Optimal Couplings on Wiener Space and an Extension of Talagrand's Transport Inequality

Hans Föllmer*

July 10, 2021

Abstract

For a probability measure Q on Wiener space, Talagrand's transport inequality takes the form $W_{\mathcal{H}}(Q,P)^2 \leq 2H(Q|P)$, where the Wasserstein distance $W_{\mathcal{H}}$ is defined in terms of the Cameron-Martin norm, and where $H(Q|P)$ denotes the relative entropy with respect to Wiener measure P. Talagrand's original proof takes a bottom-up approach, using finite-dimensional approximations. As shown by Feyel and Üstünel in [3] and Lehec in [10], the inequality can also be proved directly on Wiener space, using a suitable coupling of Q and P. We show how this top-down approach can be extended beyond the absolutely continuous case $Q \ll P$. Here the Wasserstein distance is defined in terms of quadratic variation, and $H(Q|P)$ is replaced by the specific relative entropy $h(Q|P)$ on Wiener space that was introduced by N. Gantert in [7].

1 Introduction

There are many ways of quantifying the extent to which a probability measure Q on the path space $C[0,1]$ deviates from Wiener measure P. In this paper we discuss the following two approaches and the relation between them. One involves the notion of entropy, the other uses a Wasserstein distance, that is, the solution of an optimal transport problem on Wiener space. We will do this in two stages.

In the first stage, the measure Q will be absolutely continuous with respect to Wiener measure P, and we consider the relative entropy $H(Q|P)$ of Q with respect to P. On the other hand, we use the Wasserstein distance

$$W_{\mathcal{H}}(Q,P) = \inf \Big(\int ||\omega - \eta)||_{\mathcal{H}} P(d\omega) R(\omega, d\eta) \Big)^{1/2}, \tag{1}$$

where the infimum is taken over all transition kernels R on Wiener space which transport P into Q, and where the transportation cost is defined by

*Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, 10099 Berlin, Germany. Email: foellmer@math.hu-berlin.de.

the Cameron-Martin norm. Talagrand's transport inequality

$$W_{\mathcal{H}}(Q,P) \le \sqrt{2H(Q|P)} \tag{2}$$

on Wiener space shows that these two measures of deviation are closely related. In fact, inequality (2) becomes an identity as soon as we introduce the additional constraint that the transport should be adapted to the natural filtration on Wiener space; this was first shown by R. Lassalle in [9].

On Wiener space, inequality (2) was first studied by Feyel and Üstünel [3]. In Talagrand's original version [13], the inequality is formulated on Euclidean space $\mathbb{R}^n$, including the case $n = \infty$; the Wasserstein distance is defined in terms of the Euclidean norm, and the reference measure P is the product of standard normal distributions. But the Lévy-Ciesielski construction of Brownian motion in terms of the Schauder functions shows that inequality (2) on Wiener space can be viewed as a direct translation of the Euclidean case for $n = \infty$, as explained in Section 3.

Talagrand's original proof in [13] takes a bottom-up approach, using finite-dimensional approximations. Instead, as shown by D. Feyel and A. S. Üstünel in [3] and by J. Lehec in [10], Talagrand's inequality can be proved directly on Wiener space, using a suitable coupling of Q and P. This top-down approach involves the computation of relative entropy in terms of the *intrinsic drift* of Q that was used in [4] and [5] for the analysis of time reversal and large deviations on Wiener space. The intrinsic drift b^Q is such that Q can be viewed as a weak solution of the stochastic differential equation $dW = dW^Q + b^Q(W)dt$, that is, W^Q is a Wiener process under Q. Coupling W^Q with the coordinate process W under Q immediately yields inequality (2), and it solves the optimal transport problem for the Cameron-Martin norm if the coupling is required to be adapted.

Clearly, inequality (2) is of interest only if the relative entropy is finite, and so Q should be absolutely continuous with respect to Wiener measure. In the second stage, we go beyond this restriction. Here we replace $H(Q|P)$ by the *specific relative entropy*

$$h(Q|P) := \liminf_{N\uparrow\infty} 2^{-N} H_N(Q|P),$$

where $H_N(Q|P)$ denotes the relative entropy of Q with respect to P on the σ-field generated by observing the path along the N-th dyadic partition of the unit interval. The notion of specific relative entropy on Wiener space was introduced by N. Gantert in her thesis [7], where it serves as a rate function for large deviations of the quadratic variation from its ergodic behaviour; cf. also [8]. In our context, the specific relative entropy appears if we rewrite the finite-dimensional Talagrand inequality for $n = 2^N$ in the form

$$W_N^2(Q,P) \le 2 \cdot 2^{-N} H_N(Q|P), \tag{3}$$

where the Wasserstein metric W_N is defined in terms of the discrete quadratic variation along the N-th dyadic partition. This suggests that a passage to the limit should yield an extension of Talagrand's inequality, where $H(Q|P)$ is replaced by $h(Q|P)$, and where $W_{\mathcal{H}}$ is replaced by a Wasserstein metric $W_{\mathcal{S}}$ that is defined in terms of quadratic variation. Here again, we take a top-down approach. Instead of analyzing the convergence on the left-hand side of (3), we argue directly on Wiener space, assuming that the coordinate process W is a special semimartingale under Q. We show that $h(Q|P) < \infty$ implies that Q admits the construction of an intrinsic Wiener process W^Q such that the pair (W, W^Q) defines a coupling of P and Q. This coupling solves the optimal transport problem defined by $W_{\mathcal{S}}$, and for a martingale measure Q it yields the inequality

$$W_{\mathcal{S}}(Q, P) \leq \sqrt{2h(Q|P)}. \tag{4}$$

If, more generally, Q is a semimartingale measure that admits a unique equivalent martingale measure Q^*, then we obtain the following extension of Talagrand's inequality on Wiener space:

$$W_{\mathcal{S}}(Q|P)^2 \leq 2\big(h(Q|P) + H(Q|Q^*)\big). \tag{5}$$

In this form, inequality (5) includes both (4) and Talagrand's inequality (2) as special cases.

The paper is organized as follows. In Section 2 we introduce the basic concepts of relative entropy and of a Wasserstein distance. Section 3 describes the top-down approach to inequality (2) in the absolutely continuous case; the exposition will be reasonably self-contained because we repeatedly refer to it in the sequel. In the second stage, we consider measures Q on $C[0,1]$ such that the coordinate process W is a semimartingale under Q. Section 4 shows how the semimartingale structure of Q is reflected in the specific relative entropy $h(Q|P)$; this extends Theorem 1.2 in [7] for martingale measures to the general case. In section 5 we show that the condition $h(Q|P) < \infty$ implies that Q admits the construction of an intrinsic Wiener process W^Q. Coupling W^Q with the coordinate process W under Q, we obtain the solution of an optimal transport problem on Wiener space that yields inequalities (4) and (5).

2 Preliminaries

In this section we recall some basic notions, in particular the definitions of relative entropy and of the Wasserstein distances that we are going to use.

For two probability measures μ and ν on some measurable space $(S, \mathcal{S})$, the *relative entropy* of ν with respect to μ is defined as

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

For $\nu \ll \mu$ we can write

$$H(\nu|\mu) = \int h(\frac{d\nu}{d\mu}) d\mu,$$

denoting by h the strictly convex function $h(x) = x \log x$ on $[0, \infty)$, and Jensen's inequality implies $H(\nu|\mu) \geq 0$, with equality if and only if $\mu = \nu$. Sometimes we will deal with different σ-fields $\mathcal{S}$ on the same space S, and then we will also use the notation $H_{\mathcal{S}}(\nu|\mu)$. We are going to use repeatedly the fact that

$$\lim_{n \uparrow \infty} H_{\mathcal{S}_n}(\nu|\mu) = H_{\mathcal{S}}(\nu|\mu) \tag{6}$$

if $(\mathcal{S}_n)_{n=1,2\ldots}$ is a sequence of σ-fields increasing to $\mathcal{S}$.

Consider a measurable cost function $c(\cdot, \cdot)$ on $S \times S$ with values in $[0, \infty]$; typically, $c(\cdot, \cdot)$ will be a metric on S. We define the corresponding *Wasserstein distance* between ν and μ as

$$W(\nu, \mu) = \inf_{\gamma \in \Gamma(\mu, \nu)} (\int c^2(x, y) \gamma(dx, dy))^{1/2},$$

where $\Gamma(\mu, \nu)$ denotes the class of all probability measures γ on the product space $S \times S$ with marginals μ and ν. Equivalently, we can write

$$W(\nu, \mu) = \inf \tilde{E}[c^2(\tilde{X}, \tilde{Y})]^{1/2},$$

where the infimum is taken over all couples $(\tilde{X}, \tilde{Y})$ of S-valued random variables on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $\tilde{X}$ and $\tilde{Y}$ have distributions μ and ν, respectively. Such a couple, and also any measure $\gamma \in \Gamma(\mu, \nu)$, will be called a *coupling of μ and ν*. We refer to [15] for a thorough discussion of Wasserstein distances in various contexts.

In the sequel, the space S will be either a Euclidean space $\mathbb{R}^n$, including the infinite-dimensional case $n = \infty$, or the space

$$\Omega = C_0[0, 1]$$

of all continuous functions ω on $[0, 1]$ with initial value $\omega(0) = 0$.

For $S = \mathbb{R}^n$ with $n \in \{1, \ldots, \infty\}$ we are going to use the cost function $c(x, y) = ||x - y||_n$, defined by the Euclidean norm $||x||_n = (\sum_{i=1}^n x_i^2)^{1/2}$. Thus, the corresponding Wasserstein distance is given by

$$W_n(\nu, \mu) = \inf_{\gamma \in \Gamma(\mu, \nu)} (\int ||x - y||_n^2) \gamma(dx, dy))^{1/2}.$$

Taking as reference measure the Gaussian measure

$$\mu_n = \prod_{i=1}^n N(0, 1),$$

Talagrand's inequality on Euclidean space can now be stated as follows:

Theorem 1. *For any $n \in \{0, \ldots, \infty\}$ and for any probability measure ν on $\mathbb{R}^n$,*

$$W_n(\nu, \mu_n) \leq \sqrt{2H(\nu|\mu_n)}. \tag{7}$$

Talagrand's proof in [13] takes a bottom-up approach. First the inequality is proved in the one-dimensional case, using Vallender's expression

$$W_1(\nu,\mu) = \Big(\int_0^1 \big(q_\nu(\alpha) - q_\mu(\alpha)\big)^2 d\alpha \Big)^{1/2} \tag{8}$$

in [14] of the Wasserstein distance on $\mathbb{R}^1$ in terms of the quantile functions q_ν and q_μ, followed by an integration by parts that involves the special form of the normal distribution. The finite-dimensional case is shown by induction, applying the one-dimensional inequality to the conditional distributions $\nu(dx_{n+1}|x_1,\dots,x_n)$ of ν. The infinite-dimensional case $n=\infty$ follows by applying (7) to the finite-dimensional marginals and taking the limit $n\uparrow\infty$, using a standard martingale argument to obtain convergence of the relative entropies on the right-hand side.

