 1}$ to this curve at the point $P^\beta(0) = P(0) = P_0$ in the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$, and analyze the behavior of the relative entropy functional (3.4) along the curve $(P_t)_{0 \leq t \leq 1}$. Here $(P_t)_{0 \leq t \leq 1}$ is understood to be a “straight line” in $\mathcal{P}_2(\mathbb{R}^n)$; i.e., using the terminology of McCann [McC97], as the *displacement interpolation* between the elements P_0 and P_1 in $\mathcal{P}_2(\mathbb{R}^n)$.

Once we have identified the “tangent” $(P_t)_{0 \leq t \leq 1}$, it is geometrically obvious — at least on an intuitive level — that the slope of the relative entropy functional (3.4) along the straight line $(P_t)_{0 \leq t \leq 1}$ should be equal to the slope along $(P^\beta(t))_{t \geq 0}$ at the touching point $P^\beta(0) = P_0$. This slope is given by (3.57), and we shall verify in the following Lemma 3.19 and Proposition 3.21 that — under suitable regularity assumptions — it coincides with the slope along $(P_t)_{0 \leq t \leq 1}$ as identified by Cordero-Erausquin in [CE02].

To work out the connection between (3.57) and [CE02] we shall turn things upside down; i.e., we first define the tangent $(P_t)_{0 \leq t \leq 1}$, and then find the corresponding perturbation β so that the curve $(P^\beta(t))_{t \geq 0}$ indeed has $(P_t)_{0 \leq t \leq 1}$ as tangent at the point $P^\beta(0) = P_0$. In this manner, we shall treat β more as an element of “control”, than a perturbation.

Fix an element $P \in \mathcal{P}_2(\mathbb{R}^n)$, and let $\gamma: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be such that $T(x) := x + \gamma(x)$ transports $P_0 = P(0)$ to $P_1 = P$ optimally, i.e., $T_\#(P_0) = P_1$ and $\|\gamma\|_{L^2(P_0)} = W_2(P_0, P_1)$; see also (3.60) below. Then (3.43), (3.57) suggest that the displacement interpolation $(P_t)_{0 \leq t \leq 1}$ between the two probability measures $P_0 = P(0)$ and $P_1 = P$, to be defined in (3.59) below, is tangent to the curve $(P^\beta(t))_{t \geq 0}$ as in Theorems 3.4 and 3.8, if γ and β are related via

$$\gamma(x) = -\frac{1}{2} \nabla \log \ell(0, x) - \beta(x), \quad x \in \mathbb{R}^n. \quad (3.58)$$

Then the random variable c of (3.57) becomes $c = \gamma(X(0))$.

We formalize these intuitive geometric insights in the subsequent Lemma 3.19, which provides the analogue of (3.43) for the displacement interpolation flow $(P_t)_{0 \leq t \leq 1}$ of (3.59). To this end, we impose temporarily the following strong regularity conditions. As it will turn out in the proof of Proposition 3.21, these will not restrict, eventually, the generality of the argument.

Assumptions 3.16 (REGULARITY ASSUMPTIONS OF LEMMA 3.19). In addition to the conditions (i) – (v) of Assumptions 1.2, we also impose that:

- (vi) P_0 and P_1 are probability measures in $\mathcal{P}_2(\mathbb{R}^n)$ with smooth densities, which are compactly supported and strictly positive on the interior of their respective supports. Hence there is a map $\gamma: \mathbb{R}^n \rightarrow \mathbb{R}^n$ of the form $\gamma(x) = \nabla(G(x) - |x|^2/2)$ for some convex function $G: \mathbb{R}^n \rightarrow \mathbb{R}$, uniquely defined on and supported by the support of P_0 , and smooth in the interior of this set, such that γ induces the optimal quadratic Wasserstein transport from P_0 to P_1 via

$$T_t^\gamma(x) := x + t\gamma(x) = (1-t)x + t\nabla G(x) \quad \text{and} \quad (T_t^\gamma)_\#(P_0) =: P_t \quad (3.59)$$

for $0 \leq t \leq 1$, and $\nabla G = T_1^\gamma = T$; to wit, $(P_t)_{0 \leq t \leq 1}$ is the displacement interpolation between P_0 and P_1 , and we have along it the linear growth of the quadratic Wasserstein distance

$$W_2(P_0, P_t) = t \sqrt{\int_{\mathbb{R}^n} |x - \nabla G(x)|^2 dP_0(x)} = t \|\gamma\|_{L^2(P_0)}, \quad 0 \leq t \leq 1. \quad (3.60)$$

Remark 3.17. For the existence and uniqueness of the optimal transport map $\gamma: \mathbb{R}^n \rightarrow \mathbb{R}^n$ we refer to [Vil03, Theorem 2.12], and for its smoothness to [Vil03, Theorem 4.14] as well as [Vil03, Remarks 4.15]. These results are known collectively under the rubric of *Brenier’s theorem*. \diamond

Remark 3.18. We remark at this point, that we have chosen the subscript notation for P_t in order to avoid confusion with the probability measure $P(t)$ from our Section 2 here. While $P_0 = P(0)$, the flow $(P_t)_{0 \leq t \leq 1}$ from P_0 to P_1 will have otherwise very little to do with the flow $(P(t))_{t \geq 0}$ from $P(0)$ to Q appearing in Theorems 3.1 and 3.4 (except for the tangential relation at $P_0 = P(0)$). Similarly, the likelihood ratio function

$$\ell_t(x) = \frac{p_t(x)}{q(x)}, \quad (t, x) \in [0, 1] \times \mathbb{R}^n, \quad (3.61)$$

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

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

Lemma 3.19. *Under the Assumptions 3.16, let X_0 be a random variable with probability distribution $P_0 = P(0)$, defined on some probability space (S, \mathcal{S}, ν) . Then we have*

$$\lim_{t \downarrow 0} \frac{H(P_t | Q) - H(P_0 | Q)}{t} = \langle \nabla \log \ell_0(X_0), \gamma(X_0) \rangle_{L^2(\nu)}. \quad (3.62)$$

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

We relegate the proof of Lemma 3.19, which follows a similar (but simpler) line of reasoning as the proof of Theorem 3.8, to Appendix F. Combining Lemma 3.19 with well-known arguments, in particular, with a fundamental result on displacement convexity due to McCann [McC97], we derive now the HWI inequality in its improved version of (3.64) due to Cordero-Erausquin [CE02]. This result sharpens the HWI inequality obtained by Otto and Villani [OV00] relating the fundamental quantities of relative entropy (H), Wasserstein distance (W) and relative Fisher information (I).

Proposition 3.21 (CORDERO-ERAUSQUIN). *Under the Assumptions 1.2, we set $P_0 = P(0)$ and let γ be as in (3.59). We suppose in addition that the potential $\Psi: \mathbb{R}^n \rightarrow [0, \infty)$ satisfies a curvature lower bound*

$$\text{Hess}(\Psi) \geq \kappa \text{Id}, \quad (3.63)$$

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

$$H(P_0 | Q) - H(P_1 | Q) \leq -\langle \nabla \log \ell_0(X_0), \gamma(X_0) \rangle_{L^2(\nu)} - \frac{\kappa}{2} W_2^2(P_0, P_1), \quad (3.64)$$

where the random variable X_0 , the likelihood ratio function ℓ_0 , and the probability measure ν are as in Lemma 3.19.

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

$$-\langle \nabla \log \ell_0(X_0), \gamma(X_0) \rangle_{L^2(\nu)} \leq \|\nabla \log \ell_0(X_0)\|_{L^2(\nu)} \|\gamma(X_0)\|_{L^2(\nu)}, \quad (3.65)$$

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

$$I(P_0 | Q) = \mathbb{E}_\nu \left[|\nabla \log \ell_0(X_0)|^2 \right] = \|\nabla \log \ell_0(X_0)\|_{L^2(\nu)}^2, \quad (3.66)$$

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

$$\|\gamma(X_0)\|_{L^2(\nu)} = W_2(P_0, P_1) \quad (3.67)$$

as in (3.60), along with the inequality

$$-\langle \nabla \log \ell_0(X_0), \gamma(X_0) \rangle_{L^2(\nu)} \leq \sqrt{I(P_0 | Q)} W_2(P_0, P_1). \quad (3.68)$$

Inserting (3.68) into (3.64) we obtain the usual form of the HWI inequality

$$H(P_0 | Q) - H(P_1 | Q) \leq W_2(P_0, P_1) \sqrt{I(P_0 | Q)} - \frac{\kappa}{2} W_2^2(P_0, P_1). \quad (3.69)$$

When there is a non-trivial angle between $\nabla \log \ell_0(X_0)$ and $\gamma(X_0)$ in $L^2(\nu)$, the inequality (3.64) gives a sharper bound than (3.69), as has been established by Cordero-Erausquin in [CE02]. We refer to the original paper [OV00], as well as to [Vil03, Chapter 5] and the recent paper [GLRT18], for a detailed discussion of the HWI inequality (3.69).

Remark 3.22. Let us suppose now that the strong non-degeneracy condition (3.63) holds with $\kappa > 0$, and that \mathbb{Q} is a probability measure in $\mathcal{P}_2(\mathbb{R}^n)$. Then the inequality (3.69) contains as special cases the *Talagrand [Tal96] and logarithmic Sobolev [Gro75] inequalities*, namely

$$W_2^2(P, \mathbb{Q}) \leq \frac{2}{\kappa} H(P | \mathbb{Q}), \quad H(P | \mathbb{Q}) \leq \frac{1}{2\kappa} I(P | \mathbb{Q}), \quad (3.70)$$

respectively. On the other hand, and now in the context of Section 2, the second inequality in (3.70) leads, in conjunction with the generalized de Bruijn identity (3.31) and (2.5), to

$$\frac{d}{dt} H(P(t) | \mathbb{Q}) \leq -\kappa H(P(t) | \mathbb{Q}), \quad (3.71)$$

and thence to the Bakry-Émery [BÉ85] *exponential dissipation of the relative entropy*:

$$H(P(t) | \mathbb{Q}) \leq H(P(t_0) | \mathbb{Q}) e^{-\kappa(t-t_0)}, \quad t \geq t_0. \quad (3.72)$$

For an exposition of the Bakry-Émery theory, which derives also the *exponential temporal dissipation of the Fisher information*, we refer to [BGL14] and [Gen14]. \diamond

The inequality (3.69) is yet another manifestation of the interplay between displacement in the ambient space of probability measures (the quantity $W_2(P_0, P_1)$) and fluctuations of the relative entropy (the quantity $H(P_0 | \mathbb{Q}) - H(P_1 | \mathbb{Q})$) as governed by the square root of the Fisher information $\sqrt{I(P_0 | \mathbb{Q})}$, very much in the spirit of (3.7).

Proof of Proposition 3.21. As elaborated in [Vil03, Section 9.4] we may assume without loss of generality that P_0 and P_1 satisfy the strong regularity Assumptions 3.16. For the existence and smoothness of the optimal transport map γ we refer to Remark 3.17.

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

$$f''(t) \geq \kappa W_2^2(P_0, P_1), \quad 0 \leq t \leq 1. \quad (3.73)$$

Indeed, according to condition (3.63), the potential Ψ is κ -uniformly convex. Consequently, by items (i) and (ii) of [Vil03, Theorem 5.15], the internal and potential energies

$$g(t) := \int_{\mathbb{R}^n} p_t(x) \log p_t(x) dx, \quad h(t) := 2 \int_{\mathbb{R}^n} \Psi(x) p_t(x) dx, \quad 0 \leq t \leq 1, \quad (3.74)$$

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

$$g''(t) \geq 0, \quad h''(t) \geq \kappa W_2^2(P_0, P_1), \quad 0 \leq t \leq 1. \quad (3.75)$$

By analogy with Lemma 2.3 we have $f = g + h$, and hence conclude that the relative entropy function f is κ -uniformly displacement convex, i.e., its second derivative satisfies (3.73).

We appeal now to Lemma 3.19, according to which we have

$$f'(0+) = \lim_{t \downarrow 0} \frac{f(t) - f(0)}{t} = \langle \nabla \log \ell_0(X_0), \gamma(X_0) \rangle_{L^2(\nu)}. \quad (3.76)$$

In conjunction with (3.73) and (3.76), the formula $f(1) = f(0) + f'(0+) + \int_0^1 (1-t)f''(t) dt$ now yields (3.64). \square

4. DETAILS AND PROOFS

In this section we complete the proofs of Corollary 3.11 and Proposition 3.14, and provide the proofs of our main results, Theorems 3.6 and 3.8. In fact, all we have to do in order to prove these latter theorems is to apply Itô's formula so as to calculate the dynamics, i.e., the stochastic differentials, of the “pure” and “perturbed” relative entropy processes

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

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

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

$$\log \ell(T-s, X(T-s)) = \log \left(\frac{p(T-s, X(T-s))}{q(X(T-s))} \right), \quad 0 \leq s \leq T, \quad (4.2)$$

as stated in Theorem 3.6, namely

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

for $0 \leq s \leq T$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$. Therefore, in order to motivate and illustrate the derivation of the dynamics (4.3), we first impose the additional assumption $Q(\mathbb{R}^n) < \infty$ (which precludes the case $\Psi \equiv 0$), so as to conform to the setting of [FJ16]. This is done in Subsection 4.2, which mainly serves as motivation; in the remainder of this paper we do not rely on the assumption $Q(\mathbb{R}^n) < \infty$.

4.1. SOME PRELIMINARIES

Our first task is to calculate the dynamics of the time-reversed relative entropy process (4.2) under the probability measure \mathbb{P} . In order to do this, we start by calculating the stochastic differential of the time-reversed canonical coordinate process $(X(T-s))_{0 \leq s \leq T}$ under \mathbb{P} , which is a well-known and classical theme; see e.g. [Föl85, Föl86], [HP86], [Mey94], [Nel01], and [Par86]. For the convenience of the reader we present the theory of time reversal of diffusion processes in Appendix G. The idea of time reversal goes back to the ideas of Boltzmann [Bol96, Bol98a, Bol98b] and Schrödinger [Sch31, Sch32], as well as Kolmogorov [Kol37]. In fact, as we shall recall in Appendix A, the relation between time-reversal of a Brownian motion and the quadratic Wasserstein distance may *in nuce* be traced back to an insight of Bachelier in his thesis [Bac00, Bac06] from 1900; at least, when admitting a good portion of wisdom of hindsight.

Recall that the probability measure \mathbb{P} was defined on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$ so that the canonical coordinate process $(X(t)(\omega))_{t \geq 0} = (\omega(t))_{t \geq 0}$ satisfies the stochastic differential equation (2.1) with initial probability distribution $P(0)$ for $X(0)$ under \mathbb{P} . In other words, the process

$$W(t) = X(t) - X(0) + \int_0^t \nabla \Psi(X(u)) du, \quad t \geq 0 \quad (4.4)$$

defines a Brownian motion of the forward filtration $(\mathcal{F}(t))_{t \geq 0}$ under the probability measure \mathbb{P} , where the integral in (4.4) is to be understood in a pathwise Riemann-Stieltjes sense. Passing to the reverse direction of time, the following classical result is well known to hold under the Assumptions 1.2.

Proposition 4.1. *Under the Assumptions 1.2, we let $T > 0$. The process*

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

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

$$dX(T-s) = \left(\nabla \log p(T-s, X(T-s)) + \nabla \Psi(X(T-s)) \right) ds + d\overline{W}^{\mathbb{P}}(T-s) \quad (4.6)$$

$$= \left(\nabla \log \ell(T-s, X(T-s)) - \nabla \Psi(X(T-s)) \right) ds + d\overline{W}^{\mathbb{P}}(T-s), \quad (4.7)$$

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

We provide proofs and references for this result in Theorems G.2 and G.5 of Appendix G.

Before proving Theorem 3.6 in Subsection 4.3 — as already announced — we digress now to present the following didactic, illuminating and important special case.

4.2. THE CASE $Q(\mathbb{R}^n) < \infty$

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

$$Q(\mathbb{R}^n) = \int_{\mathbb{R}^n} e^{-2\Psi(x)} dx < \infty. \quad (4.8)$$

Under this assumption, the measure Q on the Borel sets of \mathbb{R}^n , introduced in Section 2, is finite and can thus be re-normalized, so as to become a probability measure. In this manner, it induces a probability measure \mathbb{Q} on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$, under which the canonical coordinate process $(X(t)(\omega))_{t \geq 0} = (\omega(t))_{t \geq 0}$ satisfies the stochastic equation (2.1) with initial probability distribution Q for $X(0)$. And because this distribution is invariant, it is also the distribution of $X(t)$ under \mathbb{Q} for every $t \geq 0$.

