Doob Decomposition, Dirichlet Processes, and Entropies on Wiener Space

Hans Föllmer*

July 17, 2022

Abstract

As an extension of the Doob-Meyer decomposition of a semimartingale and the Fukushima representation of a Dirichlet process, we introduce a general Doob decomposition in continuous time, where a square-integrable process is represented as the sum of a martingale and a process with "vanishing local risk". For a probability measure Q on Wiener space, we discuss how entropy conditions on Q formulated with respect to Wiener measure P are connected with the Doob decomposition of the coordinate process W under Q. The situation is well understood if the relative entropy $H(Q|P)$ is finite; in this case the decomposition is classical and yields an immediate proof of Talagrand's transport inequality on Wiener space. To go beyond this restriction, we consider the specific relative entropy $h(Q|P)$ on Wiener space that was introduced by N. Gantert in [11]. We discuss its interplay with the Doob decomposition of W under Q and a corresponding version of Talagrand's inequality, with special emphasis on the case where W is a Dirichlet process under Q.

1 Introduction

Since the Sixties, the interplay between potential theory and the theory of Markov processes has been a rich source of inspiration for the general theory of stochastic processes. In particular, the Riesz decomposition of a superharmonic functions has its general counterpart in the *Doob-Meyer decomposition*

$$X = M + A \tag{1}$$

of a supermartingale. This has led to the general notion of a semimartingale X, defined by a Doob-Meyer decomposition (1) into a local martingale M and a predictable process A with paths of bounded variation. The canonical role of semimartingales is emphasized by the Bichteler-Dellacherie theorem, where they are characterized as general stochastic integrators.

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

M. Fukushima was the first to show that there are good reasons to go beyond the conceptual framework provided by the theory of semimartingales. Indeed, for a function F in the Dirichlet space of a symmetric Markov process Z, the process $X = F(Z)$ may not be a semimartingale, and thus may not admit a Doob-Meyer decomposition. However, M. Fukushima showed that X admits a decomposition of the form (1), where M is a square-integrable martingale and A is a process of "zero energy"; cf. [10]. This *Fukushima decomposition* has motivated the general notion of a Dirichlet process X; cf. [7] and [2].

But, as shown by S. E. Graversen and M. Rao in [14], representations of the form (1) have an even wider scope. In Section 2 we introduce a version that is convenient for our purpose. For a square-integrable adapted process $X = (X_t)_{0\leq t\leq 1}$ on a filtered probability space and any $N \geq 1$, we consider its Doob decomposition

$$X = M^N + A^N$$

in discrete time along the N-th dyadic partition of the unit interval. Assuming L^2-convergence of the random variables M_1^N, we are led to a *Doob decomposition in continuous time* of the form (1), where M is a square-ingrable martingale and A is "predictable" in the sense that the sum of the local prediction errors along the N-th dyadic partition converges to 0 as N increases to ∞. To avoid confusion with the standard notion of predictability, which is defined as measurability with respect to the predictable σ-field, we will say that A has *vanishing local risk*. Any process with zero energy has this property, and so our Doob decomposition in continuous time may be viewed as an extension of the Fukushima decomposition, and in particular of the Doob-Meyer decomposition. On the other hand, a process A with vanishing local risk is orthogonal to any square-integrable martingale. Thus, the Doob decomposition in continuous time may be viewed as a special case of the general decomposition obtained in [14]; cf. also the discussion of "weak Dirichlet processes"in [3] and [13].

In Section 3 we consider probability measures Q on the path space $C_0[0, 1]$. We denote by $\mathcal{Q}$ the class of all Q such that the coordinate process W admits a Doob decomposition $W = M + A$ in continuous time under Q. Our aim is to understand the impact of entropy bounds on $Q \in \mathcal{Q}$ with respect to Wiener measure P on the Doob decomposition of W under Q.