Let us now turn to the case $S=\Omega=C_0[0,1]$. We denote by $(\mathcal{F}_t)_{0\le t\le 1}$ the right-continuous filtration on Ω generated by the coordinate process

$$W = (W_t)_{0\le t\le 1}$$

defined by $W_t(\omega)=\omega(t)$. We set $\mathcal{F}=\mathcal{F}_1$ and denote by P the *Wiener measure* on $(\Omega,\mathcal{F})$. Let $\mathcal{H}$ denote the *Cameron-Martin space* of all absolutely continuous functions $\omega\in\Omega$ such that the derivative $\dot\omega$ is square integrable on $[0,1]$. First we will consider the cost function $c(\omega,\eta)=||\omega-\eta||_{\mathcal{H}}$, where

$$||\omega||_{\mathcal{H}} = \begin{cases} (\int_0^1 \dot\omega^2(t)dt)^{1/2} & \text{if } \omega\in\mathcal{H} \\ +\infty & \text{otherwise.} \end{cases}$$

The corresponding Wasserstein distance will be denoted by $W_{\mathcal{H}}$, that is,

$$W_{\mathcal{H}}(Q,P) = \inf_{\gamma\in\Gamma(P,Q)} \int ||\omega-\eta||^2_{\mathcal{H}}\gamma(d\omega,d\eta)^{1/2},$$

for any probability measure Q on $(\Omega,\mathcal{F})$. In this setting, Talagrand's inequality takes the following form, first stated by D. Feyel and A. S. Ustunel in [3].

Theorem 2. *For any probability measure Q on $(\Omega,\mathcal{F})$,*

$$W_{\mathcal{H}}(Q,P) \le \sqrt{2H(Q|P}. \tag{9}$$

In fact, inequality (9) can be viewed as a direct translation of Talagrand's inequality on $\mathbb{R}^\infty$. To see this, recall the Lévy-Ciesielski representation

$$W_t(\omega) = \sum_{i\in I} X_i(\omega)e_i(t)$$

of Brownian motion in terms of the Schauder basis $(e_i)_{i\in I}$ of $C_0[0,1]$. Under Wiener measure P, the coordinates X_i are independent with distribution $N(0,1)$. Thus, the random vector $(X_i(\omega))_{i\in I}$, viewed as a measurable map T

from Ω to R^∞, has distribution μ_∞ under P. Relative entropy is invariant under T, and so we get

$$H(\nu|\mu_\infty) = H(Q|P),$$

where ν denotes the image of Q under T. On the other hand we have $||\omega||_{\mathcal{H}} = ||(X_i(\omega))_{i\in I}||_\infty$, and this implies

$$W_{\mathcal{H}}(Q,P) = W_\infty(\nu,\mu_\infty).$$

Thus, Talagrand's inequality (7) for $n = \infty$ translates into inequality (9) on Wiener space.

Having scetched the bottom-up approach to Talagrand's inequality on Wiener space, we are now going to focus on the top-down approach. It consists in proving Talagrand's inequality (9) directly on Wiener space, using a suitable coupling of Q and P.

3 Intrinsic drift and optimal coupling in the absolutely continuous case

Take any probability measure Q on $(\Omega,\mathcal{F})$ that is absolutely continuous with respect to Wiener measure P. Let us first recall the following computation of the relative entroopy $H(Q|P)$ in terms of the *intrinsic drift* of Q ; cf. [4], [5] or, for the first two parts, Th. 7.11 in [11].

Proposition 3. *There exists a predictable process $b^Q = (b_t^Q(\omega))_{0\le t\le 1}$ with the following properties:*
1*)*

$$\int_0^1 \left(b_t^Q(\omega)\right)^2 dt < \infty \quad Q\text{-a.s.}, \tag{10}$$

that is, the process B^Q defined by $B_t^Q(\omega) = \int_0^t b_s^Q(\omega)ds$ satisfies

$$B^Q(\omega) \in \mathcal{H} \quad Q\text{-a.s.}$$

2*) $W^Q := W - B^Q$ is a Wiener process under Q, that is, W is a special semimartingale under Q with canonical decomposition*

$$W = W^Q + B^Q.$$

3) The relative entropy of Q with respect to P is given by

$$H(Q|P) = \frac{1}{2}E_Q\big[\int_0^1 (b_t^Q)^2 dt\big] = \frac{1}{2}E_Q\big[||B^Q||^2_{\mathcal{H}}\big]. \tag{11}$$

The process b^Q will be called the intrinsic drift *of Q.*

Proof. For the convenience of the reader we scetch the argument; cf., e.g., [5] for details.
1) By Itô's representation theorem, the density $\phi = \frac{dQ}{dP}$ can be represented as a stochastic integral of the Brownian motion W, that is, there exists a predictable process $(\xi_t)_{0\le t\le 1}$ such that $\int_0^1 \xi_t(\omega)dt < \infty \quad P-a.s.$ and

$$\phi = 1 + \int_0^1 \xi_t dW_t \qquad P\text{-a.s.}$$

Moreover, the process

$$\phi_t := E_P[\phi|\mathcal{F}_t] = 1 + \int_0^t \xi_s dW_s, \qquad 0 \le t \le 1,$$

is a continuous martingale with quadratic variation

$$\langle\phi\rangle_t = \int_0^t \xi_s^2 ds \qquad P\text{-a.s.}$$

and

$$\inf_{0\le t\le 1} \phi_t > 0 \quad P\text{-a.s. on } \{\phi > 0\},$$

hence Q-a.s.. Thus, the predictable process b^Q defined by

$$b_t^Q := \frac{\xi_t}{\phi_t} I_{\{\phi_t>0\}}, \qquad 0 \le t \le 1,$$

satisfies the integrability condition (10).
2) Applying Itô's formula to $\log \phi_t$, we get

$$\begin{aligned}
\log \phi_t &= \int_0^t \frac{1}{\phi_s} d\phi_s - \frac{1}{2}\int_0^t (\frac{1}{\phi_s})^2 d\langle\phi\rangle_s \\
&= \int_0^t b_s^Q dW_s - \frac{1}{2}\int_0^t (b_s^Q)^2 ds \\
&= \int_0^t b_s^Q dW_s^Q + \frac{1}{2}\int_0^t (b_s^Q)^2 ds
\end{aligned}$$

The second part now follows from Girsanov's theorem.
3) Equation (11) for $H(Q|P) = E_Q[\log \phi_1]$ follows from the preceding equation for $t = 1$. Indeed, if $E_Q\big[\int_0^1 (b_s^Q)^2 ds\big] < \infty$ then we get

$$E_Q\big[\int_0^1 b_s^Q dW_s^Q\big] = 0,$$

and this implies (11). In the general case, the same argument applies up to each stopping time $T_n = \inf\{t| \int_0^t (b_s^Q)^2 ds > n\} \wedge 1$, and for $n \uparrow \infty$ we obtain (11). $\square$

Remark 4. *Apart from our present purpose, the intrinsic drift of Q is also an efficient tool in proving a number of inequalities, including logarithmic Sobolev and Shannon-Stam inequalities; see [10] and [2].*

As observed by J. Lehec in [10], proposition 3 can be rephrased as follows in terms of coupling, and in this form it yields an immediate proof of Talagrand's inequality on Wiener space.

Proposition 5. *The processes $W^Q = W - B^Q$ and W, defined on the probability space $(\Omega, \mathcal{F}, Q)$, form a coupling of P and Q such that*

$$E_Q\big[||W - W^Q||^2_{\mathcal{H}}\big] = 2H(Q|P). \tag{12}$$

Corollary 6. *Any probability measure Q on $(\Omega, \mathcal{F})$ satisfies Talagrand's inequality*

$$W_{\mathcal{H}}(Q, P) \leq \sqrt{2H(Q|P)}. \tag{13}$$

Proof. If Q is not absolutely continuous with respect to Wiener measure P then we have $H(Q|P) = \infty$, and (13) holds trivially. In the absolutely continuous case, inequality (13) follows immediately from equation (12) and the definition of the Wasserstein distance $W_{\mathcal{H}}$. □

Note that the coupling (W^Q, W) of P and Q, which is defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$, is *adaptive* in the following sense.

Definition 7. *A coupling $(\tilde{X}, \tilde{Y})$ of P and Q will be called an* adaptive coupling*, if it is defined on a filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0\leq t\leq 1}, \tilde{P})$ such that*

1. *$\tilde{Y} = (\tilde{Y}_t)$ is adapted with respect to $\tilde{P}$ and $(\tilde{\mathcal{F}}_t)_{0\leq t\leq 1}$,*
2. *$\tilde{X}$ is a Wiener process with respect to $\tilde{P}$ and $(\tilde{\mathcal{F}}_t)_{0\leq t\leq 1}$. that is, each increment $\tilde{X}_t - \tilde{X}_s$ is independent of $\tilde{\mathcal{F}}_s$ with law $N(0, t-s)$.*

Theorem 8. *The* optimal adaptive coupling *of P and Q is given by (W^Q, W), that is,*

$$E_Q\big[||W - W^Q||^2_{\mathcal{H}}\big] \leq \tilde{E}\big[||\tilde{Y} - \tilde{X}||^2_{\mathcal{H}}\big], \tag{14}$$

for any adaptive coupling $(\tilde{X}, \tilde{Y})$ of P and Q, and equality holds iff

$$\tilde{Y} = W^Q(\tilde{Y}) + B^Q(\tilde{Y}), \qquad \tilde{P} - \text{a.s.}. \tag{15}$$

Proof. Take any adapted coupling $(\tilde{X}, \tilde{Y})$ of P and Q, defined on a filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0\leq t\leq 1}, \tilde{P})$, such that

$$\tilde{E}\big[||\tilde{Y} - \tilde{X}||^2_{\mathcal{H}}\big] < \infty.$$

Since $\tilde{Y}$ is adapted with continuous paths, $\tilde{B} := \tilde{Y} - \tilde{X}$ is an adapted continuous process such that $\tilde{E}\big[||B||^2_{\mathcal{H}}\big] < \infty$. This implies $\tilde{B}_t = \int_0^t \tilde{b}_s ds$ for some predictable process $\tilde{b} = (\tilde{b}_s)_{0\leq s\leq 1}$ such that $\tilde{E}\big[\int_0^1 \tilde{b}_s^2 ds\big] < \infty$. Since $\tilde{X}$ is a