For the present authors, the eye-opener leading to (4.3) was the subsequent remarkable insight due to Fontbona and Jourdain [FJ16]. This provided us with much of the original motivation to start this line of research. The result right below holds in much greater generality (essentially one only needs the Markovian structure of the process $(X(t))_{t \geq 0}$) but we only state it in the present setting given by (2.1) under the Assumptions 1.2 and $Q(\mathbb{R}^n) = 1$ in (4.8).

Theorem 4.2 (FONTBONA-JOURDAIN). *Under the Assumptions 1.2 and $Q(\mathbb{R}^n) = 1$, we let $T > 0$. The time-reversed likelihood ratio process*

$$\ell(T-s, X(T-s)) = \frac{p(T-s, X(T-s))}{q(X(T-s))}, \quad 0 \leq s \leq T \quad (4.9)$$

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

For the convenience of the reader we recall in Appendix E the surprisingly straightforward proof of Theorem 4.2. Since this result states that the time-reversed likelihood ratio process (4.9) is a \mathbb{Q} -martingale with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$, we will first state the analogue of Proposition 4.1 in terms of the probability measure \mathbb{Q} on the path space $\Omega = \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n)$, which is induced by the invariant probability distribution Q on \mathbb{R}^n .

Proposition 4.3. *Under the Assumptions 1.2 and $Q(\mathbb{R}^n) = 1$, we let $T > 0$. The process*

$$\overline{W}^Q(T-s) := W(T-s) - W(T) + 2 \int_0^s \nabla \Psi(X(T-u)) du \quad (4.10)$$

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

$$dX(T-s) = -\nabla \Psi(X(T-s)) ds + d\overline{W}^Q(T-s), \quad (4.11)$$

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

Again, for the proof of this result, we refer to Theorems G.2 and G.5 of Appendix G.

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

Lemma 4.4. *Under the Assumptions 1.2 and $Q(\mathbb{R}^n) = 1$, we let $T \in (0, \infty)$. For $0 \leq s \leq T$, we have*

$$d\overline{W}^Q(T-s) = \nabla \log \ell(T-s, X(T-s)) ds + d\overline{W}^P(T-s). \quad (4.12)$$

Proof. One just has to compare the equations (4.7) and (4.11). \square

The next corollary is a direct consequence of Theorem 4.2, Proposition 4.3 and Itô's formula.

Corollary 4.5. *Under the Assumptions 1.2 and $Q(\mathbb{R}^n) = 1$, we let $T > 0$. The time-reversed likelihood ratio process (4.9) and its logarithm satisfy respectively the stochastic differential equations*

$$d\ell(T-s, X(T-s)) = \nabla \ell(T-s, X(T-s)) d\overline{W}^Q(T-s) \quad (4.13)$$

and

$$d \log \ell(T-s, X(T-s)) = \left\langle \frac{\nabla \ell(T-s, X(T-s))}{\ell(T-s, X(T-s))}, d\overline{W}^Q(T-s) \right\rangle_{\mathbb{R}^n} - \frac{1}{2} \frac{|\nabla \ell(T-s, X(T-s))|^2}{\ell(T-s, X(T-s))^2} ds, \quad (4.14)$$

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

Proof. To prove (4.13), the decisive insight is provided by Theorem 4.2 due to Fontbona and Jourdain [FJ16]. This implies that the drift term in (4.13) must vanish, so that it suffices to calculate the diffusion term in front of $d\overline{W}^Q(T-s)$ in (4.13); using (4.11), this is an easy task.

We note that the fact that the drift term in (4.13) vanishes can also be obtained from mechanically applying Itô's formula to the process (4.9), and using (4.11) as well as the backwards Kolmogorov equation (4.17) for the likelihood ratio function $\ell(t, x)$ and to observe that all these terms cancel out. But such a procedure does not provide a hint as to why this miracle happens.

Having said this, we apply Itô's formula to the process (4.9) and use Theorem 4.2 to obtain (4.13). Assertion (4.14) follows once again by applying Itô's formula to the logarithm of the process (4.9), and using the dynamics (4.13). \square

We have now all the ingredients in order to compute, under the additional assumption $Q(\mathbb{R}^n) = 1$, the dynamics of the time-reversed relative entropy process (4.2) under the probability measure \mathbb{P} . Indeed, substituting (4.12) into the stochastic equation (4.14), we see that the process (4.2) satisfies the crucial stochastic differential equation (4.3), for $0 \leq s \leq T$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$.

4.3. THE PROOF OF THEOREM 3.6

We drop now the assumption $Q(\mathbb{R}^n) < \infty$, and write the Fokker-Planck equation (3.1) as

$$\partial_t p(t, x) = \frac{1}{2} \Delta p(t, x) + \langle \nabla p(t, x), \nabla \Psi(x) \rangle_{\mathbb{R}^n} + p(t, x) \Delta \Psi(x), \quad t > 0. \quad (4.15)$$

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

$$p(t, x) = \ell(t, x) q(x) = \ell(t, x) e^{-2\Psi(x)}, \quad t \geq 0, \quad (4.16)$$

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

$$\partial_t \ell(t, x) = \frac{1}{2} \Delta \ell(t, x) - \langle \nabla \ell(t, x), \nabla \Psi(x) \rangle_{\mathbb{R}^n}, \quad t > 0, \quad (4.17)$$

a feature consonant with the fact that the dynamics of the likelihood ratio process are most transparent under *time reversal*. This equation will allow us to develop an alternative way of deriving the dynamics (4.3) and proving Theorem 3.6, which does not rely on assumption (4.8) and uses exclusively the probability measure \mathbb{P} , as follows.

Proof of Theorem 3.6. We have to show that the stochastic process $(M(T-s))_{0 \leq s \leq T}$ in (3.23) is a \mathbb{P} -martingale, is bounded in $L^2(\mathbb{P})$, and admits the representation (3.24).

Applying Itô's formula to the time-reversed likelihood ratio process (4.9), and using (4.7) as well as the backwards Kolmogorov equation (4.17) for the likelihood ratio function $\ell(t, x)$, we get the stochastic differential equation

$$\frac{d\ell(T-s, X(T-s))}{\ell(T-s, X(T-s))} = \left\langle \frac{\nabla \ell(T-s, X(T-s))}{\ell(T-s, X(T-s))}, d\overline{W}^{\mathbb{P}}(T-s) \right\rangle_{\mathbb{R}^n} + \frac{|\nabla \ell(T-s, X(T-s))|^2}{\ell(T-s, X(T-s))^2} ds \quad (4.18)$$

as well as its logarithmic version

$$d \log \ell(T-s, X(T-s)) = \langle \nabla \log \ell(T-s, X(T-s)), d\overline{W}^{\mathbb{P}}(T-s) \rangle_{\mathbb{R}^n} + \frac{1}{2} |\nabla \log \ell(T-s, X(T-s))|^2 ds, \quad (4.19)$$

for $0 \leq s \leq T$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$. This equation right above is nothing other than the desired stochastic differential equation (4.3) from the beginning of Section 4. By condition (iv) of Assumptions 1.2, the likelihood ratio function $\ell(t, x)$ is continuous in the time variable t for each $x \in \mathbb{R}^n$. Hence, according to (4.3), the process $(M(T-s))_{0 \leq s \leq T}$ is a continuous local \mathbb{P} -martingale with representation (3.24).

To show that, in fact, this process is a true martingale, we have to rely on the finite free energy condition (1.11) which, in the light of Lemma 2.3, asserts that the relative entropy $H(P(0) | Q)$ is finite. We choose now a non-decreasing sequence of stopping times $(\tau_n)_{n \geq 0}$ with $\tau_n \uparrow T$, such that the stopped process $(M^{\tau_n}(T-s))_{0 \leq s \leq T}$ is a uniformly integrable \mathbb{P} -martingale, for every $n \in \mathbb{N}_0$. Then

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\int_0^T \frac{1}{2} \frac{|\nabla \ell(T-u, X(T-u))|^2}{\ell(T-u, X(T-u))^2} du \right] &= \lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}} \left[\int_0^{\tau_n} \frac{1}{2} \frac{|\nabla \ell(T-u, X(T-u))|^2}{\ell(T-u, X(T-u))^2} du \right] \\ &= \lim_{n \rightarrow \infty} H(P(T-\tau_n) | Q) - H(P(T) | Q) < \infty, \end{aligned} \quad (4.20)$$

where the second equality in (4.20) follows after taking expectations with respect to the probability measure \mathbb{P} in (4.3) at time $s = \tau_n$, and using the fact that the process $(M^{\tau_n}(T-s))_{0 \leq s \leq T}$ is a true \mathbb{P} -martingale with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T}$, for every $n \in \mathbb{N}_0$. From (4.20) we deduce now that the stochastic integral in (4.3) defines an $L^2(\mathbb{P})$ -bounded martingale for $0 \leq s \leq T$.

Summing up, we conclude that the process $(M(T-s))_{0 \leq s \leq T}$ is a $L^2(\mathbb{P})$ -bounded martingale satisfying (3.24). This completes the proof of Theorem 3.6. \square

4.4. THE PROOF OF THEOREM 3.8

The first step in the proof of Theorem 3.8 is to compute the stochastic differentials of the time-reversed perturbed likelihood ratio process

$$\ell^\beta(T-s, X(T-s)) = \frac{p^\beta(T-s, X(T-s))}{q(X(T-s))}, \quad 0 \leq s \leq T-t_0, \quad (4.21)$$

and its logarithm.

By analogy with Proposition 4.1, the following result is well known to hold under suitable regularity conditions, such as Assumptions 1.2. Recall that $(W^\beta(t))_{t \geq t_0}$ denotes the \mathbb{P}^β -Brownian motion (in the forward direction of time) defined in (3.11).

Proposition 4.6. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. The process*

$$\overline{W}^{\mathbb{P}^\beta}(T-s) := W^\beta(T-s) - W^\beta(T) - \int_0^s \nabla \log p^\beta(T-u, X(T-u)) du \quad (4.22)$$

for $0 \leq s \leq T-t_0$, is a Brownian motion of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$ under the probability measure \mathbb{P}^β . Furthermore, the semimartingale decomposition of the time-reversed canonical coordinate process $(X(T-s))_{0 \leq s \leq T-t_0}$ is given by

$$dX(T-s) = \left(\nabla \log p^\beta(T-s, X(T-s)) + (\nabla \Psi + \beta)(X(T-s)) \right) ds + d\overline{W}^{\mathbb{P}^\beta}(T-s) \quad (4.23)$$

$$= \left(\nabla \log \ell^\beta(T-s, X(T-s)) - (\nabla \Psi - \beta)(X(T-s)) \right) ds + d\overline{W}^{\mathbb{P}^\beta}(T-s), \quad (4.24)$$

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

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

Lemma 4.7. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. For $0 \leq s \leq T-t_0$, we have*

$$d(\overline{W}^{\mathbb{P}} - \overline{W}^{\mathbb{P}^\beta})(T-s) = \left(\beta(X(T-s)) + \nabla \log \left(\frac{p^\beta(T-s, X(T-s))}{p(T-s, X(T-s))} \right) \right) ds \quad (4.25)$$

$$= \left(\beta(X(T-s)) + \nabla \log \left(\frac{\ell^\beta(T-s, X(T-s))}{\ell(T-s, X(T-s))} \right) \right) ds. \quad (4.26)$$

Proof. It suffices to compare the equation (4.6) with (4.23). \square

Remark 4.8. Later we shall apply Lemma 4.7 to the situation when s is close to $T-t_0$. In this case the logarithmic gradients in (4.25) and (4.26) will become small in view of $p^\beta(t_0, \cdot) = p(t_0, \cdot)$, so that these logarithmic gradients will disappear in the limit $s \uparrow T-t_0$; see also Lemma 4.12 below. By contrast, the term $\beta(X(T-s))$ will not disappear in the limit $s \uparrow T-t_0$. Rather, it will tend to the random variable $\beta(X(t_0))$, which plays an important role in distinguishing between (3.53) and (3.54) in Proposition 3.14. \diamond

Next, by analogy with Subsection 4.3, for $t > t_0$, we write the perturbed Fokker-Planck equation (3.9) as

$$\partial_t p^\beta(t, x) = \frac{1}{2} \Delta p^\beta(t, x) + \langle \nabla p^\beta(t, x), \nabla \Psi(x) + \beta(x) \rangle_{\mathbb{R}^n} + p^\beta(t, x) (\Delta \Psi(x) + \operatorname{div} \beta(x)). \quad (4.27)$$

Using the relation

$$p^\beta(t, x) = \ell^\beta(t, x) q(x) = \ell^\beta(t, x) e^{-2\Psi(x)}, \quad t \geq t_0, \quad (4.28)$$

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

$$\begin{aligned} \partial_t \ell^\beta(t, x) &= \frac{1}{2} \Delta \ell^\beta(t, x) + \langle \nabla \ell^\beta(t, x), \beta(x) - \nabla \Psi(x) \rangle_{\mathbb{R}^n} \\ &\quad + \ell^\beta(t, x) \left(\operatorname{div} \beta(x) - 2 \langle \beta(x), \nabla \Psi(x) \rangle_{\mathbb{R}^n} \right), \quad t > t_0; \end{aligned} \quad (4.29)$$

this is the analogue of the backwards Kolmogorov equation (4.17) in this ‘‘perturbed’’ context.

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

Lemma 4.9. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. The time-reversed perturbed likelihood ratio process (4.21) and its logarithm satisfy respectively the stochastic differential equations*

$$\begin{aligned} \frac{d\ell^\beta(T-s, X(T-s))}{\ell^\beta(T-s, X(T-s))} &= \left(\langle \beta, 2\nabla \Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \right) (X(T-s)) ds \\ &\quad + \frac{|\nabla \ell^\beta(T-s, X(T-s))|^2}{\ell^\beta(T-s, X(T-s))^2} ds + \left\langle \frac{\nabla \ell^\beta(T-s, X(T-s))}{\ell^\beta(T-s, X(T-s))}, d\overline{W}^{\mathbb{P}^\beta}(T-s) \right\rangle_{\mathbb{R}^n} \end{aligned} \quad (4.30)$$

and

$$\begin{aligned} d \log \ell^\beta(T-s, X(T-s)) &= \left(\langle \beta, 2\nabla \Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \right) (X(T-s)) ds \\ &\quad + \frac{1}{2} \frac{|\nabla \ell^\beta(T-s, X(T-s))|^2}{\ell^\beta(T-s, X(T-s))^2} ds + \left\langle \frac{\nabla \ell^\beta(T-s, X(T-s))}{\ell^\beta(T-s, X(T-s))}, d\overline{W}^{\mathbb{P}^\beta}(T-s) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (4.31)$$

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

Proof. The equations (4.30), (4.31) follow from Itô’s formula together with (4.24), (4.29). \square

We have assembled now all the ingredients for the proof of Theorem 3.8.

Proof of Theorem 3.8. Formally, the stochastic differential in (4.31) of the time-reversed perturbed likelihood ratio process (4.21) amounts to the conclusions (3.27) – (3.29) of Theorem 3.8. But we still have to substantiate the claim, that the stochastic process $(M^\beta(T-s))_{0 \leq s \leq T-t_0}$ defined in (3.28) with representation (3.29), indeed yields a \mathbb{P}^β -martingale of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$, which is bounded in $L^2(\mathbb{P}^\beta)$.