If the relative entropy $H(Q|P)$ is finite then Q is absolutely continuous with respect to P, and the Doob decomposition takes the classical form

$$W = W^Q + B^Q,$$

where W^Q is a Wiener process under Q, and where the paths of the process B^Q belong to the Cameron-Martin space $\mathcal{H}$. In this case, the decomposition yields an immediate proof of Talagrand's transport inequality

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

where $W_{\mathcal{H}}$ denotes the Wasserstein distance induced by the Cameron-Martin norm; cf. [16] and [9], or Corollary 15 below.

To go beyond the absolutely continuous case, we consider 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 [11], where it serves as a rate function for large deviations of the quadratic variation from its ergodic behaviour; cf. also [12]. In our context, it allows us to prove a version of Talagrand's inequality of the form

$$W_{\mathcal{D}}(Q,P) \leq \sqrt{2\,h(Q|P)}, \tag{3}$$

where the Wasserstein distance $W_{\mathcal{D}}$ is defined in terms of quadratic variation. This involves a careful analysis of the specific relative entropy, and in particular the inequality

$$2\,h(Q|P) \geq E_Q\big[M_1^2 - 1 + H(\lambda|q(\cdot)) + < A >_1\big], \tag{4}$$

where $q(\omega, dt)$ is the random measure on $[0,1]$ whose distribution function is given by the quadratic variation of the martingale M, λ denotes the Lebesgue measure on $[0,1]$, and $< A >$ is the quadratic variation of A. For a martingale measure Q, inequality (4) with $A = 0$ is proved in [9]; in the special case where $q(\cdot)$ is absolutely continuous, it was already shown in [11]. Here we extend it to a large class of measures $Q \in \mathcal{Q}$. As a corollary we obtain our version (3) of Talagrand's inequality for measures $Q \in \mathcal{Q}$ such that W is a Dirichlet process under Q.

2 Doob Decomposition in continuous time

Let $(\Omega, \mathcal{F})$ be a measurable space endowed with a right-continuous filtration $(\mathcal{F}_t)_{0\leq t\leq 1}$, and let Q be a probability measure on $(\Omega, \mathcal{F})$. In the sequel, the measure Q will vary, and so we do not assume that the filtration is completed with respect to Q.

Throughout this paper, the index N will refer to the N-th dyadic partition of the unit interval, that is,

$$D_N = \{i2^{-N} | i = 1, \dots, 2^N\},$$

and for fixed $N \geq 1$ we use the notation $t_i = i2^N$. For any square-integrable adapted process $Z = (Z_t)$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$, defined for $t \in [0,1]$ or at least for $t \in D_N$, we denote by

$$\Delta_{N,i} Z = Z_{t_i} - Z_{t_{i-1}}$$

the increments of Z along the N-th dyadic partition, and by

$$\zeta_{N,i}^2 = E_Q[(\Delta_{N,i}Z)^2|\mathcal{F}_{t_{i-1}}] - \big(E_Q[\Delta_{N,i}Z|\mathcal{F}_{t_{i-1}}]\big)^2 \tag{5}$$

their conditional variances given the past. Note that $\zeta^2_{N,i}$ can be viewed as the conditional prediction error if the increment $\Delta_{N,i}Z$ is predicted by its conditional expectation under Q.

Definition 1. *For each $N \geq 1$, the sum*

$$R_N(Z) := \sum_{i=1}^{2^N} \zeta^2_{N,i} \; \in L^1(Q)$$

will be called the local risk *of the process Z along the N-th dyadic partition. We will say that the process Z has* vanishing local risk *if*

$$\lim_{N\uparrow\infty} R_N(Z) = 0 \quad \text{in } L^1(Q). \tag{6}$$

For two square-integrable adapted processes Z and $\tilde{Z}$ we denote by

$$CV_N(Z,\tilde{Z}) := \frac{1}{2}\big(R_N(Z+\tilde{Z}) - R_N(Z) - R_N(\tilde{Z})\big)$$

the sum of the conditional covariances of the increments along the N-th dyadic partition, and we say that Z and $\tilde{Z}$ are orthogonal *if*

$$\lim_{N\uparrow\infty} CV_N(Z,\tilde{Z}) = 0 \quad \text{in } L^1(Q). \tag{7}$$