Brownian motion with respect to the filtration $(\tilde{\mathcal{F}}_t)$, the process $\tilde{Y}$ is a special semimartingale with canonical decomposition

$$\tilde{Y}_t = \tilde{X}_t + \int_0^t \tilde{b}_s ds \tag{16}$$

under $\tilde{P}$ with respect to $(\tilde{\mathcal{F}}_t)$. On the other hand, since $\tilde{Y}$ has law Q under $\tilde{P}$ and W^Q is a Brownian motion under Q, the process $W^Q(\tilde{Y})$ is a Brownian motion under $\tilde{P}$ with respect to the smaller filtration $(\tilde{\mathcal{F}}_t^0)$ generated by the adapted process $\tilde{Y}$. Thus, $\tilde{Y}$ has the canonical decomposition

$$\tilde{Y}_t = W_t^Q(\tilde{Y}) + \int_0^t b_s^Q(\tilde{Y}) ds \tag{17}$$

under $\tilde{P}$ with respect to $(\tilde{\mathcal{F}}_t^0)$. This implies

$$b_t^Q(\tilde{Y}) = \tilde{E}\big[\,\tilde{b}_t \,|\, \tilde{\mathcal{G}}_t\big] \qquad \tilde{P} \otimes dt - \text{a.s.}; \tag{18}$$

cf., for example, Th. 8.1 in [11] or the proof of equation 68 in the general context of Proposition 35 below. Applying Jensen's inequality, we obtain

$$\begin{aligned} \tilde{E}\big[||\tilde{Y} - \tilde{X}||_{\mathcal{H}}^2\big] &= \tilde{E}\big[\int_0^1 \tilde{b}_t^2 dt\big] \\ &\geq \tilde{E}\big[\int_0^1 (b_t^Q(\tilde{Y}))^2 dt\big] = E_Q\big[\int_0^1 (b_t^Q(W))^2 dt\big] \\ &= 2H(Q|P). \end{aligned}$$

Equality holds iff

$$\tilde{b}_t = b_t^Q(\tilde{Y}) \qquad \tilde{P} \otimes dt - \text{a.s.},$$

and in this case (16) and (17) imply $\tilde{X} = W^Q(\tilde{Y})$ $\tilde{P}$-a.s.. □

Let us define $W_{\mathcal{H},ad}(Q,P)$ as the infimum of the right hand side in (14), taken only over the *adaptive* couplings of P and Q. Clearly we have

$$W_{\mathcal{H}}(Q,P) \leq W_{\mathcal{H},ad}(Q,P), \tag{19}$$

and Theorem 8 shows that the following identity holds, first proved by R. Lassalle in [9].

Corollary 9. *For any probability measure Q on $(\Omega, \mathcal{F})$ we have*

$$W_{\mathcal{H},ad}(Q,P) = \sqrt{2H(Q|P)}. \tag{20}$$

Remark 10. *For a thorough discussion of optimal transport problems on Wiener space under various constraints, with special emphasis on the effects of an enlargement of filtration, we refer to [1].*

The following example illustrates the difference between $W_{\mathcal{H}}$ and $W_{\mathcal{H},ad}$. It also shows how the finite-dimensional inequalities in (7) can be derived from Talagrand's inequality on Wiener space, thus completing the top-down approach.

For a probability measure ν on $\mathbb{R}^1$ we introduce the probability measure

$$Q^\nu = \int P^x \nu(dx)$$

on $(\Omega, \mathcal{F})$, where P^x denotes the law of the Brownian bridge from 0 to $x \in \mathbb{R}^1$. If $\nu \ll \mu := N(0,1)$, then Q^ν is absolutely continuous with respect to P with density

$$\frac{dQ^\nu}{dP} = \frac{d\nu}{d\mu}(W_1),$$

and the relative entropy is given by

$$H(Q^\nu|P) = \int \log \frac{d\nu}{d\mu}(W_1) dQ^\nu = \int \log \frac{d\nu}{d\mu} d\nu = H(\nu|\mu). \tag{21}$$

Corollary 11. *We have*

$$W_{\mathcal{H}}(Q^\nu, P) = W_1(\nu, \mu) \quad \text{and} \quad W_{\mathcal{H},ad}(Q^\nu, P) = \sqrt{2H(\nu|\mu)}. \tag{22}$$

Thus, inequality (19) *implies*

$$W_1(\nu, \mu) \leq \sqrt{2H(\nu|\mu)}. \tag{23}$$

Inequality (23) *is strict except for the case where* $\nu = N(m, 1)$ *for some* $m \in \mathbb{R}^1$.

Proof. 1) The second identity in (22) follows from Corollary 9 and equation (21).
2) To prove the first identity, take any coupling $(\tilde{X}, \tilde{Y})$ of P and Q, defined on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, such that $Z := \tilde{Y} - \tilde{X} \in \mathcal{H}$. Then the endpoints $\tilde{X}_1$ and $\tilde{Y}_1$ form a coupling of μ and ν. Since

$$(\tilde{Y}_1 - \tilde{X}_1)^2 = Z_1^2 = (\int_0^1 \dot{Z}_s ds)^2 \leq \int_0^1 \dot{Z}_s^2 ds = ||\tilde{Y} - \tilde{X}||_{\mathcal{H}}^2,$$

we obtain

$$W_1^2(\nu, \mu) \leq \tilde{E}\big[(\tilde{Y}_1 - \tilde{X}_1)^2\big] \leq \tilde{E}\big[||\tilde{Y} - \tilde{X}||_{\mathcal{H}}^2\big],$$

hence

$$W_1^2(\nu, \mu) \leq W_{\mathcal{H}}^2(Q, P). \tag{24}$$

We now show that the lower bound $W_1^2(\nu, \mu)$ is attained by the following coupling (W, Y) of P and Q^ν, defined on the Wiener space $(\Omega, \mathcal{F}, P)$. The process Y is given by

$$Y_t = W_t + t\big(f_\nu(W_1) - W_1\big), \quad 0 \leq t \leq 1,$$

where $f_\nu(x) = q_\nu(\Phi(x))$ and Φ denotes the distribution function of $\mu = N(0,1)$. The endpoint $Y_1 = f_\nu(W_1)$ has distribution ν under P, and the conditional distribution of Y given the endpoint $Y_1 = y$ coincides with the Brownian bridge P^y. Thus Y has distribution Q^ν under P, and (W,Y) is a coupling of P and Q^ν, defined on $(\Omega, \mathcal{F}, P)$. Note that this coupling is not adaptive with respect to the filtration $(\mathcal{F}_t)$, since Y anticipates the endpoint W_1 of the Brownian path. Since $||Y - W||^2_{\mathcal{H}} = (f_\nu(W_1) - W_1)^2$, we get

$$\begin{aligned} E_P\big[||Y-W||^2_{\mathcal{H}}\big] &= \int \big(f_\nu(x) - x\big)^2 \mu(dx) \\ &= \int_0^1 \big(q_\nu(\alpha) - \Phi^{-1}(\alpha)\big)^2 d\alpha = W_1^2(\nu,\mu), \end{aligned}$$

using equation (8) in the last step. This completes the proof of the first identity in (22)

3) Let us write $Q = Q^\nu$. Theorem 8 shows that the optimal adapted coupling of Q and P is given by (W, W^Q) under Q. Since

$$(W_1 - W_1^Q)^2 = (\int_0^1 b_t^Q dt)^2 \leq \int_0^1 (b_t^Q)^2 dt = ||B^Q||^2_{\mathcal{H}}$$

and

$$W_1^2(\nu,\mu) \leq E_Q\big[(W_1 - W_1^Q)^2\big] \leq E_Q\big[||B^Q||^2_{\mathcal{H}}\big] = 2H(\nu|\mu),$$

equality in (23) implies, Q-a.s., that $b_t^Q(\cdot)$ is almost everywhere constant in t, hence equal to $m(\cdot) := W_1 - W_1^Q$. Since the process b^Q is adapted to the filtration $(\mathcal{F}_t)$, $m(\cdot)$ is measurable with respect to $\mathcal{F}_0 = \bigcap_{t>0} \mathcal{F}_t$. But P is 0-1 on $\mathcal{F}_0$, and the same is true for $Q \ll P$. This implies $m(\cdot) = m \quad Q$-a.s. for some $m \in R^1$, that is, $W_1 = W_1^Q + m$ and $\nu = N(m,1)$. □

Talagrand's inequality in any finite dimension $n > 1$ follows in the same manner. For our purpose it is convenient to use the following equivalent version, where the reference measure is taken to be

$$\tilde{\mu}_n = \prod_{i=1}^n N(0, \frac{1}{n})$$

instead of $\mu_n = \prod_{i=1}^n N(0,1)$ as in (7).

Corollary 12. *For any probability measure ν on $\mathbb{R}^n$,*

$$nW_n^2(\nu, \tilde{\mu}_n) \leq 2H(\nu|\tilde{\mu}_n). \tag{25}$$

Proof. We may assume $\nu \ll \tilde{\mu}_n$. Let $T_n : \Omega \to \mathbb{R}^n$ denote the map that associates to each path ω the vector of its increments $\omega(i/n) - \omega((i-1)/n) \quad (i = 1, \ldots, n)$. Under Wiener measure P, the distribution of T_n is given by $\tilde{\mu}_n$. Define Q^ν on $(\Omega, \mathcal{F})$ by

$$\frac{dQ^\nu}{dP} = \frac{d\nu}{d\mu}(T_n).$$

For any coupling $(\tilde{X}, \tilde{Y})$ of P and Q^ν such that $Z := \tilde{Y} - \tilde{X} \in \mathcal{H}$, the vectors $X_n := T_n(\tilde{X})$ and $Y_n := T_n(\tilde{Y})$ form a coupling of ν and $\tilde{\mu}_n$. Since

$$||X_n - Y_n||^2 = \sum_{i=1}^{n} (\int_{(i-1)/n}^{i/n} \dot{Z}_s ds)^2 \leq \sum_{i=1}^{n} \frac{1}{n} \int_{(i-1)/n}^{i/n} \dot{Z}_s^2 ds = \frac{1}{n} ||\tilde{Y} - \tilde{X}||_{\mathcal{H}}^2,$$

we obtain

$$W_n^2(\nu, \tilde{\mu}_n) \leq \tilde{E}\big[||Y_n - X_n||^2\big] \leq \frac{1}{n} \tilde{E}\big[(||\tilde{Y} - \tilde{X}||_{\mathcal{H}}^2\big],$$

hence

$$W_n^2(\nu, \tilde{\mu}_n) \leq \frac{1}{n} W_{\mathcal{H}}^2(Q, P) \leq \frac{2}{n} H(Q^\nu | P).$$

due to Corollary 6. Since $H(Q^\nu|P) = H(\nu|\tilde{\mu}_n)$, we have proved (25). □

4 Specific Relative Entropy

The following concept of specific relative entropy on Wiener space was introduced by N. Gantert in her thesis [7], where it plays the role of a rate function for large deviations of the quadratic variation from its ergodic behaviour; cf. also [8]. In our context, it will allow us to extend Talagrand's inequality on Wiener space beyond the absolutely continuous case $Q \ll P$.