In order to do this, we split the perturbed Fisher information process $(F^\beta(T-s))_{0 \leq s \leq T-t_0}$ defined in (3.27) into two parts. First, we define the process

$$F^{\beta, L}(T-s) := \int_0^s \left(\langle \beta, 2\nabla \Psi \rangle_{\mathbb{R}^n} - \operatorname{div} \beta \right) (X(T-u)) du \quad (4.32)$$

for $0 \leq s \leq T - t_0$. The letter L stands for ‘‘Lipschitz’’, as the process (4.32) satisfies

$$\|F^{\beta, L}(T-s_1) - F^{\beta, L}(T-s_2)\|_{L^\infty(\mathbb{P}^\beta)} \leq L |s_1 - s_2| \quad (4.33)$$

for $0 \leq s_1, s_2 \leq T - t_0$. Here the constant $L \in (0, \infty)$ only depends on the vector field β and the restriction of the potential Ψ to the compact support of β . Consequently, the process (4.32)

has uniformly Lipschitz continuous paths. The remaining part of the perturbed cumulative Fisher information process (3.27), is the difference

$$F^\beta(T-s) - F^{\beta,L}(T-s) = \int_0^s \frac{1}{2} \frac{|\nabla \ell^\beta(T-u, X(T-u))|^2}{\ell^\beta(T-u, X(T-u))^2} du, \quad 0 \leq s \leq T-t_0, \quad (4.34)$$

and we define

$$M^{\beta,\text{sub}}(T-s) := \left(\log \ell^\beta(T-s, X(T-s)) - \log \ell^\beta(T, X(T)) \right) - F^{\beta,L}(T-s), \quad 0 \leq s \leq T-t_0. \quad (4.35)$$

Invoking (4.31), this process $(M^{\beta,\text{sub}}(T-s))_{0 \leq s \leq T-t_0}$ is seen to be a submartingale, as it has dynamics

$$dM^{\beta,\text{sub}}(T-s) = \left\langle \frac{\nabla \ell^\beta(T-s, X(T-s))}{\ell^\beta(T-s, X(T-s))}, d\overline{W}^{\mathbb{P}^\beta}(T-s) \right\rangle_{\mathbb{R}^n} + \frac{1}{2} \frac{|\nabla \ell^\beta(T-s, X(T-s))|^2}{\ell^\beta(T-s, X(T-s))^2} ds, \quad (4.36)$$

for $0 \leq s \leq T-t_0$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$. According to condition (v) of Assumptions 1.2, the perturbed likelihood ratio function $\ell^\beta(t, x)$ is continuous in the time variable $t \geq t_0$ for each $x \in \mathbb{R}^n$. Therefore the above equation (4.36) implies that the process $(M^{\beta,\text{sub}}(T-s))_{0 \leq s \leq T-t_0}$ is a local \mathbb{P}^β -submartingale.

Finally, we show that the process $(M^{\beta,\text{sub}}(T-s))_{0 \leq s \leq T-t_0}$ is a \mathbb{P}^β -submartingale of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$, whose martingale part is bounded in $L^2(\mathbb{P}^\beta)$ and whose drift part is bounded in $L^1(\mathbb{P}^\beta)$. Because it is a local \mathbb{P}^β -submartingale, we can choose a non-decreasing sequence of stopping times $(\tau_n)_{n \geq 0}$ with $\tau_n \uparrow T-t_0$, such that the stopped process $((M^{\beta,\text{sub}})^{\tau_n}(T-s))_{0 \leq s \leq T-t_0}$ is a uniformly integrable \mathbb{P}^β -submartingale of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$, for every $n \in \mathbb{N}_0$. Consequently, after taking expectations with respect to the probability measure \mathbb{P}^β in (4.31) at time $s = \tau_n$, we obtain

$$H(P^\beta(T-\tau_n) | \mathbb{Q}) - H(P^\beta(T) | \mathbb{Q}) = \mathbb{E}_{\mathbb{P}^\beta} \left[\int_0^{\tau_n} \frac{1}{2} \frac{|\nabla \ell^\beta(T-u, X(T-u))|^2}{\ell^\beta(T-u, X(T-u))^2} du \right] \quad (4.37)$$

$$+ \mathbb{E}_{\mathbb{P}^\beta} \left[\int_0^{\tau_n} \left(\langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n} - \text{div } \beta \right) (X(T-u)) du \right]. \quad (4.38)$$

Passing to the limit as $n \rightarrow \infty$ yields

$$H(P^\beta(t_0) | \mathbb{Q}) - H(P^\beta(T) | \mathbb{Q}) = \mathbb{E}_{\mathbb{P}^\beta} \left[\int_0^{T-t_0} \frac{1}{2} \frac{|\nabla \ell^\beta(T-u, X(T-u))|^2}{\ell^\beta(T-u, X(T-u))^2} du \right] \quad (4.39)$$

$$+ \mathbb{E}_{\mathbb{P}^\beta} \left[\int_0^{T-t_0} \left(\langle \beta, 2 \nabla \Psi \rangle_{\mathbb{R}^n} - \text{div } \beta \right) (X(T-u)) du \right]. \quad (4.40)$$

Recall that the relative entropy $H(P^\beta(t_0) | \mathbb{Q}) = H(P(t_0) | \mathbb{Q})$ is finite, and that $H(P^\beta(T) | \mathbb{Q})$ cannot take the value $-\infty$; see Lemma 3.3 and Appendix C. Furthermore, the quantity in (4.40) is finite, due to the compact support of the smooth perturbation β . Hence both $H(P^\beta(T) | \mathbb{Q})$ and the expression on the right-hand side of (4.39) must be finite, i.e.,

$$\mathbb{E}_{\mathbb{P}^\beta} [F^\beta(t_0)] = \mathbb{E}_{\mathbb{P}^\beta} \left[\int_0^{T-t_0} \frac{1}{2} \frac{|\nabla \ell^\beta(T-u, X(T-u))|^2}{\ell^\beta(T-u, X(T-u))^2} du \right] < \infty. \quad (4.41)$$

From (4.41) we deduce that the process $(M^{\beta,\text{sub}}(T-s))_{0 \leq s \leq T-t_0}$ is a $L^2(\mathbb{P}^\beta)$ -bounded submartingale.

Summing up, we conclude that the stochastic process $(M^\beta(T-s))_{0 \leq s \leq T-t_0}$, defined in (3.28) and admitting the (3.29), yields a \mathbb{P}^β -martingale of the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$. This martingale is bounded in $L^2(\mathbb{P}^\beta)$. \square

4.5. SOME USEFUL LEMMAS

In this subsection we collect some useful lemmas, which are needed in order to complete the proofs of Corollary 3.11 and Proposition 3.14.

First, let us recall the probability density function $(t, x) \mapsto p(t, x)$ from (4.16), its perturbed version $(t, x) \mapsto p^\beta(t, x)$ from (4.28), and the respective likelihood ratios

$$\ell(t, x) = \frac{p(t, x)}{q(x)}, \quad \ell^\beta(t, x) = \frac{p^\beta(t, x)}{q(x)} \quad (4.42)$$

from (2.2), (3.25), respectively. We introduce the ‘‘perturbed-to-unperturbed’’ ratio

$$Y^\beta(t, x) := \frac{\ell^\beta(t, x)}{\ell(t, x)} = \frac{p^\beta(t, x)}{p(t, x)}, \quad (t, x) \in [t_0, \infty) \times \mathbb{R}^n. \quad (4.43)$$

We recall the backward Kolmogorov-type equations (4.17), (4.29). These lead to the partial differential equation

$$\begin{aligned} \partial_t Y^\beta(t, x) &= \frac{1}{2} \Delta Y^\beta(t, x) + \langle \nabla Y^\beta(t, x), \beta(x) + \nabla \log p(t, x) + \nabla \Psi(x) \rangle_{\mathbb{R}^n} \\ &\quad + Y^\beta(t, x) \left(\operatorname{div} \beta(x) + \langle \beta(x), \nabla \log p(t, x) \rangle_{\mathbb{R}^n} \right), \quad t > t_0, \end{aligned} \quad (4.44)$$

with $Y^\beta(t_0, \cdot) = 1$, for the ratio in (4.43). In conjunction with (4.6), this equation leads to the following backward dynamics.

Lemma 4.10. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. The time-reversed ratio process $(Y^\beta(T - s, X(T - s)))_{0 \leq s \leq T - t_0}$ and its logarithm satisfy respectively the stochastic differential equations*

$$\begin{aligned} \frac{dY^\beta(T - s, X(T - s))}{Y^\beta(T - s, X(T - s))} &= \left\langle \frac{\nabla Y^\beta(T - s, X(T - s))}{Y^\beta(T - s, X(T - s))}, d\overline{W}^\mathbb{P}(T - s) - \beta(X(T - s)) ds \right\rangle_{\mathbb{R}^n} \\ &\quad - \left(\operatorname{div} \beta(X(T - s)) + \langle \beta(X(T - s)), \nabla \log p(T - s, X(T - s)) \rangle_{\mathbb{R}^n} \right) ds \end{aligned} \quad (4.45)$$

and

$$\begin{aligned} d \log Y^\beta(T - s, X(T - s)) &= \left\langle \frac{\nabla Y^\beta(T - s, X(T - s))}{Y^\beta(T - s, X(T - s))}, d\overline{W}^\mathbb{P}(T - s) - \beta(X(T - s)) ds \right\rangle_{\mathbb{R}^n} \\ &\quad - \left(\operatorname{div} \beta(X(T - s)) + \langle \beta(X(T - s)), \nabla \log p(T - s, X(T - s)) \rangle_{\mathbb{R}^n} \right) ds \\ &\quad - \frac{1}{2} \frac{|\nabla Y^\beta(T - s, X(T - s))|^2}{Y^\beta(T - s, X(T - s))^2} ds, \end{aligned} \quad (4.46)$$

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

We also need a preliminary control on $Y^\beta(\cdot, \cdot)$, which will be refined in Lemma 4.11 below.

Lemma 4.11. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. There is a real constant $C > 1$ such that*

$$\frac{1}{C} \leq Y^\beta(t, x) \leq C, \quad (t, x) \in [t_0, T] \times \mathbb{R}^n. \quad (4.47)$$

Proof. In the forward direction of time, the canonical coordinate process $(X(t))_{t_0 \leq t \leq T}$ on the path space $\Omega = \mathcal{C}([t_0, T]; \mathbb{R}^n)$ satisfies the stochastic equations (2.1) and (3.11) with initial probability distribution $P(t_0)$ under the probability measures \mathbb{P} and \mathbb{P}^β , respectively. Hence, the \mathbb{P} -Brownian motion $(W(t))_{t_0 \leq t \leq T}$ from (2.1) can be represented as

$$W(t) - W(t_0) = W^\beta(t) - W^\beta(t_0) - \int_{t_0}^t \beta(X(u)) \, du, \quad t_0 \leq t \leq T, \quad (4.48)$$

where $(W^\beta(t))_{t_0 \leq t \leq T}$ is the \mathbb{P}^β -Brownian motion appearing in (3.11). By the Girsanov theorem, this amounts to

$$Z(t) := \frac{d\mathbb{P}^\beta}{d\mathbb{P}} \Big|_{\mathcal{F}(t)} = \exp \left(- \int_{t_0}^t \langle \beta(X(u)), dW(u) \rangle_{\mathbb{R}^n} - \frac{1}{2} \int_{t_0}^t |\beta(X(u))|^2 \, du \right) \quad (4.49)$$

for $t_0 \leq t \leq T$.

Now, for each $(t, x) \in [t_0, T] \times \mathbb{R}^n$, the ratio $Y^\beta(t, x) = p^\beta(t, x)/p(t, x)$ equals the conditional expectation of the random variable (4.49) with respect to the probability measure \mathbb{P} , where we condition on $X(t) = x$; to wit,

$$Y^\beta(t, x) = \mathbb{E}_{\mathbb{P}}[Z(t) \mid X(t) = x], \quad (t, x) \in [t_0, T] \times \mathbb{R}^n. \quad (4.50)$$

Therefore, in order to obtain the estimate (4.47), it suffices to show that the log-density process $(\log Z(t))_{t_0 \leq t \leq T}$ is uniformly bounded. Since the perturbation β is smooth and has compact support, the Lebesgue integral inside the exponential of (4.49) is uniformly bounded, as required. In order to handle the stochastic integral with respect to the \mathbb{P} -Brownian motion $(W(u))_{t_0 \leq u \leq t}$ inside the exponential (4.49), we invoke the assumption that the vector field β equals the gradient of a potential $B: \mathbb{R}^n \rightarrow \mathbb{R}$, which is of class $C^\infty(\mathbb{R}^n; \mathbb{R})$ and has compact support. According to Itô's formula and (2.1), we can express the stochastic integral appearing in (4.49) as

$$\int_{t_0}^t \langle \beta(X(u)), dW(u) \rangle_{\mathbb{R}^n} = B(X(t)) - B(X(t_0)) + \int_{t_0}^t \left(\langle \beta, \nabla \Psi \rangle_{\mathbb{R}^n} - \frac{1}{2} \operatorname{div} \beta \right) (X(u)) \, du \quad (4.51)$$

for $t_0 \leq t \leq T$. At this stage it becomes obvious that the expression of (4.51) is uniformly bounded. This completes the proof of Lemma 4.11. \square

The following Lemma 4.12 provides the crucial estimates (3.36) and (3.37), needed in the proof of Corollary 3.11 from Theorem 3.8, and of Proposition 3.14.

Lemma 4.12. *Under the Assumptions 1.2, we let $t_0 \geq 0$ and $T > t_0$. There is a constant $C > 0$ such that*

$$|Y^\beta(T - s, x) - 1| \leq C(T - t_0 - s), \quad (4.52)$$

as well as

$$\mathbb{E}_{\mathbb{P}} \left[\int_s^{T-t_0} \left| \nabla \log Y^\beta(T - u, X(T - u)) \right|^2 \, du \mid X(T - s) = x \right] \leq C(T - t_0 - s)^2, \quad (4.53)$$

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

$$\lim_{s \uparrow T-t_0} \frac{\log Y^\beta(T - s, x)}{T - t_0 - s} = \operatorname{div} \beta(x) + \langle \beta(x), \nabla \log p(t_0, x) \rangle_{\mathbb{R}^n}, \quad (4.54)$$

where the fraction on the left-hand side of (4.54) is uniformly bounded on $[0, T - t_0] \times \mathbb{R}^n$.