Let us now fix an adapted right-continuous process X on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$ such that

$$X = (X_t)_{0\leq t\leq 1} \subset L^2(Q).$$

For any $N \geq 1$, consider its Doob decomposition

$$X_t = M^N_t + A^N_t, \qquad t \in D_N \tag{8}$$

in discrete time along the N-th dyadic partition. Thus, the process $A^N = (A^N_t)_{t\in D_N}$ is defined by $A^N_0 = 0$ and the increments

$$\Delta_{N,i}A^N = E_Q[\Delta_{N,i}X|\mathcal{F}_{t_{i-1}}],$$

and $M^N = (M^N_t)_{t\in D_N}$ is a square-integrable martingale in discrete time with initial value $M^N_0 = X_0$.

Remark 2. *The process A^N is "predictable" in discrete time, that is, $A^N_{t_i}$ is $\mathcal{F}_{t_{i-1}}$-measurable for each $t_i \in D_N$. Equivalently, this property can be expressed in terms of actual predictions. Indeed, if we predict the increments of A^N by taking conditional expectations given the past, then the local prediction errors, defined as the conditional variances $(\alpha^N)^2_{N,i}$ of the process A^N, are all equal to 0, i.e.,*

$$R_N(A^N) = \sum_{i=1}^{2^N} (\alpha^N)^2_{N,i} = 0 \quad Q-\text{a.s.}.$$

In this sense the process A^N carries no local risk. In fact, the local risk of the process X along the N-th dyadic partition is fully captured by the martingale M^N, that is,

$$R_N(X) = R_N(M^N), \tag{9}$$

since the conditional variances are the same for X and for M^N. This alternative interpretation of "predictability" in discrete time motivates our definition of vanishing local risk and the following version of the Doob decomposition in continuous time.

Theorem 3. *1) The following two properties of the process X with respect to Q are equivalent:*

i) The random variables $(M_1^N)_{N=1,2,\dots}$ in (8) form a Cauchy sequence in $L^2(Q)$,

ii) X admits a Doob decomposition in continuous time of the form

$$X = M + A, \tag{10}$$

where $M = (M_t)_{0\leq t\leq 1}$ is a square-integrable right-continuous martingale such that $M_0 = X_0$, and where the process $A = (A_t)_{0\leq t\leq 1}$ has vanishing local risk.

2) The decomposition (10) of X into a square-integrable martingale M and a process A with vanishing local risk is unique.

Proof. 1) Suppose that $(M_1^N)_{N=1,2,\dots}$ is a Cauchy sequence in $L^2(Q)$, hence convergent in $L^2(Q)$ to a random variable $M_1 \in L^2(Q)$. We denote by $M = (M_t)_{0\leq t\leq 1}$ a right-continuous version of the square-integrable martingale given by the conditional expectations $E_Q[M_1|\mathcal{F}_t]$; cf. [6] or [4], Ch. VI.5. Then the process $A = (A_t)_{0\leq t\leq 1}$ defined by $A = X - M$ is right-continuous, adapted, and square-integrable. For $N \geq 1$, the increments of A along the N-th dyadic partition satisfy

$$\begin{aligned} \Delta_{N,i}A - E_Q[\Delta_{N,i}A|\mathcal{F}_{t_{i-1}}] &= \Delta_{N,i}X - \Delta_{N,i}M - E_Q[\Delta_{N,i}X|\mathcal{F}_{t_{i-1}}] \\ &= \Delta_{N,i}M^N - \Delta_{N,i}M \\ &= \Delta_{N,i}(M^N - M). \end{aligned}$$

Thus, the conditional variance of $\Delta_{N,i}A$ is given by

$$\alpha_{N,i}^2 = E_Q[(\Delta_{N,i}(M^N - M))^2|\mathcal{F}_{t_{i-1}}]. \tag{11}$$