From now on, the index N will refer to the N-th dyadic partition of the unit interval, that is, $D_N = \{k2^{-N} | k = 1, \ldots, 2^N\}$. In particular we introduce the discretized filtration

$$\mathcal{F}_{N,t} = \sigma(\{W_s | s \in D_N, s \leq t\}), \quad 0 \leq t \leq 1$$

on $\Omega = C_0[0,1]$, and we set $\mathcal{F}_N = \mathcal{F}_{N,1} = \sigma(\{W_s | s \in D_N\})$.

Definition 13. *For any probability measure Q on $(\Omega, \mathcal{F})$, the* specific relative entropy *of Q with respect to Wiener measure P is defined as*

$$h(Q|P) = \liminf_{N \uparrow \infty} 2^{-N} H_N(Q|P), \tag{26}$$

where $H_N(Q|P)$ denotes the relative entropy of Q with respect to P on the σ-field $\mathcal{F}_N$.

Since $H(Q|P) = \lim_N H_N(Q|P)$, we get $h(Q|P) = 0$ for any Q such that $H(Q|P) < \infty$. Thus, the notion of specific relative entropy is of interest only if we look beyond the cases that we have considered so far.

Remark 14. *Note that $\mathcal{F}_N = \sigma(T_n)$ for $n = 2^N$, where $T_n : \Omega \to \mathbb{R}^n$ maps a path ω to the vector of its increments along the N-th dyadic partition; cf. the proof of Corollary 12. Identifying the restrictions of Q and P to $\mathcal{F}_N$ with their images ν and $\tilde{\mu}_n$ under T_n, Talagrand's finite-dimensional inequality* (25) *can be written in the form*

$$2^N W_N^2(Q, P) \leq 2 H_N(Q|P), \tag{27}$$

with

$$W_N(Q,P) := \inf \big(\tilde{E}_{\tilde{P}}\big[\big\langle \tilde{Y} - \tilde{X}\big\rangle_N\big]\big)^{1/2},$$

where the infimum is taken over all couplings of Q *and* P *and* $\langle \cdot \rangle_N$ *denotes the discrete quadratic variation along the* N*-th dyadic partition, that is,* $\langle \omega \rangle_N = ||T_n(\omega)||_n^2$ *for any continuous function* $\omega \in \Omega = C_0[0,1]$*. For* $N \uparrow \infty$*, the right hand side of* (27) *increases to* $2H(Q|P)$*. Thus, an alternative version of the bottom-up approach to Talagrand's inequality on Wiener space consists in showing that, in the limit* $N \uparrow \infty$*, the left hand side of* (27) *can be replaced by* $W_{\mathcal{H}}(Q,P)$ *if* $H(Q|P < \infty$*.*

In order to go beyond the absolutely continuous case $Q \ll P$, let us rewrite the finite-dimensional inequality (27) as

$$W_N^2(Q,P) \leq 2 \cdot 2^{-N} H_N(Q|P). \tag{28}$$

Taking the limit $N \uparrow \infty$, the specific relative entropy $h(Q|P)$ appears on the right hand side of (28), while the left hand side suggests to define a new Wasserstein distance on Wiener space in terms of quadratic variation. The resulting extension of Talagrand's inequality is contained in Theorems 32 and 36 below. Instead of analyzing the limit behaviour of the left hand side of (28), we are going to use again a top-down approach, arguing directly in terms of couplings on Wiener space. As a first step in that direction, we now show how the specific relative entropy $h(Q|P)$ reflects the special structure of a semimartingale measure Q on $C_0[0,1]$.

Definition 15. *Let* $\mathcal{Q}_{\mathcal{S}}$ *denote the class of all probability measures* Q *on* $\Omega = C_0[0,1]$ *such that the coordinate process* W *is a special semimartingale of the form*

$$W = M^Q + A^Q \tag{29}$$

under Q *with respect to the filtration* $(\mathcal{F}_t)$*, where*

1. $M^Q = (M^Q)_{0\leq t\leq 1}$ *is a square-integrable martingale under* Q
2. $A^Q = (A^Q)_{0\leq t\leq 1}$ *is an adapted process with continuous paths of bounded variation such that its total variation* $|A|^Q$ *satisfies* $|A|_1^Q \in L^2(Q)$*.*

A probability measure $Q \in \mathcal{Q}_{\mathcal{S}}$ *will be called a* martingale measure *if* $A^Q = 0$*, that is, if* W *is a square-integrable martingale under* Q*. The class of all such martingale measures will be denoted by* $\mathcal{Q}_{\mathcal{M}}$*.*

Remark 16. *Proposition 3 shows that any probability measure* Q *on* $(\Omega, \mathcal{F})$ *with finite relative entropy* $H(Q|P) < \infty$ *belongs to the class* $\mathcal{Q}_{\mathcal{S}}$*, with* $M^Q = W^Q$ *and* $A^Q = B^Q$*.*

Let us now fix a measure $Q \in \mathcal{Q}_{\mathcal{S}}$. We denote by

$$\langle W \rangle \;=\; \big(\langle W \rangle_t\big)_{0\leq t\leq 1}$$

the continuous quadratic variation process defined, Q-a.s., by the decomposition

$$W^2 = \int W dW + \langle W \rangle$$

of the continuous semimartingale W^2 under Q. Our assumptions for $Q \in \mathcal{Q}_{\mathcal{S}}$ imply that

$$\langle W \rangle_t = \lim_{N \uparrow \infty} \sum_{t \in D_N} \big(W_t - W_{t-2^{-N}} \big)^2 \qquad \text{in } L^1(Q) \tag{30}$$

and that

$$\lim_{N \uparrow \infty} \sum_{t \in D_N} \big(A_t - A_{t-2^{-N}} \big)^2 = 0 \qquad \text{in } L^1(Q) \tag{31}$$

cf., e.g., Ch. VI in [12].

Let us introduce the finite measure $q(\omega, dt)$ on $[0,1]$ with distribution function $\langle W \rangle(\omega)$, defined Q-a.s., and denote by

$$q(\omega, dt) = q_s(\omega, dt) + \sigma^2(\omega, t) dt \tag{32}$$

its Lebesgue decomposition into a singular and an absolutely continuous part with respect to Lebesgue measure λ on $[0,1]$; an explicit construction will be given in the second part of the following proof.

Our next aim is to derive, for a large class of probability measures $Q \in \mathcal{Q}_{\mathcal{S}}$, a lower bound for the specific relative entropy $h(Q|P)$ in terms of the quadratic variation of W under Q, that is, in terms of the random measure $q(\cdot, \cdot)$. In a first step we focus on the case $Q \in \mathcal{Q}_{\mathcal{M}}$. The following theorem for martingale measures is essentially due to N. Gantert in [7]; here we extend it to the case where the quadration variation may have a singular component.

Theorem 17. *For any martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$, the specific relative entropy of Q with respect to Wiener measure P satisfies*

$$\begin{aligned} h(Q|P) &\geq \frac{1}{2} E_Q \big[q(\omega, [0,1]) - 1 + H(\lambda | q(\omega, \cdot)) \big] \\ &= \frac{1}{2} E_Q \big[q_s(\omega, [0,1]) \big] + E_Q \Big[\int_0^1 f\big(\sigma^2(\omega, t)\big) dt \Big], \end{aligned} \tag{33}$$

where f is the convex function on $[0, \infty)$ defined by $f(x) = \frac{1}{2}(x - 1 - \log x) \geq 0$. In particular,

$$h(Q|P) < \infty \quad \Longrightarrow \quad \sigma^2(\cdot, \cdot) > 0 \quad Q \otimes \lambda - a.s. \tag{34}$$

Proof. 1) First we look at the general case $Q \in \mathcal{Q}_{\mathcal{S}}$. Thus we can write $W = M + A$, where M is a square-integrable Q-martingale and A is an adapted process with continuous paths of bounded variation such that $E_Q \big[|A|_1^2 \big] < \infty$.

For $N \geq 1$ and $i = 1, \ldots, 2^N$ we write $t_i = i2^{-N}$ and denote by $\nu_{N,i}(\omega, \cdot)$ the conditional distribution of the increment $W_{t_i} - W_{t_{i-1}}$ under Q given the σ-field $\mathcal{F}_{N,t_{i-1}}$, by

$$m_{N,i} = E_Q\big[W_{t_i} - W_{t_{i-1}} | \mathcal{F}_{N,t_{i-1}}\big] = E_Q\big[A_{t_i} - A_{t_{i-1}} | \mathcal{F}_{N,t_{i-1}}\big]$$

its conditional mean, by

$$\tilde{\sigma}^2_{N,i} = E_Q\big[(W_{t_i} - W_{t_{i-1}})^2 | \mathcal{F}_{N,t_{i-1}}\big] - m^2_{N,i}$$

its conditional variance, and by

$$\sigma^2_{N,i} = E_Q\big[(M_{t_i} - M_{t_{i-1}})^2 | \mathcal{F}_{N,t_{i-1}}\big] = E_Q\big[\langle W \rangle_{t_i} - \langle W \rangle_{t_{i-1}} | \mathcal{F}_{N,t_{i-1}}\big] \tag{35}$$

the conditional variance of the martingale increment $M_{t_i} - M_{t_{i-1}}$. We can write

$$H_N(Q|P) = \sum_{i=1}^{2^N} E_Q\big[H\big(\nu_{N,i}(\omega, \cdot) | N(0, 2^{-N})\big)\big].$$

Since

$$H\big(N(m, \alpha) | N(0, \beta)\big) = f(\frac{\alpha}{\beta}) + \frac{m^2}{2\beta}$$

for $\alpha, \beta > 0$ and $m \in R^1$, we get

$$\begin{aligned} &H\big(\nu_{N,i} | N(0, 2^{-N})\big) \\ &\quad = H\big(\nu_{N,i} | N(m_{N,i}, \tilde{\sigma}^2_{N,i})\big) + H\big(N(m_{N,i}, \tilde{\sigma}^2_{N,i}) | N(0, 2^{-N})\big) \\ &\quad = H\big(\nu_{N,i} | N(m_{N,i}, \tilde{\sigma}^2_{N,i})\big) + f(2^N \tilde{\sigma}^2_{N,i}) + \frac{1}{2} 2^N m^2_{N,i}\,, \end{aligned}$$

hence

$$H_N(Q|P) = H_N(Q|Q^*_N) + E_Q\big[\sum_{i=1}^{2^N} f(2^N \tilde{\sigma}^2_{N,i})\big] + \frac{1}{2} 2^N I_N, \tag{36}$$

where we define

$$I_N := E_Q\big[\sum_{i=1}^{2^N} m^2_{N,i}\big], \tag{37}$$

and where Q^*_N denotes the probability measure on $(\Omega, \mathcal{F}_N)$ such that the increments $W_{t_i} - W_{t_{i-1}}$ have conditional distribution $N(m_{N,i}, \tilde{\sigma}^2_{N,i})$ given the σ-field $\mathcal{F}_{N,t_{i-1}}$. Note that Jensen's inequality yields

$$I_N \leq E_Q\big[\sum_{i=1}^{2^N} (A_{t_i} - A_{t_{i-1}})^2\big],$$

hence

$$\lim_{N \uparrow \infty} I_N = 0, \tag{38}$$

due to (31). Note also that $H_N(Q|P) < \infty$ implies $\tilde{\sigma}^2_{N,i}(\omega) > 0 \quad Q$-a.s., since $f(0) = \infty$.