Remark 4.13. The pointwise limiting assertion (4.54) is the deterministic analogue of the trajectorial relation (3.55) from Proposition 3.14. In Subsection 4.6 below we will prove that the limiting assertion (3.55) holds in the norm of L^1 under both \mathbb{P} and \mathbb{P}^β , and is valid for all $t_0 > 0$. \diamond

Proof. As $\log Y^\beta = \log \ell^\beta - \log \ell$, we obtain from (4.20), (4.41) and (4.47) that the martingale part of the process in (4.46) is bounded in $L^2(\mathbb{P})$, i.e.,

$$\mathbb{E}_{\mathbb{P}} \left[\int_0^{T-t_0} \frac{|\nabla Y^\beta(T-u, X(T-u))|^2}{Y^\beta(T-u, X(T-u))^2} du \right] < \infty. \quad (4.55)$$

Once again using (4.47), we compare $\nabla Y^\beta/Y^\beta$ with ∇Y^β to see that (4.55) also implies

$$\mathbb{E}_{\mathbb{P}} \left[\int_0^{T-t_0} |\nabla Y^\beta(T-u, X(T-u))|^2 du \right] < \infty. \quad (4.56)$$

According to (4.45), the time-reversed ratio process $(Y^\beta(T-s, X(T-s)))_{0 \leq s \leq T-t_0}$ satisfies the stochastic differential equation

$$\begin{aligned} dY^\beta(T-s, X(T-s)) = & \left\langle \nabla Y^\beta(T-s, X(T-s)), d\overline{W}^\mathbb{P}(T-s) - \beta(X(T-s)) ds \right\rangle_{\mathbb{R}^n} \\ & - Y^\beta(T-s, X(T-s)) \left(\operatorname{div} \beta(X(T-s)) + \left\langle \beta(X(T-s)), \nabla \log p(T-s, X(T-s)) \right\rangle_{\mathbb{R}^n} \right) ds \end{aligned} \quad (4.57)$$

for $0 \leq s \leq T-t_0$, with respect to the backwards filtration $(\mathcal{G}(T-s))_{0 \leq s \leq T-t_0}$. In view of (4.56), the martingale part in (4.57) is bounded in $L^2(\mathbb{P})$. As regards the drift terms of this equation, we observe that it vanishes when $X(T-s)$ takes values outside the compact support of the smooth vector field β . Consequently, the drift terms are bounded, i.e., the constant

$$C_1 := \sup_{\substack{t_0 \leq t \leq T, \\ y \in \mathbb{R}^n}} \left| -Y^\beta(t, y) \left(\operatorname{div} \beta(y) + \left\langle \beta(y), \nabla \log p(t, y) + \frac{\nabla Y^\beta(t, y)}{Y^\beta(t, y)} \right\rangle_{\mathbb{R}^n} \right) \right| \quad (4.58)$$

is finite, and the processes

$$Y^\beta(T-s, X(T-s)) + C_1 s \quad \text{and} \quad Y^\beta(T-s, X(T-s)) - C_1 s \quad (4.59)$$

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

$$\left| Y^\beta(T-s, x) - \mathbb{E}_{\mathbb{P}} \left[Y^\beta(t_0, X(t_0)) \mid X(T-s) = x \right] \right| \leq C_1 (T-t_0-s) \quad (4.60)$$

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

$$|Y^\beta(T-s, x) - 1| \leq C_1 (T-t_0-s). \quad (4.61)$$

Now we turn our attention to the second estimate (4.53). We fix $0 \leq s \leq T-t_0$ and $x \in \mathbb{R}^n$. By means of the stochastic differentials in (4.46) and (4.57), we find that the expression

$$\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\int_s^{T-t_0} |\nabla \log Y^\beta(T-u, X(T-u))|^2 du \mid X(T-s) = x \right] \quad (4.62)$$

is equal to

$$\log Y^\beta(T-s, x) - Y^\beta(T-s, x) + 1 + \mathbb{E}_{\mathbb{P}} \left[\int_s^{T-t_0} G(T-u, X(T-u)) du \mid X(T-s) = x \right], \quad (4.63)$$

where we have set

$$G(t, y) := (Y^\beta(t, y) - 1) \left(\operatorname{div} \beta(y) + \left\langle \beta(y), \nabla \log p(t, y) + \frac{\nabla Y^\beta(t, y)}{Y^\beta(t, y)} \right\rangle_{\mathbb{R}^n} \right) \quad (4.64)$$

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

$$C_2 := \sup_{\substack{t_0 \leq t \leq T, \\ y \in \mathbb{R}^n}} \left| \operatorname{div} \beta(y) + \left\langle \beta(y), \nabla \log p(t, y) + \frac{\nabla Y^\beta(t, y)}{Y^\beta(t, y)} \right\rangle_{\mathbb{R}^n} \right| \quad (4.65)$$

and using the just proved estimate (4.61), we see that the absolute value of the conditional expectation appearing in (4.63) can be bounded by $C_1 C_2 (T - t_0 - s)^2$. In order to handle the remaining terms of (4.63), we apply the elementary inequality $\log p \leq p - 1$, which is valid for all $p > 0$, and obtain

$$\log Y^\beta(T - s, x) - Y^\beta(T - s, x) + 1 \leq 0. \quad (4.66)$$

This implies that the expression of (4.62) is bounded by $C_1 C_2 (T - t_0 - s)^2$, which establishes the second estimate (4.53). We also note that the elementary inequality (4.66) in conjunction with the estimate (4.61) shows that

$$\log Y^\beta(T - s, x) \leq C_1 (T - t_0 - s) \quad (4.67)$$

for all $0 \leq s \leq T - t_0$ and $x \in \mathbb{R}^n$; this implies that the fraction on the left-hand side of (4.54) is uniformly bounded on $[0, T - t_0] \times \mathbb{R}^n$.

Regarding the limiting assertion (4.54), we fix $t_0 > 0$, $x \in \mathbb{R}^n$ and $0 \leq s \leq T - t_0$, and take conditional expectations with respect to $X(T - s) = x$ in the integral version of the stochastic differential (4.46). On account of (4.55), the stochastic integral with respect to the \mathbb{P} -Brownian motion $(\overline{W}^\mathbb{P}(T - s))_{0 \leq s \leq T}$ in (4.46) vanishes. Dividing by $T - t_0 - s$ and passing to the limit as $s \uparrow T - t_0$, we can use the estimate (4.53) to observe that the expression in the third line of (4.46) vanishes in the limit. After applying the Cauchy–Schwarz inequality, we note that the perturbation β appearing in the first line of (4.46) can be uniformly bounded. The remaining term of this line is

$$\frac{1}{T - t_0 - s} \int_s^{T - t_0} \left| \nabla \log Y^\beta(T - u, X(T - u)) \right| du. \quad (4.68)$$

By conditions (iv), (v) of Assumptions 1.2, the function $(t, x) \mapsto \nabla \log Y^\beta(t, x)$ is continuous on $(0, \infty) \times \mathbb{R}^n$, thus the expression in (4.68) is uniformly bounded on the rectangle $[0, T - t_0] \times \operatorname{supp} \beta$. As $\log Y^\beta(t_0, \cdot) = 0$, it converges \mathbb{P} -almost surely to zero, hence also

$$\lim_{s \uparrow T - t_0} \mathbb{E}_\mathbb{P} \left[\frac{1}{T - t_0 - s} \int_s^{T - t_0} \left| \nabla \log Y^\beta(T - u, X(T - u)) \right| du \mid X(T - s) = x \right] = 0. \quad (4.69)$$

Finally, using continuity and uniform boundedness once again, the conditional expectations of the normalized integrals over the second line of (4.46) converge to the right-hand side of (4.54), as claimed. \square

Remark 4.14. The above Lemma 4.12 justifies the estimates (3.36) and (3.37), which we have used in the proof of Corollary 3.11. They were the crucial ingredients in the effort to show that the exceptional sets considered for the relevant limits do not change when passing from the unperturbed to the perturbed equations. It is now time to come back to this technical issue.

As a general observation, we stress that no worries about limits of difference quotients arise as long as we remain in the realm of an *integral formulation* of our results, as opposed to passing to a

differential formulation. It is precisely the spirit of our basic trajectorial Theorems 3.6 and 3.8, that they are naturally formulated in *integral terms*.

We also remark that the problem of exceptional points does not arise if we impose regularity assumptions strong enough, so that the limiting assertions (3.31) and (3.34) are valid *for all* $t_0 > 0$, or even *for all* $t_0 \geq 0$, instead of *for Lebesgue-almost every* $t_0 \geq 0$. For example, this follows if we impose, in addition to Assumptions 1.2, the *a priori* assumption that the relative Fisher information function $t \mapsto I(P(t) | Q)$ is continuous on $(0, \infty)$, or continuous on $[0, \infty)$, respectively.

Having made these general observations, let us now be more technical and have a precise look at the exceptional sets in the framework of the regularity codified by Assumptions 1.2. In the proof of Corollary 3.10 we have deduced from the Lebesgue differentiation theorem that the generalized de Bruijn identity (3.31) is valid except for a set of Lebesgue measure zero. We denote by N^H the set of exceptional points $t_0 \geq 0$ for which the the one-sided limiting assertion from the right, corresponding to (3.31), fails. We have shown in the proof of Corollary 3.11 that the limiting assertion (3.34) is valid if and only if $t_0 \in \mathbb{R}_+ \setminus N^H$. In other words, the limits from the right in (3.31) and (3.34) have the same exceptional sets. Furthermore, in the proofs of Propositions 3.12 and 3.14 we have seen that the limiting assertions therein are valid for $t_0 \in \mathbb{R}_+ \setminus N^H$. \diamond

Remark 4.15. Let us now pass to the limits of the difference quotients pertaining to the Wasserstein distance. We denote by N^W the set of exceptional points $t_0 \geq 0$ for which the the one-sided limiting assertion from the right, corresponding to (3.32), fails. In other words, for $t_0 \in \mathbb{R}_+ \setminus N^W$ we have

$$\lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0} = \frac{1}{2} \left(\mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right] \right)^{1/2}. \quad (4.70)$$

The validity of (4.70) follows from the validity of the limiting assertion

$$\lim_{t \downarrow t_0} \frac{1}{t - t_0} \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \left(-\frac{1}{2} \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} \right) du \right] = -\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right] \quad (4.71)$$

as well as

$$\lim_{t \downarrow t_0} \frac{1}{2} \left(\frac{1}{t - t_0} \mathbb{E}_{\mathbb{P}} \left[\int_{t_0}^t \frac{|\nabla \ell(u, X(u))|^2}{\ell(u, X(u))^2} du \right] \right)^{1/2} = \lim_{t \downarrow t_0} \frac{W_2(P(t), P(t_0))}{t - t_0}. \quad (4.72)$$

The limiting assertion (4.71) is tantamount to the the one-sided limiting assertion from the right, corresponding to (3.31), i.e., to the assertion that $t_0 \in \mathbb{R}_+ \setminus N^H$. In the proof of Corollary 3.11 we have seen that these equivalent statements are invariant under perturbations β , in the sense that the corresponding perturbed limiting assertion (3.34) continues to be valid for the same points $t_0 \in \mathbb{R}_+ \setminus N^H$. One can apply the estimates of Lemma 4.12 to show that the limit in (3.35) has the same exceptional points as the limit in (3.32). Put another way, for $t_0 \in \mathbb{R}_+ \setminus N^W$ we have

$$\lim_{t \downarrow t_0} \frac{W_2(P^\beta(t), P^\beta(t_0))}{t - t_0} = \frac{1}{2} \left(\mathbb{E}_{\mathbb{P}} \left[\left| \frac{\nabla \ell(t_0, X(t_0))}{\ell(t_0, X(t_0))} + 2\beta(X(t_0)) \right|^2 \right] \right)^{1/2}. \quad (4.73)$$

Summing up, not only the limiting assertions (3.31), (3.32) hold for $t_0 \in \mathbb{R}^+ \setminus N$, where we have set $N := N^H \cup N^W$, but also the limiting assertions (3.34), (3.35) belonging to the perturbed situation are valid for those points t_0 . \diamond

4.6. COMPLETING THE PROOF OF PROPOSITION 3.14

On account of the preparations in Subsection 4.5 above, we are now able to complete the proof of Proposition 3.14, by establishing the remaining limiting assertions (3.54) and (3.55) therein.

Proof of the assertion (3.55) in Proposition 3.14: Let $t_0 > 0$ and select $T > t_0$. Using the notation of (4.43) above, we have to calculate the limit

$$\lim_{s \uparrow T-t_0} \frac{\log Y^\beta(T-s, X(T-s))}{T-t_0-s}. \quad (4.74)$$

Fix $0 \leq s \leq T-t_0$. According to the integral version of the stochastic differential (4.46), the fraction in (4.74) is equal to the sum of the following four normalized integral terms, whose behavior as $s \uparrow T-t_0$ we will study separately below. By conditions (iv), (v) of Assumptions 1.2, the function $(t, x) \mapsto \nabla \log Y^\beta(t, x)$ is continuous on $(0, \infty) \times \mathbb{R}^n$, thus the first expression

$$\frac{1}{T-t_0-s} \int_s^{T-t_0} \left(\operatorname{div} \beta(X(T-u)) + \left\langle \beta(X(T-u)), \nabla \log p(T-u, X(T-u)) \right\rangle_{\mathbb{R}^n} \right) du \quad (4.75)$$

is uniformly bounded on $[0, T-t_0] \times \operatorname{supp} \beta$. Using continuity and uniform boundedness, we conclude that (4.75) converges \mathbb{P} -almost surely as well as in $L^1(\mathbb{P})$ to the right-hand side of (3.55), as required. Thus it remains to show that the three remaining terms converge to zero. Using continuity and uniform boundedness once again, we see that the second integral term

$$\frac{1}{T-t_0-s} \int_s^{T-t_0} \left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}, \beta(X(T-u)) \right\rangle_{\mathbb{R}^n} du \quad (4.76)$$

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

$$\frac{1}{T-t_0-s} \int_s^{T-t_0} \frac{1}{2} \frac{|\nabla Y^\beta(T-u, X(T-u))|^2}{Y^\beta(T-u, X(T-u))^2} du \quad (4.77)$$

converges \mathbb{P} -almost surely to zero. Furthermore, owing to Lemma 4.12, there is a constant $C > 0$ such that

$$\mathbb{E}_{\mathbb{P}} \left[\frac{1}{T-t_0-s} \int_s^{T-t_0} \frac{|\nabla Y^\beta(T-u, X(T-u))|^2}{Y^\beta(T-u, X(T-u))^2} du \right] \leq C(T-t_0-s) \quad (4.78)$$

holds for all $0 \leq s \leq T-t_0$, which implies that (4.77) converges to zero also in $L^1(\mathbb{P})$. The fourth and last term is the stochastic integral

$$-\frac{1}{T-t_0-s} \int_s^{T-t_0} \left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}, d\overline{W}^{\mathbb{P}}(T-u) \right\rangle_{\mathbb{R}^n}. \quad (4.79)$$

The expression (4.77) converges to zero \mathbb{P} -almost surely, and according to (4.78) we have

$$\mathbb{E}_{\mathbb{P}} \left[\frac{1}{(T-t_0-s)^2} \int_s^{T-t_0} \frac{|\nabla Y^\beta(T-u, X(T-u))|^2}{Y^\beta(T-u, X(T-u))^2} du \right] \leq C. \quad (4.80)$$

By means of the Itô isometry, we deduce that

$$\lim_{s \uparrow T-t_0} \mathbb{E}_{\mathbb{P}} \left[\left(\frac{1}{T-t_0-s} \int_s^{T-t_0} \left\langle \frac{\nabla Y^\beta(T-u, X(T-u))}{Y^\beta(T-u, X(T-u))}, d\overline{W}^{\mathbb{P}}(T-u) \right\rangle_{\mathbb{R}^n} \right)^2 \right] = 0. \quad (4.81)$$

In other words, the normalized stochastic integral of (4.79) converges to zero in $L^2(\mathbb{P})$.

Summing up, we have shown that the limiting assertion (3.55) holds in the norm of $L^1(\mathbb{P})$ and is valid for all $t_0 > 0$. As we have seen in Lemma 4.11, the probability measures \mathbb{P} and \mathbb{P}^β are equivalent, the Radon-Nikodým derivatives $\frac{d\mathbb{P}^\beta}{d\mathbb{P}}$ and $\frac{d\mathbb{P}}{d\mathbb{P}^\beta}$ are bounded on the σ -algebra $\mathcal{F}(T) = \mathcal{G}(0)$, and therefore convergence in $L^1(\mathbb{P})$ is equivalent to convergence in $L^1(\mathbb{P}^\beta)$. This completes the proof of the limiting assertion (3.55). \square

Proof of the assertion (3.54) in Proposition 3.14: This is proved in very much the same way, as assertions (3.53), (3.55). The only novelty here, is the use of (4.25) to pass to the \mathbb{P} -Brownian motion $(\overline{W}^\mathbb{P}(T-s))_{0 \leq s \leq T-t_0}$ from the \mathbb{P}^β -Brownian motion $(\overline{W}^{\mathbb{P}^\beta}(T-s))_{0 \leq s \leq T-t_0}$, and the reliance on (4.41) to ensure that the resulting stochastic integral is a (square-integrable) \mathbb{P} -martingale. We leave the details to the care of the diligent reader. \square

4.7. THE DYNAMICS IN THE FORWARD DIRECTION OF TIME

For the sake of completeness, we calculate now the stochastic differentials of the relative entropy process and its perturbed counterpart of (4.1) also in the forward direction of time, under the measures \mathbb{P} and \mathbb{P}^β , respectively. It will turn out that we are able to derive Theorems 3.1 and 3.4 also by applying Itô's formula in the forward direction of time. But the relations between these theorems and the stochastic differentials will become less transparent than in reverse time. One may still take expectations, but to obtain Theorems 3.1 and 3.4 one also has to rely on integration by parts. Furthermore, the second derivatives $\Delta \ell$ and $\Delta \ell^\beta$ show up in the forward direction of time, which on the contrary did not appear in the backward direction.

We first compute the differentials of the likelihood ratio process

$$\ell(t, X(t)) = \frac{p(t, X(t))}{q(X(t))}, \quad t \geq 0 \quad (4.82)$$

and its logarithm in the forward direction of time. We start by recalling the backwards Kolmogorov equation (4.17). With its help, we can compute the forward dynamics of the likelihood ratio process (4.82) in the following manner.