Since $M^N - M$ is a square-integrable martingale along D_N with initial value $(M^N - M)_0 = 0$, we obtain

$$E_Q[\sum_{i=1}^{2^N} \alpha_{N,i}^2] = E_Q[\sum_{i=1}^{2^N}(\Delta_{N,i}(M^N - M))^2] = E_Q[(M_1^N - M_1)^2].$$

This implies

$$\lim_{N\uparrow\infty} E_Q[\sum_{i=1}^{2^N} \alpha_{N,i}^2] = 0, \tag{12}$$

and so we have shown that the process A has vanishing local risk.

Conversely, if X admits a decomposition (10) then the preceding equation (12) holds again, and so (6) implies that $(M_1^N)_{N=1,2,\ldots}$ is a Cauchy sequence in $L^2(Q)$.

2) To check uniqueness of the decomposition (10), suppose that

$$X = M + A = \tilde{M} + \tilde{A},$$

where M and $\tilde{M}$ are square-integrable martingales, and A and $\tilde{A}$ are processes with vanishing local risk. For any $N \geq 1$ we obtain

$$\begin{aligned} E_Q[(M_1 - \tilde{M}_1)^2] &= \sum_{i=1}^{2^N} E_Q[(\Delta_{N,i}(M - \tilde{M}))^2] = \sum_{i=1}^{2^N} E_Q[(\Delta_{N,i}(\tilde{A} - A))^2] \\ &\leq 2\sum_{i=1}^{2^N} E_P[\tilde{\alpha}_{N,i}^2 + \alpha_{N,i}^2], \end{aligned} \tag{13}$$

denoting by $\tilde{\alpha}_{N,i}^2$ and $\alpha_{N,i}^2$ the conditional variances of $\tilde{A}$ and A; in the last step we use the fact that

$$E_Q[\Delta_{N,i}\tilde{A}|\mathcal{F}_{t_i}] = E_Q[\Delta_{N,i}A|\mathcal{F}_{t_i}],$$

since both terms are equal to $E_Q[\Delta_{N,i}X|\mathcal{F}_{t_i}]$. For $N \uparrow \infty$ the right hand side of (13) converges to 0, and this implies $M_1 = \tilde{M}_1$ Q-a.s., hence $M = \tilde{M}$ and $A = \tilde{A}$. □

Lemma 4. *Let $A = (A_t)_{0\leq t\leq 1}$ be a square-integrable, right-continuous, and adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$, and consider the following properties of A:*

i) A has continuous paths of bounded variation, and the total variation process $|A|$ satisfies $|A|_1 \in L^2(Q)$.

ii) A has "zero energy" in the sense that

$$\lim_{N\uparrow\infty} E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A)^2\big] = 0, \tag{14}$$

iii) A has vanishing local risk,

iv) A is orthogonal to any square-integrable martingale L.

Then we have

$$i) \Longrightarrow ii) \Longrightarrow iii) \Longrightarrow iv).$$

Proof. Since $\alpha_{N,i}^2 \leq E_Q\big[(\Delta_{N,i}A)^2|\mathcal{F}_{t_i}\big]$, the process A has vanishing local risk as soon as it has zero energy. As to the first implication, note that

$$E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A)^2\big] \leq E_Q\big[\sum_{i=1}^{2^N} C_N|\Delta_{N,i}A|\big] = E_Q\big[C_N|A|_1\big], \tag{15}$$

where $C_N := \max_i |\Delta_{N,i} A| \leq |A|_1$. Property i) implies $\lim_{N\uparrow\infty} C_N = 0$ and $C_N |A|_1 \leq |A|_1^2 \in L^1(Q)$. By Lebesgue's theorem, the right hand side of (15) converges to 0, and so A has energy 0. As to the last implication, note that

$$CV_N(A, L) \leq R_N(A)^{1/2} R_N(L)^{1/2},$$

hence

$$E_Q\big[CV_N(A, L)\big] \leq E_Q\big[R_N(A)\big]^{1/2} E_Q\big[(L_1 - L_0)^2\big]^{1/2}.$$

□

Remark 5. *Suppose that X admits a continuous Doob decomposition* (10)*.*
1) The preceding implication iii) $\Longrightarrow$ iv) shows that this can be viewed as a special case of the decomposition derived in [14]; see also the discussion of "weak Dirichlet processes" in [3] and [13].