2) Let $Q \otimes q$ denote the finite measure on $\bar{\Omega} = \Omega \times [0,1]$ defined by $Q \otimes q(d\omega, dt) = Q(d\omega)q(\omega, dt)$. On the σ-field

$$\mathcal{P}_N := \sigma(\{A_t \times (t,1] \mid t \in D_N, A_t \in \mathcal{F}_{N,t}\}),$$

the measure $Q \otimes q$ is absolutely continuous with respect to the product measure $Q \otimes \lambda$, where λ denotes the Lebesgue measure on $(0,1]$, and the density is given by

$$\sigma^2_N(\omega, t) := \sum_{i=1}^{2^N} 2^N \sigma^2_{N,i}(\omega) I_{(t_{i-1}, t_i]}(t).$$

The σ-fields $\mathcal{P}_N$ increase to the predictable σ-field $\mathcal{P}$ on $\bar{\Omega}$, generated by the sets $A_t \times (t,1]$ with $t \in [0,1]$ and $A_t \in \mathcal{F}_t$. Applying the first part of Lemma 19 with $\mu = Q \otimes \lambda$ and $\nu = Q \otimes q$, we see that the limit

$$\sigma^2(\omega, t) = \lim_{N \uparrow \infty} \sigma^2_N(\omega, t)$$

exists both $Q \otimes q$ -a.s. and $Q \otimes \lambda$ -a.s., with

$$\sigma^2(\omega, t) \in [0, \infty) \qquad Q \otimes \lambda - a.s.$$

and

$$\sigma^2(\omega, t) \in (0, \infty] \qquad Q \otimes q - a.s..$$

Moreover, the Lebesgue decomposition of $Q \otimes q$ with respect to $Q \otimes \lambda$ on the predictable σ-field $\mathcal{P}$ takes the form

$$Q \otimes q\,[\bar{A}] = Q \otimes q\,[\bar{A} \cap \{\sigma^2 = \infty\}] + E_{Q \otimes \lambda}\big[\sigma^2; \bar{A}\big],$$

for $\bar{A} \in \mathcal{P}$. This implies, Q-a.s., the Lebesgue decomposition

$$q(\omega, dt) = q_s(\omega, dt) + \sigma^2(\omega, t)\lambda(dt),$$

of $q(\omega, \cdot)$ with respect to Lebesgue measure λ, where the singular part $q_s(\omega, \cdot)$ is given by the restriction of $q(\omega, \cdot)$ to the λ-null set

$$N(\omega) := \{t \mid \sigma^2(\omega, t) = \infty\}. \tag{39}$$

3) Let us now focus on the case where Q is a martingale measure. For $Q \in \mathcal{Q}_{\mathcal{M}}$, we have $\tilde{\sigma}^2_{N,i} = \sigma^2_{N,i}$ and $A = 0$, hence $I_N = 0$. Thus, equation (36) can be written as

$$2^{-N} H_N(Q|P) = 2^{-N} H_N(Q|Q^*_N) + E_Q\Big[\int_0^1 f(\sigma^2_N(\cdot, t))dt\Big]. \tag{40}$$

Since $H_N(Q|Q_N^*) \geq 0$, we obtain

$$\begin{aligned} h(Q|P) &\geq \lim_{N\uparrow\infty} E_Q\big[\int_0^1 f(\sigma_N^2(\cdot,t))dt\big] \\ &= \frac{1}{2}E_Q\big[q_s(\omega,(0,1])\big] + E_Q\big[\int_0^1 f(\sigma^2(\cdot,t))dt\big]. \end{aligned} \tag{41}$$

where we apply the second part of Lemma 19 below, with $\mu = Q \otimes \lambda$ and $\nu = Q \otimes q$. Since $f(0) = \infty$, we see that $h(Q|P) < \infty$ implies that $\sigma^2(\cdot,\cdot)$ is strictly positive $Q \otimes \lambda$-a.s.. □

Remark 18. *The proof of Theorem 17 shows that we obtain* existence of the limit

$$h(Q|P) = \lim_{N\uparrow\infty} 2^{-N} H_N(Q|P) \tag{42}$$

together with the equality

$$h(Q|P) = \frac{1}{2}E_Q\big[q_s(\omega,[0,1])\big] + E_Q\big[\int_0^1 f\big(\sigma^2(\omega,t)\big)dt\big], \tag{43}$$

if and only if Q is "almost locally Gaussian" in the sense that the measures Q_N^* appearing in (36) *satisfy*

$$\lim_{N\uparrow\infty} 2^{-N} H_N(Q|Q_N^*) = 0. \tag{44}$$

This was already observed by N. Gantert in [7].

In the proof of Theorem 17 we have used the following general lemma.

Lemma 19. *Consider two probability measures μ and ν on a measurable space $(S,\mathcal{S})$ and a sequence of $(\mathcal{S}_n)_{n=1,2,...}$ of sub-σ-fields increasing to $\mathcal{S}_\infty$. Suppose that ν is equivalent to μ on $\mathcal{S}_n$ with density ϕ_n.*
1) The limit $\phi_\infty = \lim_n \phi_n$ exists both μ-a.s. and ν-a,s,, with

$$\phi_\infty \in [0,\infty)\ \mu - a.s. \quad \text{and} \quad \phi_\infty \in (0,\infty]\ \nu - a.s.,$$

and the Lebesgue decomposition $\nu = \nu_s + \nu_a$ of ν with respect to μ on $\mathcal{S}_\infty$ is given by

$$\nu_s(A) = \nu(A \cap \{\phi_\infty = \infty\}) \quad \text{and} \quad \nu_a(A) = \int_A \phi_\infty d\mu.$$

2) If $\sup_n \int f(\phi_n)d\mu < \infty$ for $f(x) = \frac{1}{2}(x - 1 - \log x)$ then we have

$$\lim_{n\uparrow\infty} \int f(\phi_n)d\mu = \frac{1}{2}\nu_s(S) + \int f(\phi_\infty)d\mu. \tag{45}$$

Proof. The first part is well-known; the proof uses standard martingale arguments. To prove the second part, we write

$$\begin{aligned}\int 2f(\phi_n)d\mu &= \int \phi_n d\mu - 1 + \int \log(\phi_n^{-1})d\mu \\ &= \nu_s(S) + \int \phi_\infty d\mu - 1 + H_{\mathcal{S}_n}(\nu|\mu).\end{aligned}$$

Due to (6), we get

$$\lim_{n\uparrow\infty} \int f(\phi_n)d\mu = \frac{1}{2}\big(\nu_s(S) + \int \phi_\infty d\mu - 1 + H_{\mathcal{S}_\infty}(\nu|\mu)\big)$$

If the left hand side is finite, the relative entropy is finite and reduces to $\int \log(\phi_\infty^{-1})d\mu$, and this yields equation (45). □

Let us now go beyond the case of a martingale measure. Take $Q \in \mathcal{Q}_{\mathcal{S}}$ and let $W = M + A$ be the canonical decomposition of the semimartingale W under Q. As soon as the process A is non-deterministic, the conditional variances $\sigma_{N,i}^2$ of M defined in (35) do no longer coincide with the conditional variances $\tilde{\sigma}_{N,i}^2$ of W along the N-th dyadic partition. Instead we have

$$\tilde{\sigma}_{N,i}^2 = \sigma_{N,i}^2 + \delta_{N,i},$$

where

$$\delta_{N,i} = \alpha_{N,i}^2 + 2E_Q\big[(M_{t_i} - M_{t_{i-1}})(A_{t_i} - A_{t_{i-1}})|\mathcal{F}_{N,t_{i-1}}\big],$$

and where we denote by

$$\alpha_{N,i}^2 = E_Q\big[(A_{t_i} - A_{t_{i-1}})^2|\mathcal{F}_{N,t_{i-1}}\big] - m_{N,i}^2$$

the conditional variances of A along the N-th dyadic partition.

Lemma 20. *The differences $\delta_{N,i}$ and the conditional variances $\alpha_{N,i}^2$ satisfy*

$$\lim_{n\uparrow\infty} E_Q\big[\sum_{i=1}^{2^N} |\delta_{N,i}|\big] = \lim_{n\uparrow\infty} E_Q\big[\sum_{i=1}^{2^N} \alpha_{N,i}^2\big] = 0.$$

Proof. Since

$$J_N := E_Q\big[\sum_{i=1}^{2^N} \alpha_{N,i}^2\big] \le E_Q\big[\sum_{i=1}^{2^N} (A_{t_i} - A_{t_{i-1}})^2\big],$$

we obtain

$$\lim_{n\uparrow\infty} J_N = 0 \tag{46}$$

due to (31). On the other hand, since

$$|\delta_{N,i}| \le \alpha_{N,i}^2 + 2\sigma_{N,i}\alpha_{N,i}, \tag{47}$$

we get

$$
\begin{aligned}
E_Q\big[\sum_{i=1}^{2^N}|\delta_{N,i}|\big] &\le E_Q\big[\sum_{i=1}^{2^N}\alpha_{N,i}^2\big]+2\sum_{i=1}^{2^N}E_Q\big[\sigma_{N,i}^2\big]^{1/2}E_Q\big[\alpha_{N,i}^2\big]^{1/2}\\
&\le J_N+2\,E_Q\big[M_1^2\big]^{1/2}J_N^{1/2},
\end{aligned}
$$

hence

$$
\lim_{N\uparrow\infty}E_Q\big[\sum_{i=1}^{2^N}|\delta_{N,i}|=0, \tag{48}
$$

due to (46). □

To prove our extended version of Theorem 17, we use an additional assumption.