Lemma 4.16. *Under the Assumptions 1.2, the likelihood ratio process (4.82) and its logarithm satisfy respectively the stochastic differential equations*

$$d\ell(t, X(t)) = \Delta \ell(t, X(t)) dt + \left\langle \nabla \ell(t, X(t)), dW(t) - 2 \nabla \Psi(X(t)) dt \right\rangle_{\mathbb{R}^n} \quad (4.83)$$

and

$$\begin{aligned} d \log \ell(t, X(t)) = & \left(\frac{\Delta \ell(t, X(t))}{\ell(t, X(t))} - \frac{1}{2} \frac{|\nabla \ell(t, X(t))|^2}{\ell(t, X(t))^2} \right) dt \\ & - \left\langle \frac{\nabla \ell(t, X(t))}{\ell(t, X(t))}, 2 \nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} dt + \left\langle \frac{\nabla \ell(t, X(t))}{\ell(t, X(t))}, dW(t) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (4.84)$$

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

Proof. The canonical coordinate process $(X(t))_{t \geq 0}$ satisfies the stochastic equation (2.1). Applying Itô's formula, using (2.1) and (4.17), we obtain (4.83). One more application of Itô's formula leads to (4.84). \square

In order to deduce Theorem 3.1 from Lemma 4.16 — at least formally — we take expectations in (4.84) and observe that integration by parts shows that

$$\mathbb{E}_{\mathbb{P}} \left[\frac{\Delta \ell(t, X(t))}{\ell(t, X(t))} - \left\langle \frac{\nabla \ell(t, X(t))}{\ell(t, X(t))}, 2 \nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} \right] = 0. \quad (4.85)$$

But as opposed to the backward direction of time, the identity (4.85) does not hold any more when we take expectations conditionally on $X(t)$.

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

$$\ell^\beta(t, X(t)) = \frac{p^\beta(t, X(t))}{q(X(t))}, \quad t \geq t_0 \quad (4.86)$$

and its logarithm, again in the forward direction. With the help of the “perturbed” backwards Kolmogorov equation (4.29), we obtain the forward dynamics of the perturbed likelihood ratio process (4.86), as follows.

Lemma 4.17. *Under the Assumptions 1.2, let $t_0 \geq 0$. The perturbed likelihood ratio process (4.86) and its logarithm satisfy respectively the stochastic differential equations*

$$\begin{aligned} \frac{d\ell^\beta(t, X(t))}{\ell^\beta(t, X(t))} &= \left(\frac{\Delta \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))} + \operatorname{div} \beta(X(t)) \right) dt \\ &\quad - \left\langle \frac{\nabla \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))} + \beta(X(t)), 2 \nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} dt + \left\langle \frac{\nabla \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))}, dW^\beta(t) \right\rangle_{\mathbb{R}^n} \end{aligned} \quad (4.87)$$

and

$$\begin{aligned} d \log \ell^\beta(t, X(t)) &= \left(\frac{\Delta \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))} + \operatorname{div} \beta(X(t)) - \frac{1}{2} \frac{|\nabla \ell^\beta(t, X(t))|^2}{\ell^\beta(t, X(t))^2} \right) dt \\ &\quad - \left\langle \frac{\nabla \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))} + \beta(X(t)), 2 \nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n} dt + \left\langle \frac{\nabla \ell^\beta(t, X(t))}{\ell^\beta(t, X(t))}, dW^\beta(t) \right\rangle_{\mathbb{R}^n}, \end{aligned} \quad (4.88)$$

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

Proof. The canonical coordinate process $(X(t))_{t \geq 0}$ satisfies the stochastic equation (3.11). Together with (4.29) and Itô’s formula, this yields the stochastic equations (4.87) and (4.88). \square

Similarly as described after Lemma 4.16, taking expectations in (4.88) and integrating by parts allows to derive Theorem 3.4 from Lemma 4.17, at least on a formal level.

5. THE WASSERSTEIN TRANSPORT

For the convenience of the reader we review in this section some well-known results on quadratic Wasserstein transport, in order to establish the limits (3.32) and (3.35) and complete the proofs of Corollaries 3.10 and 3.11.

We recall below the definitions of the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$, and the quadratic Wasserstein distance W_2 . We follow the setting of [AGS08], 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 $\mathcal{P}(\mathbb{R}^n)$ the collection of probability measures on the Borel sets of \mathbb{R}^n . The *quadratic Wasserstein space* $\mathcal{P}_2(\mathbb{R}^n)$ is the subset of $\mathcal{P}(\mathbb{R}^n)$ consisting of the probability measures with finite second moment, i.e.,

$$\mathcal{P}_2(\mathbb{R}^n) := \left\{ P \in \mathcal{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} |x|^2 dP(x) < \infty \right\}. \quad (5.1)$$

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

$$\int_{\mathbb{R}^n} |x|^2 p(x) dx < \infty, \quad (5.2)$$

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

We denote by $\Gamma(P_0, P_1)$ the collection of Kantorovich transport plans, that is, probability measures γ in $\mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with given marginals $P_1, P_2 \in \mathcal{P}(\mathbb{R}^n)$. More precisely, if $\pi^i: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ are the canonical projections, then $\pi_{\#}^i \gamma = P_i$, for $i \in \{1, 2\}$. The *quadratic Wasserstein distance* between two probability measures $P_1, P_2 \in \mathcal{P}_2(\mathbb{R}^n)$ is defined by

$$W_2^2(P_1, P_2) := \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\gamma(x, y) : \gamma \in \Gamma(P_1, P_2) \right\}. \quad (5.3)$$

The quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$, endowed with the quadratic Wasserstein distance W_2 just introduced, is a Polish space [AGS08, Proposition 7.1.5].

In this section we consider the solution $(p(t))_{t \geq 0}$ of the Fokker-Planck equation (3.1) with initial condition (3.2) as a curve in the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$. This is justified by Lemma 2.1, which tells us that the Assumptions 1.2 guarantee that $p(t) \in \mathcal{P}_2(\mathbb{R}^n)$ for all $t \geq 0$. For fixed $T \in (0, \infty)$, we define now the time-dependent velocity field

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

Then the Fokker-Planck equation (3.1), satisfied by the curve $(p(t))_{0 \leq t \leq T}$ in $\mathcal{P}_2(\mathbb{R}^n)$ of probability density functions, can be written as a *continuity equation*, namely,

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

According to (4.20), we have

$$2 \int_0^T \left(\int_{\mathbb{R}^n} |v(t, x)|^2 p(t, x) dx \right) dt < \infty, \quad (5.6)$$

since the expressions in (4.20) and (5.6) are simply the same. In particular, (5.6) implies that we have $\|v(t)\|_{L^1(\mathbb{R}^n, p(t))} \in L^1[0, T]$, and we can apply [AGS08, Lemma 8.1.2] in order to choose a continuous representative. In other words, there exists a narrowly continuous curve $(\tilde{p}(t))_{0 \leq t \leq T}$ in $\mathcal{P}_2(\mathbb{R}^n)$ such that $p(t) = \tilde{p}(t)$ for Lebesgue-almost every $t \in [0, T]$. For convenience, we denote the continuous representative $(\tilde{p}(t))_{0 \leq t \leq T}$ again by $(p(t))_{0 \leq t \leq T}$.

As $(p(t))_{0 \leq t \leq T}$ is a narrowly continuous curve in $\mathcal{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (5.5) and $\|v(t)\|_{L^2(\mathbb{R}^n, p(t))} \in L^1[0, T]$, according to (5.6), we can invoke the second implication of [AGS08, Theorem 8.3.1]. The cited theorem relates absolutely continuous curves and the continuity equation. In particular, it tells us that the curve $(p(t))_{0 \leq t \leq T}$ is absolutely continuous [AGS08, Definition 1.1.1]. As a consequence, for Lebesgue-almost every $t \in [0, T]$, its *metric derivative* [AGS08, Theorem 1.1.2]

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

exists. Furthermore, [AGS08, Theorem 8.3.1] provides for Lebesgue-almost every $t \in [0, T]$ the estimate

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

On the other hand, the time-dependent velocity field $v: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ of (5.4) is a *gradient*, and therefore an element of the tangent space [AGS08, Definition 8.4.1] of $\mathcal{P}_2(\mathbb{R}^n)$ at the point $p(t) \in \mathcal{P}_2(\mathbb{R}^n)$, i.e.,

$$v(t) \in \text{Tan}_{p(t)}\mathcal{P}_2(\mathbb{R}^n) := \overline{\{\nabla\varphi: \varphi \in C_c^\infty(\mathbb{R}^n)\}}^{L^2(\mathbb{R}^n, p(t))}. \quad (5.9)$$

Since $(p(t))_{0 \leq t \leq T}$ is an absolutely continuous curve in $\mathcal{P}_2(\mathbb{R}^n)$ satisfying the continuity equation (5.5) for the time-dependent velocity field $v(t) \equiv v(t, \cdot)$, which is tangent to the curve, we can apply [AGS08, Proposition 8.4.5]. This result characterizes tangent vectors to absolutely continuous curves, and entails for Lebesgue-almost every $t \in [0, T]$ the inequality

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

Combining (5.8) and (5.10), we obtain for Lebesgue-almost every $t \in [0, T]$ the equality

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

This relates the strength of the local velocity field v in (5.5), to the rate of change, or “metric derivative” as in (5.7), of the Wasserstein distance along the curve $(p(t))_{0 \leq t \leq T}$ — justifying in this manner the introduction and the relevance of the Wasserstein distance in this context.

Using the metric derivative (5.7) of the curve $(p(t))_{0 \leq t \leq T}$, we can compute the arc length L of the curve with respect to the quadratic Wasserstein distance W_2 by

$$L = \int_0^T |p'(t)| dt. \quad (5.12)$$

This arc length L is nothing other than the quadratic Wasserstein distance between $p(0)$ and $p(T)$ along the curve $(p(t))_{0 \leq t \leq T}$. Let $t_1, t_2 \geq 0$. Motivated by (5.12), we define the *Wasserstein transportation cost* of moving $p(t_1)$ to $p(t_2)$ along the curve $(p(t))_{t \geq 0}$ as

$$\mathcal{T}_c(p(t_1), p(t_2)) := \int_{t_1}^{t_2} |p'(t)| dt \quad (5.13)$$

so that, in particular $\mathcal{T}_c(p(0), p(T)) = L$ is the quantity of (5.12). According to (5.11), this transportation cost can be computed as

$$\mathcal{T}_c(p(t_1), p(t_2)) = \int_{t_1}^{t_2} \left(\int_{\mathbb{R}^n} |v(t, x)|^2 p(t, x) dx \right)^{1/2} dt. \quad (5.14)$$

Furthermore, we note that

$$W_2(p(t_1), p(t_2)) \leq \mathcal{T}_c(p(t_1), p(t_2)) \quad (5.15)$$

for $0 \leq t_1 \leq t_2$, see [AGS08, p. 186].

We rephrase these well-known results as follows.

Theorem 5.1. *Under the Assumptions 1.2, for Lebesgue-almost every $t_0 \geq 0$ we have the local behavior of the Wasserstein distance*

$$\lim_{t \rightarrow t_0} \frac{W_2(p(t), p(t_0))}{|t - t_0|} = \frac{1}{2} \left(\mathbb{E}_{\mathbb{P}} \left[\frac{|\nabla \ell(t_0, X(t_0))|^2}{\ell(t_0, X(t_0))^2} \right] \right)^{1/2}. \quad (5.16)$$

Furthermore, for $t_1, t_2 \geq 0$, the Wasserstein transportation cost of moving $p(t_1)$ to $p(t_2)$ along the curve $(p(t))_{t \geq 0}$ amounts to

$$\mathcal{T}_c(p(t_1), p(t_2)) = \frac{1}{2} \int_{t_1}^{t_2} \left(\int_{\mathbb{R}^n} \frac{|\nabla \ell(t, x)|^2}{\ell(t, x)^2} p(t, x) dx \right)^{1/2} dt. \quad (5.17)$$

Proof. The identity (5.16) is just another way of phrasing the equality (5.11). The Wasserstein transportation cost (5.17) was derived in (5.14). \square

Now we consider the solution $(p^\beta(t))_{t \geq t_0}$ of the perturbed Fokker-Planck equation (3.9) with initial condition (3.10). Again, according to Lemma 3.3, this solution can be viewed as a curve in the quadratic Wasserstein space $\mathcal{P}_2(\mathbb{R}^n)$. Similarly as before, we define the time-dependent perturbed velocity field

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

Then the perturbed Fokker-Planck equation (3.9), satisfied by the perturbed curve $(p^\beta(t))_{t_0 \leq t \leq T}$, can be written as

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

To follow the same reasoning as above, we need that $v(t, \cdot)$ be a gradient, and this is why we have required the perturbation $\beta: \mathbb{R}^n \rightarrow \mathbb{R}^n$ to be a *gradient* vector field, i.e., of the form $\beta = \nabla B$ for some potential $B: \mathbb{R}^n \rightarrow \mathbb{R}$. Now, by the same token as above, and using the regularity assumption that the gradient vector field β is compactly supported and of class $C^\infty(\mathbb{R}^n; \mathbb{R}^n)$, we obtain the following result.

Theorem 5.2. *Under the Assumptions 1.2, for Lebesgue-almost every $t_0 \geq 0$ we have*

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

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

$$\mathcal{T}_c(p^\beta(t_1), p^\beta(t_2)) = \frac{1}{2} \int_{t_1}^{t_2} \left(\int_{\mathbb{R}^n} \left| \frac{\nabla \ell^\beta(t, x)}{\ell^\beta(t, x)} + 2\beta(x) \right|^2 p(t, x) dx \right)^{1/2} dt. \quad (5.21)$$

Remark 5.3. Since $X(t_0)$ has the same probability distribution under \mathbb{P} , as it does under \mathbb{P}^β , the expectation $\mathbb{E}_{\mathbb{P}}$ appearing in (5.20) can be replaced by $\mathbb{E}_{\mathbb{P}^\beta}$. \diamond

APPENDICES

A. BACHELIER'S WORK RELATING BROWNIAN MOTION TO THE HEAT EQUATION

In this section, which is only of historical interest, we point out that Bachelier already had some thoughts on “*horizontal transport of probability measures*” in his dissertation “*Théorie de la spéculation*” [Bac00, Bac06], 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 $(W(t))_{t \geq 0}$ as an infinitesimal version of a random walk. Suppose that the grid in space is given by

$$\dots, x_{n-2}, x_{n-1}, x_n, x_{n+1}, x_{n+2}, \dots \quad (\text{A.1})$$

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

$$\dots, p_{n-2}^t, p_{n-1}^t, p_n^t, p_{n+1}^t, p_{n+2}^t, \dots \quad (\text{A.2})$$

under the random walk under consideration. What are the probabilities

$$\dots, p_{n-2}^{t+\Delta t}, p_{n-1}^{t+\Delta t}, p_n^{t+\Delta t}, p_{n+1}^{t+\Delta t}, p_{n+2}^{t+\Delta t}, \dots \quad (\text{A.3})$$

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

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

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

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

$$P(t, x) = \int_{-\infty}^x p(t, z) dz \quad (\text{A.5})$$

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

$$\frac{\partial P}{\partial t} = \frac{1}{2} \frac{\partial p}{\partial x}, \quad (\text{A.6})$$

where we have normalized the relation between Δx and Δt to obtain the constant $1/2$. By differentiating (A.6) with respect to x one obtains the usual heat equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \quad (\text{A.7})$$

for the density function $p(t, x)$. Of course, the heat equation was known to Bachelier, and he notes regarding (A.7): “*C'est une équation de Fourier.*”

But let us still remain with the form (A.6) of the heat equation and analyze its message in terms of “horizontal transport of probability measures”. To accomplish the movement of mass $-\frac{1}{2}p'(t, x) dx$ from $(-\infty, x]$ to $[x, \infty)$ one is naturally led to define the flow induced by the velocity field

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

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 (A.1) and the corresponding probabilities (A.2) and (A.3), 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 $\frac{1}{2} + \frac{p'(t, x)}{p(t, x)} dx$ and from x_{n+1} to x_n with probability $\frac{1}{2} - \frac{p'(t, x)}{p(t, x)} dx$, with the identifications $x = x_n = x_{n+1}$, and $dx = \Delta x$. In other words, the above Brownian motion $(W(t))_{t \geq 0}$ considered in reverse direction $(W(T-s))_{0 \leq s \leq T}$ is *not* a Brownian motion, as the transition probabilities are not $(1/2, 1/2)$ any more. Rather, one has to correct these probabilities by a term which — once again — involves our familiar score function $\nabla p(t, x)/p(t, x)$ (compare (4.5) above). At this stage, it should come as no surprise, that the passage to reverse time is closely related to the Wasserstein transport induced by $p(t, x)$.