2) Applying property iv) to the martingale M in (10)*, we see that*

$$\lim_{N\uparrow\infty} R_N(X) = \lim_{N\uparrow\infty} R_N(M) \quad \text{in } L^1(Q),$$

that is, in the limit the local risk of X is carried by M. This can be seen as the continuous-time version of equation 9.

Definition 6. *Let us say that X is a* Dirichlet process *if it admits a Doob decomposition* (10) *such that A is a process with zero energy. In this case,* (10) *is also called the* Fukushima decomposition *of X.*

The preceding lemma shows that the notion of vanishing local risk has a wide scope. Combined with the uniqueness of the decomposition (10), it implies the following corollary.

Corollary 7. *The class of processes X that admit a Doob decomposition of the form* (10) *includes*

i) a large class of semimartingales, and in that case (10) *reduces to the Doob-Meyer decomposition of X,*

ii) the class of Dirichlet processes, and in that case (10) *reduces to the Fukushima decomposition of X.*

Remark 8. *Suppose that X admits a Doob decomposition* (10) *under Q. Then X is a Dirichlet process if and only if*

$$\lim_{N\uparrow\infty} E_Q[\sum_{i=1}^{2^N} (\Delta_{N,i} A^N)^2] = 0. \tag{16}$$

Indeed, the weaker condition (16) *is in fact equivalent to condition* (15) *as soon as A has vanishing local risk.*

The following equivalence was stated in [7], where condition (17) is taken as the definition of a Dirichlet process.

Theorem 9. *The process X is is Dirichlet process if and only if the processes A^N appearing in the discrete Doob decompositions* (8) *satisfy the condition*

$$\lim_{N\uparrow\infty} \sup_{K\geq N} E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^K)^2\big] = 0. \tag{17}$$

Proof. We include a proof, since the proof in [7] contains several typos.

For each $L \geq N$, we obtain

$$\begin{aligned} E_Q\big[(A_1^L - A_1^N)^2\big] &= E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^L - \Delta_{N,i}A^N)^2\big] \\ &\leq 2\big(E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^L)^2 + E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^N)^2\big) \\ &\leq 4\sup_{K\geq N} E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^K)^2\big], \end{aligned}$$

since $A^L - A^N = M^N - M^L$ is a martingale in discrete time along D_N. Thus, condition (17) implies that $(A_1^N)_{N=1,2,...}$, and hence $(M_1^N)_{N=1,2,...}$, is a Cauchy sequence in $L^2(Q)$. It also implies condition (16). In view of Theorem 3 and the preceding remark, it follows that X is a Dirichlet process.

Conversely, since

$$\Delta_{N,i}A^K = \Delta_{N,i}A^N + \Delta_{N,i}(M^N - M^K),$$

we obtain

$$\begin{aligned} E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^K)^2 &\leq 2\big(E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^N)^2\big] + E_Q\big[(M_1^N - M_1^K)^2\big]\big) \\ &\leq 2\big(E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A^N)^2\big] + \sup_{L\geq N} E_Q\big[(M_1^N - M_1^L)^2\big]\big) \end{aligned}$$

for any $K \geq N$. Thus, condition (17) is satisfied as soon as X is a Dirichlet process. □

3 Entropies and Couplings on Wiener space

From now on, the underlying measurable space will be the path space

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

of all continuous functions ω on $[0,1]$ with initial value $\omega(0) = 0$. We denote by $(\mathcal{F}_t)_{0\leq t\leq 1}$ the right-continuous filtration on Ω generated by the coordinate process