Definition 21. *We denote by $\mathcal{Q}_{\mathcal{S}}^0$ the class of all probability measures $Q\in\mathcal{Q}_{\mathcal{S}}$ such that*

$$
\lim_{n\uparrow\infty}E_Q\big[2^{-N}\sum_{i=1}^{2^N}\alpha_{N,i}^2\sigma_{N,i}^{-2}\big]=0. \tag{49}
$$

Remark 22. *Condition (49) is satisfied if $\sigma^2(\cdot,\cdot)$ is bounded away from 0. Indeed, if $\sigma^2(\cdot,\cdot)\ge c\quad Q\otimes\lambda$-a.s. for some $c>0$ then*

$$
\sum_{i=1}^{2^N}2^N\sigma_{N,i}^2(\omega)I_{(t_{i-1},t_i]}(t)=\sigma_N^2(\omega,t)\ge E_{Q\otimes\lambda}\big[\sigma^2|\mathcal{P}_N\big]\ge c\quad Q\otimes\lambda-a.s.;
$$

cf. the second part of the proof of Theorem 17. Thus, (49) follows from Lemma 20.

Theorem 23. *For any $Q\in\mathcal{Q}_{\mathcal{S}}^0$,*

$$
h(Q|P)\ge\frac{1}{2}E_Q\big[q_s(\omega,[0,1])\big]+E_Q\big[\int_0^1 f\big(\sigma^2(\omega,t)\big)dt\big]. \tag{50}
$$

Proof. 1) Let us return to the first part of the proof of Theorem 17. Since $H_N(Q|Q_N^*)\ge 0$, equation (36) yields

$$
2^{-N}H_N(Q|P)\ge E_Q\big[\sum_{i=1}^{2^N}f(2^N\tilde\sigma_{N,i}^2)2^{-N}\big]+\frac{1}{2}I_N.
$$

Since f is convex with $f'(x)=\frac{1}{2}(1-x^{-1})$, we obtain

$$
f(2^N\tilde\sigma_{N,i}^2)\ge f(2^N\sigma_{N,i}^2)+\frac{1}{2}(1-2^{-N}\sigma_{N,i}^{-2})2^N\delta_{N,i}.
$$

Due to (38), this implies

$$h(Q|P) \geq \liminf_{N\uparrow\infty} E_Q\big[\int_0^1 f\big(\sigma_N^2(\omega,t)\big)dt + \frac{1}{2}\Delta_N\big],$$

where

$$\Delta_N = \sum_{i=1}^{2^N}\big(\delta_{N,i} - 2^{-N}\sigma_{N,i}^{-2}\,\delta_{N,i}\big).$$

Applying the second part of Lemma 19 as in the proof of Theorem 17, we see that inequality (50) holds as soon as we show that

$$\lim_{N\uparrow\infty}\Delta_N = 0 \quad \text{in} \quad L^1(Q). \tag{51}$$

2) In view of Lemma 20 it is enough to show convergence to 0 for

$$\begin{aligned} E_Q\big[&\sum_{i=1}^{2^N} 2^{-N}\sigma_{N,i}^{-2}|\delta_{N,i}|\big] \\ &\leq E_Q\big[\sum_{i=1}^{2^N} 2^{-N}\sigma_{N,i}^{-2}\alpha_{N,i}^2\big] + 2E_Q\big[2^{-N}\sum_{i=1}^{2^N}\alpha_{N,i}\sigma_{N,i}^{-1}\big] \\ &\leq E_Q\big[\sum_{i=1}^{2^N} 2^{-N}\sigma_{N,i}^{-2}\alpha_{N,i}^2\big] + 2E_Q\big[2^{-N}\sum_{i=1}^{2^N}\sigma_{N,i}^{-2}\alpha_{N,i}^2\big]^{1/2}. \end{aligned}$$

But the last two terms converge to 0 due to our assumption (49), and this completes the proof of (51). □

Corollary 24. *Let $Q \in \mathcal{Q}_{\mathcal{S}}$ be such that $||A^Q||_{\mathcal{H}} \in L^2(Q)$. Then we have*

$$h(Q|P) = 0 \quad \Longleftrightarrow \quad H(Q|P) < \infty,$$

and in this case the canonical decomposition (29) *of W under Q takes the form $M^Q = W^Q$ and $A^Q = B^Q$.*

Proof. Let us assume $h(Q|P) = 0$. Inequality (33) implies $q_s(\omega,\cdot) = 0$ Q-a.s and $f\big(\sigma^2(\omega,t)\big) = 0$ $Q\otimes\lambda$-a.s, hence $\sigma^2(\omega,t) = 1$ $Q\otimes\lambda$-a.s. Thus, W has quadratic variation

$$\langle W\rangle_t = \langle M^Q\rangle_t = t$$

under Q, and so M^Q is a Wiener process under Q. Uniqueness of the canonical decomposition of W under Q yields $M^Q = W^Q$ and $A^Q = B^Q$, hence

$$H(Q|P) = \frac{1}{2}E_Q\big[||A^Q||_{\mathcal{H}}^2\big] < \infty$$

due to Proposition 3. Conversely, $H(Q|P) < \infty$ implies $h(Q|P) = 0$, as we have already observed above, following the definition of $h(Q|P)$. □

5 Intrinsic Wiener Process and Optimal Coupling for Semimartingale Measures

We fix a probability measure $Q \in \mathcal{Q}_{\mathcal{S}}$ and denote by

$$W = M + A \tag{52}$$

the canonical decomposition of the coordinate process W under Q. Recall the Lebesgue decomposition

$$q(\omega, dt) = q(\omega, dt) + \sigma^2(\omega, t)dt$$

of the random measure $q(\omega, \cdot)$ on $[0, 1]$ with distribution function $\langle W \rangle(\omega)$, and put

$$A(\omega) := \{t \in [0, 1] \mid \sigma^2(\omega, t) < \infty\}.$$

The following construction of an intrinsic Wiener process W^Q for Q extends the definition in Proposition 3 beyond the absolutely continuous case $Q \ll P$.

Lemma 25. *If $h(Q|P) < \infty$ then the process $W^Q = (W_t^Q)_{0 \le t \le 1}$, defined Q-a.s. by*

$$W_t^Q := \int_0^t \sigma(\cdot, s)^{-1} I_{A(\cdot)}(s) dM_s, \tag{53}$$

is a Wiener process under Q.

Proof. By Theorem 17, our assumption $h(Q|P) < \infty$ implies

$$E_Q\Big[\int_0^1 f(\sigma^2(\omega, t))dt\Big] < \infty,$$

where $f(x) = \frac{1}{2}(x - 1 - \log x)$, and in particular

$$0 < \sigma^2(\cdot, \cdot) < \infty \quad Q \otimes \lambda - a.s..$$

since $f(0) = \infty$. Since $\langle M \rangle = \langle W \rangle$ and $\lambda\big(A(\cdot)\big) = 1$ Q-a.s., the predictable integrand $\phi_s = \sigma(\cdot, s)^{-1} I_{A(\cdot)}(s)$ in (53) satisfies

$$\int_0^t \phi_s^2 d\langle M \rangle_s = \int_0^t \sigma_s^{-2} I_{A(\cdot)}(s) \sigma_s^2 ds = \int_0^t I_{A(\cdot)}(s) ds = t.$$

Thus, the stochastic integrals in (53) are well defined, and they define a continuous martingale W^Q under Q with quadratic variation $\langle W^Q \rangle_t = t$. This implies that W^Q is a Wiener process under Q. □

For the rest of this section we assume that $Q \in \mathcal{Q}_{\mathcal{S}}$ satisfies the condition

$$h(Q|P) < \infty, \tag{54}$$

and so W^Q will be a Wiener process under Q.

Definition 26. *W^Q will be called the* intrinsic Wiener process *of Q.*

Remark 27. *If $H(Q|P) < \infty$ then the intrinsic Wiener process coincides with the Wiener process $W^Q := W - B^Q$ defined in Proposition 3; cf. the proof of corollary 24.*

Definition 28. *An adaptive coupling $(\tilde{X}, \tilde{Y})$ of P and Q on a filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0 \le t \le 1}, \tilde{P})$ will be called a* semimartingale coupling *if $\tilde{Y}$ is a special semimartingale with respect to $\tilde{P}$ and $(\tilde{\mathcal{F}}_t)_{0 \le t \le 1}$, and if the canonical decomposition $\tilde{Y} = \tilde{M} + \tilde{A}$ is such that*

1. *$\tilde{M}$ is a square-integrable martingale,*
2. *$\tilde{A}$ is an adapted process with continuous paths of bounded variation such that its total variation $|\tilde{A}|$ satisfies $|\tilde{A}|_1 \in L^2(\tilde{P})$.*

Clearly, the pair (W^Q, W) is a semimartingale coupling of P and Q, defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le 1}, Q)$. In fact, we are going to show that (W^Q, W) is the *optimal* semimartingale coupling for the Wasserstein distance $W_{\mathcal{S}}(Q, P)$ defined below.

Proposition 29. *For any semimartingale coupling $(\tilde{X}, \tilde{Y})$ of P and Q on some filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0 \le t \le 1}, \tilde{P})$ we have*

$$\tilde{E}\big[\big\langle \tilde{Y} - \tilde{X} \big\rangle_1\big] \ge E_Q\big[\big\langle W - W^Q \big\rangle_1\big], \tag{55}$$

and equality holds if and only if $\tilde{X} = W^Q(\tilde{Y})$ $\quad \tilde{P}$-a.s.. Moreover,

$$E_Q\big[\big\langle W - W^Q \big\rangle_1\big] = E_Q\Big[\int_0^1 \big(\sigma(\cdot, s) - 1\big)^2 ds + q_s(\cdot, (0,1])\Big]. \tag{56}$$

.