Let us play this infinitesimal reasoning one more time, in order to visualize the Fontbona-Jourdain result (Theorem 4.2). Arguing in the reverse direction of time, we may ask the following question: how do we have to choose the transition probabilities to go from x at time $t + \Delta t$ to either $x + dx$ or $x - dx$ at time t , so that the density process $p(t, x)$ becomes a martingale in reverse time under these transition probabilities? As the difference between the probabilities $p(t, x + dx)$ and $p(t, x - dx)$ equals $2p'(t, x) dx$ (up to terms of smaller order than dx) we conclude that the transition probabilities have to be changed from $(1/2, 1/2)$ to

$$\left(\frac{1}{2} - \frac{p'(t, x)}{p(t, x)} dx, \frac{1}{2} + \frac{p'(t, x)}{p(t, x)} dx \right) \quad (\text{A.9})$$

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

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

B. THE PROOFS OF LEMMAS 2.1 AND 3.3

Proof of Lemma 2.1. Let the real constants $c, R \geq 0$ be as in condition (iii) of Assumptions 1.2, and denote

$$m_R := \max_{|x| \leq R} |1 - \langle x, 2 \nabla \Psi(x) \rangle_{\mathbb{R}^n}| < \infty, \quad \tau_k := \inf \{t \geq 0: |X(t)| > k\} \quad (\text{B.1})$$

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

$$d|X(t)|^2 = \left(1 - \left\langle X(t), 2 \nabla \Psi(X(t)) \right\rangle_{\mathbb{R}^n}\right) dt + \langle 2X(t), dW(t) \rangle_{\mathbb{R}^n} \quad (\text{B.2})$$

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

$$\varphi_k(t) = \varphi(0) + \mathbb{E}_{\mathbb{P}} \left[\int_0^{t \wedge \tau_k} \left(1 - \left\langle X(u), 2 \nabla \Psi(X(u)) \right\rangle_{\mathbb{R}^n}\right) \mathbf{1}_{\{|X(u)| \leq R\}} du \right] \quad (\text{B.3})$$

$$+ \mathbb{E}_{\mathbb{P}} \left[\int_0^{t \wedge \tau_k} \left(1 - \left\langle X(u), 2 \nabla \Psi(X(u)) \right\rangle_{\mathbb{R}^n}\right) \mathbf{1}_{\{|X(u)| > R\}} du \right] \quad (\text{B.4})$$

$$\leq \varphi(0) + m_R \mathbb{E}_{\mathbb{P}}[t \wedge \tau_k] + \mathbb{E}_{\mathbb{P}} \left[\int_0^{t \wedge \tau_k} (1 + 2c|X(u)|^2) du \right] \quad (\text{B.5})$$

$$\leq \varphi(0) + (1 + m_R)t + 2c \int_0^t \varphi_k(u) du. \quad (\text{B.6})$$

The Gronwall inequality gives now

$$\varphi_k(t) \leq g(t) := \varphi(0) + (1 + m_R)t + 2c \int_0^t (\varphi(0) + (1 + m_R)u) e^{2c(t-u)} du. \quad (\text{B.7})$$

According to the second-moment condition in (1.11), the quantity $g(t)$ is finite for all $t \geq 0$, and independent of k ; letting $k \uparrow \infty$ in (B.7), we get

$$\varphi(t) = \mathbb{E}_{\mathbb{P}}[|X(t)|^2] \leq g(t) < \infty, \quad t \geq 0. \quad (\text{B.8})$$

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

Proof of Lemma 3.3. The proof of Lemma 3.3 follows by analogy with the proof of Lemma 2.1 above. Indeed, one just has to add the perturbation β to the gradient $\nabla \Psi$ in the expressions (B.1) – (B.4), write $W^\beta(t)$ instead of $W(t)$ in (B.2), and replace all the \mathbb{P} -expectations by expectations with respect to the probability measure \mathbb{P}^β . Then the constant m_R in (B.1) is still finite and, because of its compact support, the perturbation β in the expression (B.4) vanishes, provided R is chosen large enough. Hence, by the same token as above, we conclude that $P^\beta(t) \in \mathcal{P}_2(\mathbb{R}^n)$ for all $t \geq t_0$. \square

C. RELATIVE ENTROPY WITH RESPECT TO A σ -FINITE MEASURE \mathbb{Q}

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

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

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

Let us consider the σ -finite measure \mathbb{Q} on the Borel sets of \mathbb{R}^n with density $q(x) = e^{-2\Psi(x)}$ for $x \in \mathbb{R}^n$, introduced in Section 2. Following the approach of [Léo14, Section 2], we shall demonstrate that the same definition of relative entropy $H(P | \mathbb{Q})$ applies to the reference measure \mathbb{Q} , provided that the probability measure P is an element of $\mathcal{P}_2(\mathbb{R}^n)$, with the only difference that the quantity (C.1) now takes values in $(-\infty, \infty]$.

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

$f(x) := e^{-|x|^2}$, $x \in \mathbb{R}$, we have $\mathbb{E}_Q[f] \in (0, \infty)$. According to [Léo14, Section 2], we let \mathcal{Q} be the probability measure on the Borel sets of \mathbb{R}^n having probability density function $f/\mathbb{E}_Q[f]$ with respect to the measure \mathcal{Q} , so that

$$\frac{dP}{d\mathcal{Q}} = \frac{f}{\mathbb{E}_Q[f]} \frac{dP}{d\mathcal{Q}}. \quad (\text{C.2})$$

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

$$H(P|Q) = H(P|\mathcal{Q}) - \int_{\mathbb{R}^n} p(x) |x|^2 dx - \log \left(\int_{\mathbb{R}^n} e^{-|x|^2 - 2\Psi(x)} dx \right), \quad (\text{C.3})$$

which is justified by (C.1) and the fact that $P \in \mathcal{P}_2(\mathbb{R}^n)$ as well as $\mathbb{E}_Q[f] \in (0, \infty)$. In particular, we see that the right-hand side of (C.3) takes values in the interval $(-\infty, \infty]$. Summing up, we can define well the relative entropy $H(P|Q)$ as in (C.1) provided that $P \in \mathcal{P}_2(\mathbb{R}^n)$, even when the σ -finite measure \mathcal{Q} has infinite total mass.

Remark C.1. Wherever in this paper the relative entropy $H(P|Q)$ is considered for some σ -finite measure \mathcal{Q} on the Borel sets of \mathbb{R}^n with density $q(x) = e^{-2\Psi(x)}$, $x \in \mathbb{R}^n$, the probability measure P will always be assumed to belong to $\mathcal{P}_2(\mathbb{R}^n)$. This is in accordance with Lemmas 2.1 and 3.3, as well as Lemma 3.19. In the latter, the constant-speed geodesic $(P_t)_{0 \leq t \leq 1}$ joining two probability measures P_0 and P_1 in $\mathcal{P}_2(\mathbb{R}^n)$ is considered. Therefore, in all situations relevant to us, the relative entropy $H(P|Q)$ is well-defined and takes values in the interval $(-\infty, \infty]$. \diamond

D. A MEASURE-THEORETIC RESULT

In the proofs of Propositions 3.12 and 3.14 we have used a result on conditional expectations, which we will formulate and prove below. We place ourselves on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a left-continuous filtration $(\mathcal{F}(t))_{t \geq 0}$. We first state the following result, which is known as *Scheffé's lemma* [Wil91, 5.10].

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

Proposition D.2. *Let $(B(t))_{0 \leq t \leq T}$ be a continuous, adapted, non-negative process. Define the process $(A(t))_{0 \leq t \leq T}$ as its primitive, i.e.,*

$$A(t) = \int_0^t B(u) du, \quad 0 \leq t \leq T \quad (\text{D.1})$$

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

$$\lim_{t \rightarrow t_0} \mathbb{E} \left[\frac{A(t) - A(t_0)}{t - t_0} \right] = \lim_{t \rightarrow t_0} \mathbb{E} \left[\frac{1}{t - t_0} \int_{t_0}^t B(u) du \right] = \mathbb{E}[B(t_0)]. \quad (\text{D.2})$$

Now fix a "Lebesgue point" $t_0 \in [0, T]$ for which (D.2) does hold. Then we have the analogous limiting assertion for the conditional expectations, i.e.,

$$\lim_{t \uparrow t_0} \frac{\mathbb{E}[A(t_0) - A(t) | \mathcal{F}(t)]}{t_0 - t} = \lim_{t \downarrow t_0} \frac{\mathbb{E}[A(t) - A(t_0) | \mathcal{F}(t_0)]}{t - t_0} = B(t_0), \quad (\text{D.3})$$

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

Proof. Fix a Lebesgue point $t_0 \in [0, T]$ for which (D.2) does hold. As the process $(B(t))_{0 \leq t \leq T}$ is continuous, the fundamental theorem of calculus ensures that the limit

$$\lim_{t \rightarrow t_0} \frac{A(t) - A(t_0)}{t - t_0} = \lim_{t \rightarrow t_0} \frac{1}{t - t_0} \int_{t_0}^t B(u) du = B(t_0) \quad (\text{D.4})$$

exists almost surely. Since the random variables appearing in (D.4) are integrable, and we already have the convergence of the $L^1(\mathbb{P})$ -norms from (D.2), Lemma D.1 allows us to conclude that the convergence of (D.4) holds also in the norm of $L^1(\mathbb{P})$, i.e.,

$$\lim_{t \rightarrow t_0} \mathbb{E} \left[\frac{A(t) - A(t_0)}{t - t_0} - B(t_0) \right] = 0. \quad (\text{D.5})$$

From (D.5) we can deduce now the $L^1(\mathbb{P})$ -convergence of (D.3) as follows. Regarding the second limit in (D.3), for $t > t_0$, we find

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

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

and because of (D.5), the expression in (D.7) converges to zero as $t \downarrow t_0$. Similarly, to handle the first limit in (D.3) we use for $t < t_0$ the estimate

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

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

As $t \uparrow t_0$, the expression on the right-hand side of (D.8) converges to zero as before. The same is true also for the term in (D.9), on account of [Chu01, Theorem 9.4.8] and the left-continuity of the filtration $(\mathcal{F}(t))_{0 \leq t \leq T}$. This completes the proof of Proposition D.2. \square

E. THE PROOF OF THE FONTBONA-JOURDAIN THEOREM

Proof of Theorem 4.2 [FJ16]. For $0 \leq s \leq T$, we define the random variable $M(T - s)$ as the conditional expectation of the random variable

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

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

$$M(T - s) := \mathbb{E}_{\mathbb{Q}} \left[\ell(0, X(0)) \mid \mathcal{G}(T - s) \right], \quad 0 \leq s \leq T. \quad (\text{E.2})$$

Obviously the process $(M(T - s))_{0 \leq s \leq T}$ is a martingale of the backwards filtration $(\mathcal{G}(T - s))_{0 \leq s \leq T}$ under the probability measure \mathbb{Q} . Now we make the following elementary, but crucial, observation: as the stochastic process $(X(t))_{0 \leq t \leq T}$, which solves the stochastic differential equation (2.1), is a *Markov process*, the time-reversed process $(X(T - s))_{0 \leq s \leq T}$ is a Markov process, too, under the probability measure \mathbb{P} as well as under \mathbb{Q} . Hence

$$M(T - s) = \mathbb{E}_{\mathbb{Q}} \left[\ell(0, X(0)) \mid X(T - s) \right], \quad 0 \leq s \leq T. \quad (\text{E.3})$$

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

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

Note also that

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

Because the Borel set $A \subseteq \mathbb{R}^n$ is arbitrary, we deduce from (E.4) and (E.5) that

$$\mathbb{E}_{\mathbb{Q}} \left[\frac{p(0, X(0))}{q(X(0))} \mid X(T-s) \right] = \frac{p(T-s, X(T-s))}{q(X(T-s))} = \ell(T-s, X(T-s)). \quad (\text{E.6})$$

This completes the proof of Theorem 4.2. \square

F. THE PROOF OF LEMMA 3.19

Proof of Lemma 3.19. In order to show (3.62), we recall the notation of (3.59) and consider the time-dependent velocity field

$$[0, 1] \times \mathbb{R}^n \ni (t, x) \longmapsto v_t(x) := \gamma\left((T_t^\gamma)^{-1}(x)\right) \in \mathbb{R}^n, \quad (\text{F.1})$$

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

$$\frac{d}{dt} T_t^\gamma(x) = v_t(T_t^\gamma(x)). \quad (\text{F.2})$$

Let $p_t(\cdot)$ be the probability density function of the probability measure P_t in (3.59). Then, according to [Vil03, Theorem 5.34], the function $p_t(\cdot)$ satisfies the continuity equation

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

which can be written equivalently as

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

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

$$X_t = X_0 + \int_0^t v_\theta(X_\theta) d\theta, \quad 0 \leq t \leq 1 \quad (\text{F.5})$$

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

$$dp_t(X_t) = \partial_t p_t(X_t) dt + \langle \nabla p_t(X_t), dX_t \rangle_{\mathbb{R}^n} = -p_t(X_t) \operatorname{div}(v_t(X_t)) dt \quad (\text{F.6})$$

on account of (F.4), (F.5), thus also

$$d \log p_t(X_t) = -\operatorname{div}(v_t(X_t)) dt, \quad 0 \leq t \leq 1. \quad (\text{F.7})$$

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

$$d \log q(X_t) = -\langle 2 \nabla \Psi(X_t), dX_t \rangle_{\mathbb{R}^n} = -\langle 2 \nabla \Psi(X_t), v_t(X_t) \rangle_{\mathbb{R}^n} dt. \quad (\text{F.8})$$

For the likelihood ratio function $\ell_t(\cdot)$ of (3.61) we get from (F.7) and (F.8) that

$$d \log \ell_t(X_t) = \langle 2 \nabla \Psi(X_t), v_t(X_t) \rangle_{\mathbb{R}^n} dt - \operatorname{div}(v_t(X_t)) dt, \quad 0 \leq t \leq 1. \quad (\text{F.9})$$

Taking expectations in the integral version of (F.9), we obtain that the difference

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

is equal to

$$\mathbb{E}_\nu \left[\int_0^t \left(\langle 2 \nabla \Psi(X_\theta), v_\theta(X_\theta) \rangle_{\mathbb{R}^n} - \operatorname{div}(v_\theta(X_\theta)) \right) d\theta \right] \quad (\text{F.11})$$

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

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

Integrating by parts, we see that

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

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

Recalling (F.12), and combining it with the relation $\nabla \log \ell_t(x) = \nabla \log p_t(x) + 2 \nabla \Psi(x)$, as well as with (F.13) and (F.14), we get

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

Since $v_0 = \gamma$, this leads to (3.62). □

G. TIME REVERSAL OF DIFFUSIONS

We review in the present section the theory of time reversal for diffusion processes developed by Föllmer [Föl85, Föl86], Haussmann and Pardoux [HP86], and Pardoux [Par86]. This section can be read independently of the rest of the paper; it does not present novel results.

G.1. INTRODUCTION

It is very well known that the Markov property is invariant under time reversal. In other words, a Markov process remains a Markov process under time reversal (e.g., [RW00a, Exercise E60.41, p. 162]). On the other hand, it is also well known that the strong Markov property is not necessarily preserved under time reversal (e.g., [RW00a, p. 330]), and neither is the semimartingale property (e.g., [Wal82]). The reason for such failure is the same in both cases: after reversing time, “we may know too much”. Thus, the following questions arise rather naturally:

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

$$\widehat{X}(s) := X(T - s), \quad 0 \leq s \leq T, \quad (\text{G.1})$$

also be a diffusion? if it happens to be, what are the characteristics of the time-reversed diffusion?