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

defined by $W_t(\omega) = \omega(t)$. We set $\mathcal{F} = \mathcal{F}_1$, and we 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 $[0, 1]$. For $\omega \in \Omega$ we write

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

Definition 10. *We denote by $\mathcal{Q}$ the class of all probability measures Q on $(\Omega, \mathcal{F})$ such that the process W admits a Doob decomposition* (10) *under Q with continuous paths, that is,*

$$W = M + A,$$

where M is a square-integrable continuous martingale under Q, and where A is a continuous adapted process with vanishing local risk under Q. For $Q \in \mathcal{Q}$ we will write

- $Q \in \mathcal{Q}_{\mathcal{M}}$ *if $A = 0$, that is, Q is a martingale measure,*
- $Q \in \mathcal{Q}_{\mathcal{H}}$ *if A satisfies $E_Q\left[||A||^2_{\mathcal{H}}\right] < \infty$,*
- $Q \in \mathcal{Q}_{\mathcal{S}}$ *if A has continuous paths of bounded variation, and the total variation process $|A|$ satisfies $|A|_1 \in L^2(Q)$,*
- $Q \in \mathcal{Q}_{\mathcal{D}}$ *if A has zero energy, that is, W is a Dirichlet process under Q.*

Lemma 4 shows that

$$\mathcal{Q}_{\mathcal{M}} \subset \mathcal{Q}_{\mathcal{H}} \subset \mathcal{Q}_{\mathcal{S}} \subset \mathcal{Q}_{\mathcal{D}} \subset \mathcal{Q}. \tag{18}$$

For a given measure $Q \in \mathcal{Q}$, we are now going to study the impact of entropy bounds on the Doob decomposition (10) of the process W under Q. These bounds will be formulated in terms of relative entropies with respect to Wiener measure P.

Remark 11. *Recall that, 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*

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

and that $H(\nu|\mu) \geq 0$, with equality if and only if $\mu = \nu$. Moreover,

$$\lim_{n\uparrow\infty} H_n(\nu|\mu) = H(\nu|\mu) \tag{19}$$

if $(\mathcal{S}_n)_{n=1,2...}$ is a sequence of σ-fields increasing to $\mathcal{S}$ and $H_n(\nu|\mu)$ denotes the relative entropy of ν with respect to μ on $(S, \mathcal{S}_n)$. Note also that equation (19) *extends to the case where ν or μ is a non-negative finite measure on $(S, \mathcal{S})$.*

First we review the case where Q has finite relative entropy $H(Q|P)$ with respect to Wiener measure P. The following proposition is well known; cf., for example, [8] or [9].

Proposition 12. *For any probability measure Q on $(\Omega, \mathcal{F})$,*

$$H(Q|P) < \infty \quad \Longleftrightarrow \quad Q \ll P \text{ and } Q \in \mathcal{Q}_{\mathcal{H}}.$$

In this case, the Doob decomposition (10) *takes the form*

$$W = W^Q + B^Q, \tag{20}$$

where W^Q is a Wiener process under Q and the process B^Q has paths in $\mathcal{H}$, and the relative entropy is given by

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

Remark 13. *The process $b^Q := \dot{B}^Q$ will be called the* intrinsic drift *of Q. Note that equation* (20) *can be read as*

$$dW_t = dW_t^Q + b_t^Q(W)dt.$$

Thus, any measure Q on path space such that $H(Q|P) < \infty$ can be viewed as a weak solution of the stochastic differential equation

$$dX = dZ + b_t^Q(X)dt,$$

where Z is required to be a Wiener process, and its relative entropy takes the form

$$H(Q|P) = \frac{1}{2} E_Q\Big[\int_0^1 (b^Q)_t^2 dt\Big].$$

As first observed by J. Lehec in [16], Proposition 12 yields an immediate proof of *Talagrand's inequality on Wiener space*, which relates the relative entropy $H(Q|P)$ to the Wasserstein distance $W_{\mathcal{H}}(Q, P)$ defined in terms of the Cameron-Martin norm.