Proof. 1) First we show that the last equality holds. Recall from the proof of Theorem 17 that $q_s(\omega, \cdot)$ is given, Q-a.s., by the restriction of $q(\omega, \cdot)$ to the λ-nullset $N(\omega)$ defined in (39). Since $A(\cdot) \cup N(\cdot) = [0,1]$, we have

$$\begin{aligned} W_t &= \int_0^t I_{A(\cdot)}(s) dW_s + \int_0^1 I_{N(\cdot)}(s) dW_s \\ &= \int_0^t \sigma(\cdot, s) dW_s^Q + \int_0^1 I_{N(\cdot)}(s) dW_s, \end{aligned}$$

hence

$$(W - W^Q)_t = \int_0^t \big(\sigma(\cdot, s) - 1\big) dW_s^Q + \int_0^1 I_{N(\cdot)}(s) dW_s$$

and

$$\begin{aligned} \big\langle W - W^Q \big\rangle_t &= \int_0^t \big(\sigma(\cdot, s) - 1\big)^2 ds + \int_0^1 I_{N(\cdot)}(s) d\big\langle W \big\rangle_s \\ &+ 2 \int_0^t \big(\sigma(\cdot, s) - 1\big) I_{N(\cdot)}(s) d\big\langle W^Q, W \big\rangle_s. \end{aligned}$$

The last term vanishes since, Q-a.s., $N(\omega)$ is a nullset with respect to $d\langle W^Q, W\rangle(\omega) \ll d\langle W^Q\rangle(\omega) = dt$. This implies

$$E_Q\big[\langle W - W^Q\rangle_1\big] = E_Q\big[\int_0^1 \big(\sigma(\cdot, s) - 1\big)^2 ds + q_s(\cdot, (0,1])\big].$$

2) Consider any semimartingale coupling $(\tilde{X}, \tilde{Y})$ of P and Q, defined on some filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0\le t\le 1}, \tilde{P})$. Both $\tilde{X}$ and the process $\tilde{W} := W^Q(\tilde{Y})$, defined by

$$\tilde{W}_t := \int_0^t \sigma(\tilde{Y}, s)^{-1} I_{A(\tilde{Y})}(s) d\tilde{Y}_s\,,$$

are Wiener processes under $\tilde{P}$ with respect to the filtration $(\tilde{\mathcal{F}}_t)$. Projecting the first on the second, we can write

$$\tilde{X}_t = \int_0^t \rho_s d\tilde{W}_s + \tilde{L}_t\,,$$

where $\tilde{L} = (\tilde{L}_t)_{0\le t\le 1}$ is a martingale orthogonal to $\tilde{W}$. Since

$$t = \langle \tilde{X}\rangle_t = \int_0^t \rho_s^2 ds + \langle \tilde{L}\rangle_t,$$

we get $\rho_t^2 \le 1$ and $d\langle \tilde{L}\rangle_t = (1 - \rho_t^2)dt$. This implies

$$\begin{aligned} d\langle \tilde{X}, \tilde{Y}\rangle &= \rho_t d\langle \tilde{W}, \tilde{Y}\rangle \\ &= \rho_t \sigma^{-1}(\tilde{Y}, t) I_{A(\tilde{Y})}(t) \sigma^2(\tilde{Y}, t) dt \\ &\le \sigma(\tilde{Y}, t) dt, \end{aligned}$$

hence

$$\begin{aligned} \langle \tilde{Y} - \tilde{X}\rangle_1 &= \langle \tilde{Y}\rangle_1 + \langle \tilde{X}\rangle_1 - 2\langle \tilde{X}, \tilde{Y}\rangle_1 \\ &\ge \int_0^1 \sigma^2(\tilde{Y}, t) dt + q_s(\tilde{Y}, (0,1]) + 1 - 2\int_0^1 \sigma(\tilde{Y}, t) dt \\ &= \int_0^1 \big(\sigma(\tilde{Y}, t) - 1\big)^2 dt + q_s(\tilde{Y}, (0,1]). \end{aligned}$$

Thus,

$$\begin{aligned} \tilde{E}\big[\langle \tilde{Y} - \tilde{X}\rangle_1\big] &\ge \tilde{E}\big[\int_0^1 (\sigma\big(\tilde{Y}, t) - 1\big)^2 dt + q_s(\tilde{Y}, (0,1])\big] \\ &= E_Q\big[\int_0^1 \big(\sigma(\cdot, t) - 1\big)^2 dt + q_s(\cdot, (0,1])\big] \\ . &= E_Q\big[\langle W - W^Q\rangle_1\big], \end{aligned}$$

and equality holds iff $\rho_t(\cdot) = 1$ $\tilde{P} \otimes dt$ -a.s., that is, iff $\tilde{X} = \tilde{W} = W^Q(\tilde{Y})$ $\tilde{P}$-a.s.. □

Now consider the following Wasserstein distance $W_{\mathcal{S}}(Q,P)$, where the cost function is defined in terms of quadratic variation.

Definition 30. *The* Wasserstein distance $W_{\mathcal{S}}(Q,P)$ *between* Q *and Wiener measure* P *is defined as*

$$W_{\mathcal{S}}(Q,P) = \inf \big(\tilde{E}\big[\langle \tilde{Y} - \tilde{X}\rangle_1 + ||\tilde{A}||^2_{\mathcal{S}}\big]\big)^{\frac{1}{2}}, \tag{57}$$

where the infimum is taken over all semimartingale couplings $(\tilde{Y},\tilde{X})$ *of* Q *and* P *on some filtered probability space, where* $\tilde{M}+\tilde{A}$ *is the canonical decomposition of* $\tilde{Y}$*, and where we set*

$$||\tilde{A}||_{\mathcal{S}} = \Big(\int_0^1 \tilde{a}_t^2 d\langle \tilde{Y}\rangle_t\Big)^{1/2}$$

if $\tilde{A}$ *is absolutely continuous with respect to* $\langle \tilde{Y}\rangle$ *with density process* $\tilde{a}$*, and* $||\tilde{A}||_{\mathcal{S}} = \infty$ *otherwise.*

Remark 31. *In the absolutely continuous case* $Q \ll P$ *we have*

$$d\langle \tilde{Y}\rangle = d\langle \tilde{X}\rangle = dt \qquad Q\text{−a.s.},$$

and so the norm $||\tilde{A}||_{\mathcal{S}}$ *reduces to the Cameron-Martin norm* $||\tilde{A}||_{\mathcal{H}}$*.*

As an immediate corollary to the preceding proposition we obtain the following inequality for martingale measures. It provides a first extension of Talagrand's inequality (13) on Wiener space beyond the absolutely continuous case.

Theorem 32. *For a martingale measure* $Q \in \mathcal{Q}_{\mathcal{M}}$*,*

$$W^2_{\mathcal{S}}(Q,P) = E_Q\big[\langle W - W^Q\rangle_1\big] \leq 2h(Q|P), \tag{58}$$

and equality holds iff $Q = P$*.*

Proof. 1) For $Q \in \mathcal{Q}_{\mathcal{M}}$, the pair (W, W^Q) is a semimartingale coupling of Q and P, defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$, such that $W - W^Q = M - W^Q$ is a martingale under Q. Thus, the expected cost in (57) only involves the quadratic variation component, and Proposition 29 implies

$$W^2_{\mathcal{S}}(Q,P) = E_Q\big[\langle W - W^Q\rangle_1\big] = E_Q\Big[\int_0^1 \big(\sigma(\cdot,s) - 1\big)^2 ds + q_s(\cdot,(0,1])\Big]. \tag{59}$$

Note that

$$(\sigma - 1)^2 \leq \sigma^2 - 1 - \log \sigma^2 = 2f(\sigma^2),$$

with equality iff $\sigma^2 = 1$. Thus,

$$\begin{aligned} E_Q\big[\langle W - W^Q\rangle_1\big] &\leq E_Q\Big[2\int_0^1 f\big(\sigma^2(\cdot,s)\big)dt + q_s(\cdot,(0,1])\Big] \\ &\leq 2\,h(Q|P), \end{aligned} \tag{60}$$

where the second inequality follows from Theorem 17.

2) Equality in (58) implies equality in (60). It follows from part 1) that $\sigma^2(\cdot,\cdot) = 1 \ \ Q\otimes\lambda$-a.s.. This implies $W = M = W^Q$ under Q, hence $Q = P$. The converse is obvious. $\square$

Definition 33. *We write $Q \in \mathcal{Q}_{\mathcal{S}}^*$ if the canonical decomposition $W = M + A$ of the coordinate process W under $Q \in \mathcal{Q}_{\mathcal{S}}$ is such that*

$$E_Q\big[||A||_{\mathcal{S}}^2\big] < \infty, \tag{61}$$

that is, $dA_t = a_t d\langle W\rangle_t$ with $\int_0^1 a_t^2 d\langle W\rangle_t \in L^1(Q)$, and if

$$G^* := \exp\big(-\int_0^1 a_t dM - \frac{1}{2}\int_0^1 a_t^2 d\langle M\rangle_t\big)$$

satisfies

$$G^* \in L^2(Q) \quad \text{and} \quad E_Q[G^*] = 1. \tag{62}$$

Remark 34. *For $Q \in \mathcal{Q}_{\mathcal{S}}^*$, the probability measure Q^* defined by*

$$dQ^* = G^* dQ \tag{63}$$

is an equivalent martingale measure *for Q; cf., for example, [6]. Note that $\mathcal{Q}_{\mathcal{M}} \subset \mathcal{Q}_{\mathcal{S}}^*$, and that $Q^* = Q$ for $Q \in \mathcal{Q}_{\mathcal{M}}$.*

Proposition 35. *For $Q \in \mathcal{Q}_{\mathcal{S}}^*$, the coupling (W, W^Q) of Q and P is optimal for $W_{\mathcal{S}}$, that is,*

$$W_{\mathcal{S}}^2(Q,P) = E_Q\big[\langle W - W^Q\rangle_1 + ||A||_{\mathcal{S}}^2\big]. \tag{64}$$

Proof. For $Q \in \mathcal{Q}_{\mathcal{S}}^*$, the right-hand side in (64) is finite, and so we have $W_{\mathcal{S}}(Q,P) < \infty$. Now take any semimartingale coupling $(\tilde{Y}, \tilde{X})$ of Q and P, defined on some filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0\le t\le 1}, \tilde{P})$, such that

$$\tilde{E}\big[\langle \tilde{Y} - \tilde{X}\rangle_1 + ||\tilde{A}||_{\mathcal{S}}^2\big] < \infty.$$

Since

$$\tilde{E}\big[\langle \tilde{Y} - \tilde{X}\rangle_1\big] \geq E_Q\big[\langle W - W^Q\rangle_1\big] \tag{65}$$

by Proposition 29, it only remains to show that

$$\tilde{E}\big[||\tilde{A}||_{\mathcal{S}}^2\big] \geq E_Q\big[||A||_{\mathcal{S}}^2\big],$$

that is,

$$\tilde{E}\big[\int_0^1 \tilde{a}_t^2 d\langle \tilde{Y}\rangle_t\big] \geq E_Q\big[\int_0^1 a_t^2 d\langle W\rangle_t\big]. \tag{66}$$