Such questions go back at least to Boltzmann [Bol96, Bol98a, Bol98b], Schrödinger [Sch31, Sch32] and Kolmogorov [Kol37]; they were dealt with systematically by Nelson [Nel01] (see also Carlen [Car84]) in the context of Nelson’s dynamical theories for Brownian motion and diffusion. There is now a rather complete theory that answers these questions and provides, as a kind of “bonus”, some rather unexpected results as well. It was developed in the context of theories of filtering, interpolation and extrapolation, where such issues arise naturally — most notably Haussmann and Pardoux [HP86], and Pardoux [Par86]. Very interesting related results in a non-Markovian context, but with dispersion structure given by the identity matrix, have been obtained by Föllmer [Föl85, Föl86]. Here, this theory is presented in the spirit of the expository paper by Meyer [Mey94].

G.2. THE SETTING

We place ourselves on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{F} = (\mathcal{F}(t))_{0 \leq t \leq T}$ rich enough to support an \mathbb{R}^d -valued Brownian motion $W = (W_1, \dots, W_d)'$ adapted to \mathbb{F} , as well as an independent $\mathcal{F}(0)$ -measurable random vector $\xi = (\xi_1, \dots, \xi_n)': \Omega \rightarrow \mathbb{R}^n$. In fact, we shall assume that \mathbb{F} is the filtration generated by these two objects, in the sense that we shall take

$$\mathcal{F}(t) = \sigma(\xi, W(\theta): 0 \leq \theta \leq t), \quad 0 \leq t \leq T,$$

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

$$X_i(t) = \xi_i + \int_0^t b_i(\theta, X(\theta)) d\theta + \sum_{\nu=1}^d \int_0^t s_{i\nu}(\theta, X(\theta)) dW_\nu(\theta), \quad 0 \leq t \leq T, \quad (\text{G.2})$$

for $i = 1, \dots, n$ admits a pathwise unique, strong solution. It is then well known that the resulting continuous process $X = (X_1, \dots, X_n)'$ is \mathbb{F} -adapted (the strong solvability of the equation (G.2)), which implies that we have also

$$\mathcal{F}(t) = \sigma(X(\theta), W(\theta): 0 \leq \theta \leq t) = \sigma(X(0), W(t) - W(0): 0 \leq \theta \leq t) \quad (\text{G.3})$$

modulo \mathbb{P} -augmentation, for $0 \leq t \leq T$; as well as that X has the strong Markov property, and is thus a diffusion process with drifts $b_i(\cdot, \cdot)$ and dispersions $s_{i\nu}(\cdot, \cdot)$, $i = 1, \dots, n$, $\nu = 1, \dots, d$. We shall denote the (i, j) th entry of the covariance matrix $a(t, x) := s(t, x) s'(t, x)$ by

$$a_{ij}(t, x) := \sum_{\nu=1}^d s_{i\nu}(t, x) s_{j\nu}(t, x), \quad 1 \leq i, j \leq n.$$

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 $p(t, \cdot): \mathbb{R}^n \rightarrow (0, \infty)$ in

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

is smooth. Sufficient conditions on the drift $b_i(\cdot, \cdot)$ and dispersion $s_{i\nu}(\cdot, \cdot)$ characteristics that lead to such smoothness, are provided by the Hörmander hypoellipticity conditions; see for instance [Bel95], [Nua06] for this result, as well as [Rog85] for a very simple argument in the one-dimensional case ($n = d = 1$), and to the case of Langevin type equation (2.1) for arbitrary $n \in \mathbb{N}$. We refer to [Fri75], [RW00b] or [KS91] for the basics of the theory of stochastic equations of the form (G.2).

The probability density function $p(t, \cdot): \mathbb{R}^n \rightarrow (0, \infty)$ solves the forward Kolmogorov [Kol31] equation [Fri75, p. 149]

$$\partial_t p(t, x) = \frac{1}{2} \sum_{i,j=1}^n D_{ij}^2 (a_{ij}(t, x) p(t, x)) - \sum_{i=1}^n D_i (b_i(t, x) p(t, x)), \quad (t, x) \in (0, T] \times \mathbb{R}^n. \quad (\text{G.4})$$

If the drift and dispersion characteristics do not depend on time, and an invariant probability measure exists for the diffusion process of (G.2), the density function $p(\cdot)$ of this measure solves the stationary version of this forward Kolmogorov equation, to wit

$$\frac{1}{2} \sum_{i,j=1}^n D_{ij}^2(a_{ij}(x)p(x)) = \sum_{i=1}^n D_i(b_i(x)p(x)), \quad x \in \mathbb{R}^n. \quad (\text{G.5})$$

G.3. TIME REVERSAL AND THE BACKWARDS FILTRATION

Consider now the family of σ -algebras $(\widehat{\mathcal{F}}(t))_{0 \leq t \leq T}$ given by

$$\widehat{\mathcal{F}}(t) := \sigma(X(\theta), W(\theta) - W(t) : t \leq \theta \leq T), \quad 0 \leq t \leq T. \quad (\text{G.6})$$

It is not hard to see that the σ -algebra in (G.6) is expressed equivalently as

$$\begin{aligned} \widehat{\mathcal{F}}(t) &= \sigma(X(t), W(\theta) - W(t) : t \leq \theta \leq T) = \sigma(X(t), W(\theta) - W(T) : t \leq \theta \leq T) \\ &= \sigma(X(T), W(\theta) - W(t) : t \leq \theta \leq T) = \sigma(X(T)) \vee \mathcal{H}(t). \end{aligned} \quad (\text{G.7})$$

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

$$\mathcal{H}(t) := \sigma(W(\theta) - W(t) : t \leq \theta \leq T), \quad 0 \leq t \leq T, \quad (\text{G.8})$$

is independent of the random vector $X(t)$. The time-reversed processes \widehat{X} as in (G.1), as well as

$$\widetilde{W}(s) := W(T - s) - W(T), \quad 0 \leq s \leq T, \quad (\text{G.9})$$

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

$$\begin{aligned} \widehat{\mathcal{F}}(T - s) &= \sigma(X(T - u), W(T - u) - W(T - s) : 0 \leq u \leq s) \\ &= \sigma(\widehat{X}(u), \widetilde{W}(u) - \widetilde{W}(s) : 0 \leq u \leq s) \end{aligned} \quad (\text{G.10})$$

from (G.6). Note that, by complete analogy with (G.3), we have also

$$\widehat{\mathcal{F}}(T - s) = \sigma(X(T), W(T - u) - W(T - s) : 0 \leq u \leq s) = \sigma(\widehat{X}(0)) \vee \mathcal{H}(T - s) \quad (\text{G.11})$$

on account of (G.7), where

$$\mathcal{H}(T - s) = \sigma(W(T - u) - W(T - s) : 0 \leq u \leq s) = \sigma(\widetilde{W}(u) - \widetilde{W}(s) : 0 \leq u \leq s). \quad (\text{G.12})$$

In words: the σ -algebra $\widehat{\mathcal{F}}(T - s)$ is generated by the terminal value $X(T)$ of the forward process (i.e., by the original value $\widehat{X}(0)$ of the backward process) and by the increments of the time-reversed process \widetilde{W} on $[0, s]$; see the expressions right above. Furthermore, the σ -algebra $\widehat{\mathcal{F}}(T - s)$ measures all the random variables $\widehat{X}(u)$, $u \in [0, s]$.

Remark G.1. In fact, the time-reversed process \widetilde{W} is a *Brownian motion of the backwards filtration* $\mathbb{H} := (\mathcal{H}(T - s))_{0 \leq s \leq T} \subseteq \widehat{\mathbb{F}}$ as in (G.12), generated by the increments of W after time $T - s$, $0 \leq s \leq T$. This is because it is a martingale of this filtration, has continuous paths, and its quadratic variation is that of Brownian motion (Lévy's theorem [KS91, Theorem 5.1]). In the next subsection we shall see that the process \widetilde{W} is *only a semimartingale of the larger backwards filtration* $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}(T - s))_{0 \leq s \leq T}$, and identify its semimartingale decomposition. \diamond

G.4. SOME REMARKABLE BROWNIAN MOTIONS

Following the exposition and ideas in [Mey94], we start with a couple of observations. First, for every $t \in [0, T]$ and every integrable, $\widehat{\mathcal{F}}(t)$ -measurable random variable \mathcal{K} , we have

$$\mathbb{E}[\mathcal{K} \mid \mathcal{F}(t)] = \mathbb{E}[\mathcal{K} \mid X(t)], \quad \text{almost surely.} \quad (\text{G.13})$$

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

$$g(\theta, x) := \mathbb{E}[G(X(t)) \mid X(\theta) = x], \quad (\theta, x) \in [0, t] \times \mathbb{R}^n.$$

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

$$g(\theta, X(\theta)) = \mathbb{E}[G(X(t)) \mid X(\theta)] = \mathbb{E}[G(X(t)) \mid \mathcal{F}(\theta)], \quad 0 \leq \theta \leq t$$

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

$$G(X(t)) - g(\theta, X(\theta)) = g(t, X(t)) - g(\theta, X(\theta)) = \sum_{i=1}^n \sum_{\nu=1}^d \int_{\theta}^t D_i g(v, X(v)) s_{i\nu}(v, X(v)) dW_\nu(v).$$

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

$$\begin{aligned} \mathbb{E}[(W_\nu(t) - W_\nu(\theta)) \cdot G(X(t))] &= \mathbb{E}\left[(W_\nu(t) - W_\nu(\theta)) \cdot (g(t, X(t)) - g(\theta, X(\theta)))\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \int_{\theta}^t D_i g(v, X(v)) s_{i\nu}(v, X(v)) dv\right] = \sum_{i=1}^n \int_{\theta}^t \int_{\mathbb{R}^n} (D_i g \cdot s_{i\nu})(v, x) p(v, x) dx dv \\ &= - \sum_{i=1}^n \int_{\theta}^t \int_{\mathbb{R}^n} g(v, x) D_i(p(v, x) s_{i\nu}(v, x)) dx dv = - \int_{\theta}^t \int_{\mathbb{R}^n} g(v, x) \operatorname{div}(p(v, x) \bar{s}_\nu(v, x)) dx dv \\ &= - \int_{\theta}^t \mathbb{E}\left[g(v, X(v)) \cdot \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(v, X(v))\right] dv = - \mathbb{E}\left[G(X(t)) \cdot \int_{\theta}^t \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(v, X(v)) dv\right]. \end{aligned}$$

Here $\bar{s}_\nu(v, \cdot)$ is the ν^{th} column vector of the dispersion matrix. Comparing the first and last expressions in the above string of equalities, we see that with $0 \leq \theta \leq t$ we have

$$\mathbb{E}\left[G(X(t)) \cdot \left(W_\nu(t) - W_\nu(\theta) + \int_{\theta}^t \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(v, X(v)) dv\right)\right] = 0 \quad (\text{G.14})$$

for every $G \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, and thus by extension for every bounded, measurable $G: \mathbb{R}^n \rightarrow \mathbb{R}$.

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

$$B_\nu(s) := \widetilde{W}_\nu(s) - \int_0^s \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(T - u, \widehat{X}(u)) du \quad (\text{G.15})$$

$$= W_\nu(T - s) - W_\nu(T) - \int_{T-s}^T \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(v, X(v)) dv, \quad 0 \leq s \leq T, \quad (\text{G.16})$$

for $\nu = 1, \dots, d$, is a Brownian motion of the backwards filtration $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}(T - s))_{0 \leq s \leq T}$.

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

$$\{B(T - \theta) - B(T - t) : 0 \leq \theta \leq t\} \quad \text{is independent of} \quad \widehat{\mathcal{F}}(t) \supseteq \sigma(X(v) : t \leq v \leq T).$$

Note also from (G.16) that

$$B_\nu(T - \theta) - B_\nu(T - t) = W_\nu(\theta) - W_\nu(t) - \int_{\theta}^t \frac{\operatorname{div}(p \bar{s}_\nu)}{p}(v, X(v)) dv, \quad 0 \leq \theta \leq t. \quad \diamond$$

Reversing time once again, we obtain the following corollary of Theorem G.2.

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

$$V_\nu(t) := B_\nu(T-t) - B_\nu(T) = W_\nu(t) + \int_0^t \frac{\operatorname{div}(p\bar{s}_\nu)}{p}(v, X(v)) \, dv, \quad 0 \leq t \leq T, \quad (\text{G.17})$$

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

$$\mathcal{F}^V(t) := \sigma(V(\theta) : 0 \leq \theta \leq t) \quad (\text{G.18})$$

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

Proof of Theorem G.2. It suffices to show that each component process B_ν is a martingale of the backwards filtration $\widehat{\mathbb{F}}$; because then, in view of the continuity of paths and the easily checked property $\langle B_\nu, B_\ell \rangle(s) = s \delta_{\nu\ell}$, we can deduce that each B_ν is a Brownian motion in the backwards filtration $\widehat{\mathbb{F}}$ (and of course also in its own filtration), and that B_ν, B_ℓ are independent for $\ell \neq \nu$, by appealing to Lévy's theorem once again.

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

$$\mathbb{E} \left[\mathbb{E}[\mathcal{K} | \mathcal{F}(t)] \cdot \left(W_\nu(t) - W_\nu(\theta) + \int_\theta^t \frac{\operatorname{div}(p\bar{s}_\nu)}{p}(v, X(v)) \, dv \right) \right] = 0,$$

as the expression inside the curved braces is $\mathcal{F}(t)$ -measurable. But recalling (G.13) we have that $\mathbb{E}[\mathcal{K} | \mathcal{F}(t)] = \mathbb{E}[\mathcal{K} | X(t)] = G(X(t))$ for some bounded, measurable $G: \mathbb{R}^n \rightarrow \mathbb{R}$, and the desired result follows from (G.14). \square

G.5. THE DIFFUSION PROPERTY UNDER TIME REVERSAL

Let us return now to the question, whether the time-reversed process \widehat{X} of (G.1), (G.2) is a diffusion. We start by expressing X_i of (G.2) in terms of a backwards Itô integral (see Subsection G.6) as

$$\begin{aligned} X_i(t) - \xi_i - \int_0^t b_i(\theta, X(\theta)) \, d\theta &= \sum_{\nu=1}^d \int_0^t s_{i\nu}(\theta, X(\theta)) \, dW_\nu(\theta) \\ &= \sum_{\nu=1}^d \left(\int_0^t s_{i\nu}(\theta, X(\theta)) \bullet dW_\nu(\theta) - \langle s_{i\nu}(\cdot, X), W_\nu \rangle(t) \right). \end{aligned}$$

From (G.2), we have by Itô's formula that the process

$$s_{i\nu}(\cdot, X) - s_{i\nu}(0, \xi) - \sum_{j=1}^n \sum_{\kappa=1}^d \int_0^\cdot D_j s_{i\nu}(\theta, X(\theta)) \cdot s_{j\kappa}(\theta, X(\theta)) \, dW_\kappa(\theta)$$

is of finite first variation, therefore $\langle s_{i\nu}(\cdot, X), W_\nu \rangle(t) = \sum_{j=1}^n \int_0^t s_{j\nu}(\theta, X(\theta)) \, D_j s_{i\nu}(\theta, X(\theta)) \, d\theta$. We conclude

$$X_i(t) = \xi_i - \int_0^t \left(\sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu} \, D_j s_{i\nu} - b_i \right) (\theta, X(\theta)) \, d\theta + \sum_{\nu=1}^d \int_0^t s_{i\nu}(\theta, X(\theta)) \bullet dW_\nu(\theta).$$

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

$$X_i(t) = X_i(T) + \int_t^T \left(\sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu} D_j s_{i\nu} - b_i \right) (\theta, X(\theta)) d\theta - \sum_{\nu=1}^d \int_t^T s_{i\nu} (\theta, X(\theta)) \bullet dW_\nu(\theta),$$

as well as

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \left(\sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu} D_j s_{i\nu} - b_i \right) (T - u, \widehat{X}(u)) du + \sum_{\nu=1}^d \int_0^s s_{i\nu} (T - u, \widehat{X}(u)) d\widetilde{W}_\nu(u)$$

by reversing time. It is important here to note that the backward Itô integral for W becomes a forward Itô integral for the process \widetilde{W} , the time-reversal of W in the manner of (G.9).