Definition 14. *For any probability measure Q on $(\Omega, \mathcal{F})$, we define the Wasserstein distance $W_{\mathcal{H}}(Q, P)$ between Q and P as*

$$W_{\mathcal{H}}(Q, P) = \inf_{\gamma \in \Gamma(P,Q)} \int ||\omega - \eta||^2_{\mathcal{H}} \gamma(d\omega, d\eta)^{1/2}, \tag{21}$$

where $\Gamma(P, Q)$ denotes the class of all probability measures γ on the product space $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$ with marginals P and Q.

Equivalently, we can write

$$W_{\mathcal{H}}(Q, P) = \inf \ \tilde{E}[||\tilde{X} - \tilde{Y}||^2_{\mathcal{H}}]^{1/2}, \tag{22}$$

where the infimum is taken over all couples $(\tilde{X}, \tilde{Y})$ of Ω-valued random variables on some probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $\tilde{X}$ and $\tilde{Y}$ have distributions P and Q, respectively. Such a couple, and also any measure $\gamma \in \Gamma(P,Q)$, will be called a *coupling of P and Q*. We refer to [18] for a thorough discussion of Wasserstein distances in various contexts.

Corollary 15. *Any probability measure Q on $(\Omega, \mathcal{F})$ satisfies Talagrand's inequality*

$$W_{\mathcal{H}}(Q,P) \leq \sqrt{2H(Q|P)}. \tag{23}$$

Proof. If $H(Q|P) < \infty$ then 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

$$E_Q\left[||W - W^Q||_{\mathcal{H}}^2\right] = 2H(Q|P). \tag{24}$$

Thus, (23) follows from the definition of the Wasserstein distance $W_{\mathcal{H}}(Q,P)$. □

Remark 16. *Inequality* (23) *on Wiener space was first stated by Feyel and Üstunel in* [5]. *However, using the Lévy-Ciesielski representation of Brownian motion in terms of Schauder functions, it can also be seen as a direct translation, for $n = \infty$, of Talagrand's original inequality in* [17], *where Q is a probability measure on Euclidean space $\mathbb{R}^n$ with $n \in \{1, \ldots, \infty\}$, P is the product of standard normal distributions, and the Wasserstein distance is defined in terms of the Euclidean norm; cf.* [9].

Remark 17. *Note that the coupling (W^Q, W) of P and Q, defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq 1}, Q)$, is* adaptive *in the sense that both processes are adapted and the first is a Wiener process with respect to the given filtration. As shown by Lassalle in* [15], *(W^Q, W) is in fact the* optimal *adaptive coupling of P and Q. Thus, equation* (24) *shows that Talagrand's inequality reduces to the equality*

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

if the left hand side is defined as in (22), *but taking the infimum only over the* adaptive *couplings of P and Q; cf.* [15] *or* [9]. *For a systematic discussion of the optimal transport problem* (22) *under various constraints we refer to* [1].

Let us now go beyond the case where Q has finite entropy with respect to Wiener measure P. For any $N \geq 1$, consider the discretized filtration

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

on $\Omega = C_0[0,1]$. We set $\mathcal{F}_N = \mathcal{F}_{N,1} = \sigma(\{W_s | s \in D_N\})$, and we denote by $H_N(Q|P)$ the relative entropy of Q with respect to P on the σ-field $\mathcal{F}_N$. Since the σ-fields $\mathcal{F}_N$ increase to $\mathcal{F}$, we have

$$H(Q|P) = \lim_{N\uparrow\infty} H_N(Q|P).$$

From now on we assume that the finite-dimensional marginals of Q are such that

$$H_N(Q|P) < \infty, \qquad N = 1, 2, \dots \tag{25}$$

and we focus on the case $H(Q|P) = \infty$. It is then natural to rescale the finite-dimensional entropies $H_N(Q|P)$ in order to obtain meaningful results.