We denote by $\tilde{\mathcal{P}}$ the predictable σ-field on $\tilde{\Omega}\times(0,1]$ corresponding to the filtration $(\tilde{\mathcal{F}}_t)$, and by $\mathcal{P}^0 \subseteq \mathcal{P}$ the predictable σ-field corresponding to the smaller filtration $(\tilde{\mathcal{F}}_t^0)$ generated by $(\tilde{Y}_t)$. Since $\tilde{E}\big[||\tilde{A}||_{\mathcal{S}}^2\big] < \infty$, we have

$$d\tilde{A}_t = \tilde{a}_t d\langle \tilde{Y}\rangle_t = \tilde{a}_t dq(\tilde{Y}, t),$$

where $\tilde{a} = (\tilde{a}_t)$ is $\mathcal{P}$-measurable and square-integrable with respect to the finite measure $\tilde{P} \otimes q(\tilde{Y}, \cdot)$ on $\tilde{\mathcal{P}}$. Let $\tilde{a}^0 = (\tilde{a}^0_t)$ denote the process defined by the conditional expectation

$$\tilde{a}^0 := E_{\tilde{P}\otimes q(\tilde{Y},\cdot)}\big[\tilde{a} \,\big|\, \mathcal{P}^0\big],$$

and note that Jensen's inequality implies

$$E_{\tilde{P}\otimes q(\tilde{Y},\cdot)}\big[(\tilde{a}^0)^2\big] \leq E_{\tilde{P}\otimes q(\tilde{Y},\cdot)}\big[\tilde{a}^2\big]. \tag{67}$$

For any $A^0_t \in \mathcal{F}^0_t$ we can write

$$\begin{aligned}
\tilde{E}\big[\tilde{Y}_{t+h} - \tilde{Y}_t; A^0_t\big] &= \tilde{E}\big[\tilde{M}_{t+h} - \tilde{M}_t; A^0_t\big] + \tilde{E}\big[\tilde{A}_{t+h} - \tilde{A}_t; A^0_t\big] \\
&= \tilde{E}\Big[\int_t^{t+h} \tilde{a}_s d\big\langle \tilde{Y}\big\rangle_s; A^0_t\Big] = E_{\tilde{P}\otimes q(\tilde{Y},\cdot)}\big[\tilde{a}; A^0_t \times (t, t+h]\big] \\
&= E_{\tilde{P}\otimes q(\tilde{Y},\cdot)}\big[\tilde{a}^0; A^0_t \times (t, t+h]\big] = \tilde{E}\Big[\int_t^{t+h} \tilde{a}^0_s d\big\langle \tilde{Y}\big\rangle_s; A^0_t\Big].
\end{aligned}$$

This implies that the canonical decomposition of the semimartingale $\tilde{Y}$ in the smaller filtration $(\tilde{\mathcal{F}}^0_t)$ is of the form

$$\tilde{Y}_t = \tilde{M}^0_t + \int_0^t \tilde{a}^0_s d\big\langle \tilde{Y}\big\rangle_s.$$

where $\tilde{M}^0$ is a martingale with respect to $(\tilde{\mathcal{F}}^0_t)$. On the other hand, since the law of $\tilde{Y}$ under $\tilde{P}$ is given by Q, we have

$$\tilde{Y}_t = M_t(\tilde{Y}) + \int_0^t a_s(\tilde{Y}) d\big\langle \tilde{Y}\big\rangle_s.$$

Uniqueness of the canonical decomposition implies

$$\tilde{a}^0 = a(\tilde{Y}) \qquad \tilde{P} \otimes q(\tilde{Y}, \cdot) - a.s. \tag{68}$$

Thus, inequality (67) yields

$$\tilde{E}\Big[\int_0^1 \tilde{a}_t^2 d\big\langle \tilde{Y}\big\rangle_t\Big] \geq \tilde{E}\Big[\int_0^1 a_t^2(\tilde{Y}) d\big\langle \tilde{Y}\big\rangle_t\Big] = E_Q\Big[\int_0^1 a_t^2(W) d\big\langle W\big\rangle_t\Big],$$

and so we have shown inequality (66). □

The following inequality extends Theorem 32 beyond the case of a martingale measure. As explained in Remark 37 below, it contains inequality (58) for $Q \in \mathcal{Q}_{\mathcal{M}}$, Talagrand's inequality (9) for $Q \ll P$, and Corollary 9 for $W_{\mathcal{H},ad}$ as special cases.

Theorem 36. *For $Q \in \mathcal{Q}^*_{\mathcal{S}}$,*

$$W^2_{\mathcal{S}}(Q,P) \le 2\big(h(Q|P) + H(Q|Q^*)\big), \tag{69}$$

where Q^* is the equivalent martingale measure for Q defined by (63)*. Equality holds iff $H(Q|P) < \infty$.*

Proof. 1) Proposition 35 combined with inequality (60) shows that

$$\begin{aligned} W^2_{\mathcal{S}}(Q,P) &= E_Q\big[\langle W - W^Q\rangle_1 + ||A||^2_{\mathcal{S}}\big] \\ &\le 2h(Q|P) + E_Q\Big[\int_0^1 a_t^2 d\langle W\rangle_t\Big]. \end{aligned} \tag{70}$$

Since Q^* is equivalent to Q, we have

$$\begin{aligned} H(Q|Q^*) &= E_Q\big[\log\big(dQ^*/dQ)^{-1}\big] \\ &= E_Q\Big[\int_0^1 a_t dM_t + \frac{1}{2}\int_0^1 a_t^2 d\langle M\rangle_t\Big]. \end{aligned}$$

But M is a square-integrable martingale under Q and $E_Q[\int_0^1 a_t^2 d\langle M\rangle_t] < \infty$ for $Q \in \mathcal{Q}^*_{\mathcal{S}}$. This implies $E_Q\big[\int_0^1 a_t dM_t\big] = 0$, hence

$$H(Q|Q^*) = \frac{1}{2}E_Q\Big[\int_0^1 a_t^2 d\langle M\rangle_t\Big].$$

Thus,

$$W^2_{\mathcal{S}}(Q,P) \le E_Q\big[\langle W - W^Q\rangle_1 + ||A||^2_{\mathcal{S}}\big] \le 2h(Q|P) + 2H(Q|Q^*).$$

and so we have shown inequality (69).
2) Equality in (69) implies equality in (70), hence

$$E_Q\big[\langle W - W^Q\rangle_1\big] = 2h(Q|P).$$

Recall that the left-hand side satisfies equation (56). As in the proof of Theorem 32, it follows that $M = W^Q$. This implies $W = W^Q + A$ and $||A||_{\mathcal{H}} = ||A||_{\mathcal{S}} \in L^2(Q)$, hence

$$H(Q|P) = \frac{1}{2}E_Q\big[||A||^2_{\mathcal{H}}\big] < \infty,$$

due to Proposition 3.
Conversely, $H(Q|P) < \infty$ implies $h(Q|P) = 0$ and $Q \in \mathcal{Q}^*_{\mathcal{S}}$ with $\mathcal{Q}^* = P$, hence $H(Q|Q^*) = H(Q|P)$. Thus, the right-hand side of (69) reduces to $2H(Q|P) = E_Q\big[||B^Q||^2_{\mathcal{H}}\big]$. Moreover, since $W = W^Q + B^Q$ and $\langle W\rangle_t = t$ under Q, we get $A = B^Q$, and the left-hand side becomes $W^2_{\mathcal{S}}(Q,P) = W^2_{\mathcal{H},ad}(Q,P) = E_Q\big[||B^Q||^2_{\mathcal{H}}\big]$. Thus, equality holds in (69). □

Remark 37. *Inequality* (69) *includes inequality* (58) *for martingale measures as a special case. Indeed, for* $Q \in \mathcal{Q}_{\mathcal{M}} \subset \mathcal{Q}^*_{\mathcal{S}}$ *we have* $Q = Q^*$*, hence* $H(Q|Q^*) = 0$ *and*

$$W^2_{\mathcal{S}}(Q, P) \leq 2h(Q|P).$$

Part 2) of the preceding proof shows how Talagrand's inequality (9) *and the identity* (20) *for* $W_{\mathcal{H},ad}$ *follow from Theorem 36.*

Acknowledgments: This paper is dedicated to the memory of Mark Davis, a cherished fellow explorer of the rich landscape of Stochastic Analysis, including a long excursion to the uncharted territory of Mathematical Finance.

Work on this paper began in the fall of 2018 while I was giving lectures on "Optimal Transport on Wiener Space" at Jiao Tong University, Shanghai. It is a pleasure to express my thanks to Samuel Drapeau and his colleagues in the School of Mathematical Sciences and in the China Academy of Finance Research for their kind hospitality.

References

[1] B. Acciaio, J. Backhoff Veraguas, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Processes and their Applications*, 9(3):203–228, 2019.

[2] R. Eldan and D. Mikulincer. Stability of the Shannon-Stam inequality via the Föllmer process. *Probab. Theor. Relat. Fields*, 177:891–922, 2020.

[3] D. Feyel and A.S. Üstünel. Monge-Kantorovich measure transportation and Monge-Ampere equation on Wiener Space. *Probab. Theor. Relat. Fields*, 128(3):347–385, 2004.

[4] H. Föllmer. Time reversal on Wiener Space. In *Stochastic Processes - Mathematics and Physics (Bielefeld 1984)*, volume 1158 of *Lecture Notes in Mathematics*, pages 119–129. Springer, 1986.

[5] H. Föllmer. Random Fields and Diffusion Processes. In *Ecole d' été de Probabilités de St. Flour XV- XVII, 1985-87*, volume 1362 of *Lecture Notes in Mathematics*, pages 101–203. Springer, Berlin, 1988.

[6] H. Föllmer and M. Schweizer. Hedging of contingent claims under incomplete information. In *Applied Stochastic Analysis*, volume 5 of *Stochastic Monographs*, pages 389–414. Gordon and Breach, London, New York, 1986.

[7] N. Gantert. *Einige grosse Abweichungen der Brownschen Bewegung.* PhD thesis, Universität Bonn, 1991.

[8] N. Gantert. Self-similarity of Brownian motion and a large deviation principle for random fields on a binary tree. *Probability Theory and Related Fields*, 98:7–20, 1994.

[9] R. Lassalle. Causal transference plans and their Monge-Kantorovich problems. *Stochastic Analysis and Applications*, 36(1):1–33, 2018.

[10] J. Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré - Probabilités et Statistiques*, 49(3):885–899, 2013.

[11] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes I, General Theory*. Applications of Mathematics 5. Springer, 1977.

[12] P. A. Meyer. Un Cours sur les Intégrales Stochastiques. In *Séminaire de Probabilités X*, volume 511 of *Lecture Notes in Mathematics*, pages 246–400. Springer, 1976.

[13] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[14] S.S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory Probab. Appl.*, 18:784–788, 1973.

[15] C. Villani. *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften 338. Springer, 2009.