But now let us recall (G.15), on the strength of which the above expression takes the form

$$\begin{aligned} \widehat{X}_i(s) &= \widehat{X}_i(0) + \sum_{\nu=1}^d \int_0^s s_{i\nu} (T - u, \widehat{X}(u)) dB_\nu(u) \\ &\quad + \int_0^s \left(\sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu} D_j s_{i\nu} + \sum_{\nu=1}^d s_{i\nu} \frac{\operatorname{div}(p \overline{s}_\nu)}{p} - b_i \right) (T - u, \widehat{X}(u)) du, \quad 0 \leq s \leq T. \end{aligned}$$

But in conjunction with Theorem G.2, this means that the time-reversed process \widehat{X} of (G.1), (G.2) is a semimartingale of the backwards filtration $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}(T - s))_{0 \leq s \leq T}$, with decomposition

$$\widehat{X}_i(s) = \widehat{X}_i(0) + \int_0^s \widehat{b}_i(T - u, \widehat{X}(u)) du + \sum_{\nu=1}^d \int_0^s s_{i\nu} (T - u, \widehat{X}(u)) dB_\nu(u) \quad (\text{G.19})$$

for $0 \leq s \leq T$, where, for each $i = 1, \dots, n$, the function $\widehat{b}_i(\cdot, \cdot)$ is specified by

$$\begin{aligned} \widehat{b}_i(t, x) + b_i(t, x) &= \sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu}(t, x) D_j s_{i\nu}(t, x) + \sum_{\nu=1}^d s_{i\nu}(t, x) \frac{\operatorname{div}(p(t, x) \overline{s}_\nu(t, x))}{p(t, x)} \\ &= \sum_{j=1}^n \sum_{\nu=1}^d s_{j\nu}(t, x) D_j s_{i\nu}(t, x) + \sum_{\nu=1}^d \frac{s_{i\nu}(t, x)}{p(t, x)} \left(\sum_{j=1}^n D_j (p(t, x) s_{j\nu}(t, x)) \right) \\ &= \sum_{j=1}^n (D_j a_{ij}(t, x) + a_{ij}(t, x) \cdot D_j \log p(t, x)). \end{aligned}$$

Theorem G.5. *Under the assumptions of this section, the time-reversed process \widehat{X} of (G.1), (G.2) is a diffusion in the backwards filtration $\widehat{\mathbb{F}} = (\widehat{\mathcal{F}}(T - s))_{0 \leq s \leq T}$, with characteristics as in (G.19), namely, dispersions $s_{i\nu}(T - s, x)$ and drifts $\widehat{b}_i(T - s, x)$ given by the generalized Nelson equation*

$$\widehat{b}_i(t, x) + b_i(t, x) = \sum_{j=1}^n \left(D_j a_{ij}(t, x) + a_{ij}(t, x) \cdot D_j \log p(t, x) \right), \quad i = 1, \dots, n. \quad (\text{G.20})$$

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

$$\widehat{b}(t, x) + b(t, x) = \operatorname{div}(a(t, x)) + a(t, x) \cdot \nabla \log p(t, x). \quad (\text{G.21})$$

Remark G.6. This result can be extended to the case where the sums of the distributional derivatives $\sum_{j=1}^n D_j (a_{ij}(t, x) p(t, x))$, $i = 1, \dots, n$, are only assumed to be locally integrable functions of $x \in \mathbb{R}^n$; see [MNS89, RVW01]. \diamond

Remark G.7 (SOME FILTRATION COMPARISONS). For an invertible dispersion matrix $s(\cdot, \cdot)$, it follows from (G.19) that the Brownian motion B is adapted to the filtration generated by \widehat{X} ; that is,

$$\mathcal{F}^B(s) \subseteq \mathcal{F}^{\widehat{X}}(s) := \sigma(\widehat{X}(u): 0 \leq u \leq s) = \sigma(X(T-u): 0 \leq u \leq s), \quad 0 \leq s \leq T. \quad (\text{G.22})$$

Now recall (G.15); in its light, the filtration comparison in (G.22) implies $\mathcal{F}^{\widehat{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s)$, thus $\mathcal{H}(T-s) \subseteq \mathcal{F}^{\widehat{W}}(s) \subseteq \mathcal{F}^{\widehat{X}}(s)$ from (G.12), for $0 \leq s \leq T$, and from (G.11) also

$$\widehat{\mathcal{F}}(T-s) \subseteq \mathcal{F}^{\widehat{X}}(s), \quad 0 \leq s \leq T. \quad (\text{G.23})$$

But we have also the reverse inclusion $\mathcal{F}^{\widehat{X}}(s) \subseteq \widehat{\mathcal{F}}(T-s)$ on account of (G.10) and (G.22); therefore, $\mathcal{F}^{\widehat{X}}(s) = \widehat{\mathcal{F}}(T-s)$ holds for all $0 \leq s \leq T$ when $s(\cdot, \cdot)$ is invertible. These considerations inform our choice of backwards filtration $\mathcal{G}(T-s) \equiv \mathcal{F}^{\widehat{X}}(s)$, $0 \leq s \leq T$, in (3.21). \diamond

G.6. THE BACKWARDS ITÔ INTEGRAL

For two continuous semimartingales $X = X(0) + M + B$ and $Y = Y(0) + N + C$, with B, C continuous adapted processes of finite variation and M, N continuous local martingales, let us recall the definition of the Fisk-Stratonovich integral in [KS91, Definition 3.3.13, p. 156], as well as its properties in [KS91, Problem 3.3.14] and [KS91, Problem 3.3.15].

By analogy with this definition, we introduce the *backwards Itô integral*

$$\int_0^\cdot Y(t) \bullet dX(t) := \int_0^\cdot Y(t) dM(t) + \int_0^\cdot Y(t) dB(t) + \langle M, N \rangle, \quad (\text{G.24})$$

where the first (respectively, the second) integral on the right-hand side is to be interpreted in the Itô (respectively, the Lebesgue-Stieltjes) sense.

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

$$\sum_{j=0}^{m-1} Y(t_{j+1})(X(t_{j+1}) - X(t_j)) \quad (\text{G.25})$$

converge in probability to $\int_0^T Y(t) \bullet dX(t)$ 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

$$f(X) = f(X(0)) + \sum_{i=1}^n \int_0^\cdot D_i f(X(t)) \bullet dX_i(t) - \frac{1}{2} \sum_{i,j=1}^n \int_0^\cdot D_{ij}^2 f(X(t)) d\langle M_i, M_j \rangle(t), \quad (\text{G.26})$$

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.

REFERENCES

- [ABP⁺13] B. Acciaio, M. Beiglböck, F. Penkner, W. Schachermayer, and J. Temme. *A trajectorial interpretation of Doob's martingale inequalities*. The Annals of Applied Probability, 23(4):1494–1505, 2013.
- [AG13] L. Ambrosio and N. Gigli. *A User's Guide to Optimal Transport*. In Modelling and Optimisation of Flows on Networks, volume 2062 of Lecture Notes in Mathematics, pages 1–155. Springer, Berlin Heidelberg, 2013.
- [AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser, Basel, second edition, 2008.
- [Bac00] L. Bachelier. *Théorie de la spéculation*. Annales scientifiques de l'École Normale Supérieure, 17:21–86, 1900.
- [Bac06] L. Bachelier. *Louis Bachelier's Theory of Speculation — The Origins of Modern Finance. Translated and with Commentary by Mark Davis and Alison Etheridge*. Princeton University Press, Princeton, New Jersey, 2006.
- [BB00] J.-D. Benamou and Y. Brenier. *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*. Numerische Mathematik, 84(3):375–393, 2000.
- [BÉ85] D. Bakry and M. Émery. *Diffusions hypercontractives*. In Séminaire de Probabilités XIX, volume 1123 of Lecture Notes in Mathematics, pages 177–206. Springer, Berlin Heidelberg, 1985.
- [Bel95] D. R. Bell. *Degenerate stochastic differential equations and hypoellipticity*, volume 79 of Pitman Monographs and Surveys in Pure and Applied Mathematics. Longman, 1995.
- [BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*, volume 348 of Grundlehren der mathematischen Wissenschaften. Springer, 2014.
- [Bol96] L. Boltzmann. *Vorlesungen über Gastheorie — I. Theil*. Johann Ambrosius Barth Verlag, Leipzig, 1896.
- [Bol98a] L. Boltzmann. *Ueber die sogenannte H-Curve*. Mathematische Annalen, 50(2–3):325–332, 1898.
- [Bol98b] L. Boltzmann. *Vorlesungen über Gastheorie — II. Theil*. Johann Ambrosius Barth Verlag, Leipzig, 1898.
- [BS15] M. Beiglböck and P. Siorpaes. *Pathwise versions of the Burkholder-Davis-Gundy inequality*. Bernoulli, 21(1):360–373, 2015.
- [Car84] E. A. Carlen. *Conservative diffusions*. Communications in Mathematical Physics, 94(3):293–315, 1984.
- [CE02] D. Cordero-Erausquin. *Some Applications of Mass Transport to Gaussian-Type Inequalities*. Archive for Rational Mechanics and Analysis, 161(3):257–269, 2002.
- [Chu01] K. L. Chung. *A Course in Probability Theory*. Academic Press, 2001.

- [CS91] E. A. Carlen and A. Soffer. *Entropy Production by Block Variable Summation and Central Limit Theorems*. *Communications in Mathematical Physics*, 140(2):339–371, 1991.
- [CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley, Hoboken, New Jersey, second edition, 2006.
- [DK94] M. H. A. Davis and I. Karatzas. *A Deterministic Approach to Optimal Stopping*. In *Probability, Statistics and Optimization: A Tribute to Peter Whittle*, volume 104 of Wiley Series in Probability and Statistics — Applied Probability and Statistics Section, pages 455–466. Wiley, Chichester, New York, 1994.
- [Fer02] E. R. Fernholz. *Stochastic Portfolio Theory*, volume 48 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2002.
- [FJ16] J. Fontbona and B. Jourdain. *A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations*. *The Annals of Probability*, 44(1):131–170, 2016.
- [Föll85] H. Föllmer. *An entropy approach to the time reversal of diffusion processes*. In *Stochastic Differential Systems — Filtering and Control*, volume 69 of *Lecture Notes in Control and Information Sciences*, pages 156–163. Springer, Berlin Heidelberg, 1985.
- [Föll86] H. Föllmer. *Time reversal in Wiener space*. In *Stochastic Processes — Mathematics and Physics*, volume 1158 of *Lecture Notes in Mathematics*, pages 119–129. Springer, Berlin Heidelberg, 1986.
- [Fri75] A. Friedman. *Stochastic Differential Equations and Applications*, volume 1. Academic Press, New York, 1975.
- [Gar09] C. W. Gardiner. *Stochastic Methods. A Handbook for the Natural and Social Sciences*, volume 13 of *Springer Series in Synergetics*. Springer, Berlin Heidelberg, fourth edition, 2009.
- [Gen14] I. Gentil. *Logarithmic Sobolev inequality for diffusion semigroups*. In *Optimal Transportation — Theory and Applications*, volume 413 of *London Mathematical Society Lecture Note Series*, pages 41–57. Cambridge University Press, 2014.
- [GL10] N. Gozlan and C. Léonard. *Transport Inequalities. A Survey*. *Markov Processes and Related Fields*, 16(4):635–736, 2010.
- [GLR18] I. Gentil, C. Léonard, and L. Ripani. *Dynamical aspects of generalized Schrödinger problem via Otto calculus – A heuristic point of view*. *ArXiv e-prints*, 2018.
- [GLRT18] I. Gentil, C. Léonard, L. Ripani, and L. Tamanini. *An entropic interpolation proof of the HWI inequality*. *ArXiv e-prints*, 2018.
- [Gro75] L. Gross. *Logarithmic Sobolev Inequalities*. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
- [HP86] U. G. Hausmann and É. Pardoux. *Time Reversal of Diffusions*. *The Annals of Probability*, 14(4):1188–1205, 1986.
- [JK96] R. Jordan and D. Kinderlehrer. *An extended variational principle*. In *Partial Differential Equations and Applications*, volume 177 of *Lecture Notes in Pure and Applied Mathematics*, chapter 18, pages 187–200. CRC Press, 1996.

- [JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. *The variational formulation of the Fokker-Planck equation*. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
- [Kol31] A. N. Kolmogorov. *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*. *Mathematische Annalen*, 104(1):415–458, 1931.
- [Kol37] A. N. Kolmogorov. *Zur Umkehrbarkeit der statistischen Naturgesetze*. *Mathematische Annalen*, 113(1):766–772, 1937.
- [KS91] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 1991.
- [Léo14] C. Léonard. *Some Properties of Path Measures*. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Mathematics*, pages 207–230. Springer, 2014.
- [McC97] R. J. McCann. *A Convexity Principle for Interacting Gases*. *Advances in Mathematics*, 128(1):153–179, 1997.
- [Mey94] P. A. Meyer. *Sur une transformation du mouvement brownien due à Jeulin et Yor*. In *Séminaire de Probabilités XXVIII*, volume 1583 of *Lecture Notes in Mathematics*, pages 98–101. Springer, Berlin Heidelberg, 1994.
- [MNS89] A. Millet, D. Nualart, and M. Sanz. *Integration by Parts and Time Reversal for Diffusion Processes*. *The Annals of Probability*, 17(1):208–238, 1989.
- [MV99] P. A. Markowich and C. Villani. *On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis*. In *Physics and Functional Analysis, Matematica Contemporanea (SBM) 19*, pages 1–29, 1999.
- [Nel01] E. Nelson. *Dynamical Theories of Brownian motion*. Princeton University Press, second edition, 2001.
- [Nua06] D. Nualart. *The Malliavin Calculus and Related Topics*. *Probability and Its Applications*. Springer, Berlin Heidelberg, second edition, 2006.
- [Ott01] F. Otto. *The geometry of dissipative evolution equations: the porous medium equation*. *Communications in Partial Differential Equations*, 26(1–2):101–174, 2001.
- [OV00] F. Otto and C. Villani. *Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality*. *Journal of Functional Analysis*, 173(2):361–400, 2000.
- [Par86] É. Pardoux. *Grossissement d’une filtration et retournement du temps d’une diffusion*. In *Séminaire de Probabilités XX*, volume 1204 of *Lecture Notes in Mathematics*, pages 48–55. Springer, Berlin Heidelberg, 1986.
- [Ris96] H. Risken. *The Fokker-Planck equation. Methods of Solution and Applications*, volume 18 of *Springer Series in Synergetics*. Springer, Berlin Heidelberg, second edition, 1996.
- [Rog85] L. C. G. Rogers. *Smooth Transition Densities for One-Dimensional Diffusions*. *Bulletin of the London Mathematical Society*, 17(2):157–161, 1985.
- [RVW01] F. Russo, P. Vallois, and J. Wolf. *A generalized class of Lyons-Zheng processes*. *Bernoulli*, 7(2):363–379, 2001.
- [RW00a] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Volume 1: Foundations*. *Cambridge Mathematical Library*. Cambridge University Press, second edition, 2000.

- [RW00b] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Volume 2: Itô Calculus*. Cambridge Mathematical Library. Cambridge University Press, second edition, 2000.
- [Sch31] E. Schrödinger. *Über die Umkehrung der Naturgesetze*. Sitzungsberichte der Preussischen Akademie der Wissenschaften: Physikalisch-Mathematische Klasse, pages 144–153, 1931.
- [Sch32] E. Schrödinger. *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique*. Annales de l'Institut Henri Poincaré, 2(4):269–310, 1932.
- [Sch80] Z. Schuss. *Singular Perturbation Methods in Stochastic Differential Equations of Mathematical Physics*. SIAM Review, 22(2):119–155, 1980.
- [Sta59] A. J. Stam. *Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon*. Information and Control, 2:101–112, 1959.
- [Stu06a] K.-T. Sturm. *On the geometry of metric measure spaces. I*. Acta Mathematica, 196(1):65–131, 2006.
- [Stu06b] K.-T. Sturm. *On the geometry of metric measure spaces. II*. Acta Mathematica, 196(1):133–177, 2006.
- [Tal96] M. Talagrand. *Transportation cost for Gaussian and other product measures*. Geometric and Functional Analysis, 6(3):587–600, 1996.
- [Vil03] C. Villani. *Topics in Optimal Transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island, 2003.
- [Wal82] J. B. Walsh. *A non-reversible semimartingale*. In Séminaire de Probabilités XVI, volume 920 of Lecture Notes in Mathematics, page 212. Springer, Berlin Heidelberg, 1982.
- [Wil91] D. Williams. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, 1991.