The following concept of specific relative entropy on Wiener space was introduced by N. Gantert in her thesis [11], where it plays the role of a rate function for large deviations of the quadratic variation from its ergodic behaviour; cf. also [12]. In our context, it will allow us to extend Talagrand's inequality on Wiener space beyond the absolutely continuous case $Q \ll P$, and to throw new light on the Doob decomposition in continuous time.

Definition 18. *For any probability measure Q on $(\Omega, \mathcal{F})$, the* specific relative entropy *of Q with respect to Wiener measure P is defined as*

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

To illustrate the role of specific relative entropy $h(Q|P)$, and in particular its connection with Dirichlet processes, we first consider the particularly transparent case where the coordinate process W is square-integrable and has *independent increments* under Q. Then the increments are normally distributed under Q, there are functions a and β in $C_0[0,1]$ such that

$$E_Q[W_t] = a(t) \quad \text{and} \quad \mathrm{var}_Q(W_t) = \beta(t),$$

and the function β is strictly increasing due to our assumption (25). In this case, let us write

$$Q = Q_{a,\beta}.$$

Note that $Q_{a,\beta} \in \mathcal{Q}$, and that the Doob decomposition (10) under $Q_{a,\beta}$ takes the form $W = M + A$, where M is a Gaussian martingale with quadratic variation

$$<M>_t = \beta(t),$$

and where the deterministic process A given by $A_t(\omega) = a(t)$ clearly carries no local risk.

Let q denote the finite measure on $[0,1]$ with distribution function β, and denote by

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

its Lebesgue decomposition with respect to Lebesgue measure λ on $[0,1]$, where q_s denotes the singular part and $\sigma^2(\cdot)$ is the density of the absolutely continuous part.

Proposition 19. *For $Q = Q_{a,\beta}$, the specific relative entropy $h(Q|P)$ is given by*

$$h(Q|P) = \frac{1}{2}\big(\beta(1) - 1 + H(\lambda|q) + \liminf_{N\uparrow\infty} \sum_{i=1}^{2^N} (\Delta_{N,i}a)^2\big). \tag{27}$$

In particular, $H(\lambda|q) < \infty$ implies that $h(Q|P)$ exists as a finite limit if and only if the function a has "finite energy", that is,

$$h(Q|P) = \lim_{N\uparrow\infty} 2^{-N} H_N(Q|P) < \infty \iff \exists <a>_1 := \lim_{N\uparrow\infty} \sum_{i=1}^{2^N} (\Delta_{N,i}a)^2 < \infty.$$

In this case,

$$h(Q|P) = \frac{1}{2} q_s([0,1]) + \int_0^1 f(\sigma^2(t))dt + \frac{1}{2} <a>_1, \tag{28}$$

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

$$Q \in \mathcal{Q}_{\mathcal{D}} \iff h(Q|P) = \frac{1}{2} q_s([0,1]) + \int_0^1 f(\sigma^2(t))dt, \tag{29}$$

that is, W is a Dirichlet process under Q iff $h(Q|P)$ only depends on β and not on a.

Proof. For two normal distributions $N(m, \sigma^2)$ and $N(\tilde{m}, \tilde{\sigma}^2)$ on $\mathbb{R}^1$, the relative entropy is given by

$$H\big(N(\tilde{m}, \tilde{\sigma}^2)|N(m, \sigma^2)\big) = f(\tilde{\sigma}^2/\sigma^2) + \frac{1}{2} \frac{(\tilde{m} - m)^2}{\sigma^2}. \tag{30}$$

Since the increments $\Delta_{N,i}W$ along the N-th dyadic partition are independent under both Q and P, with distribution $N(\Delta_{N,i}a, \Delta_{N,i}\beta)$ under Q and $N(0, 2^{-N})$ under P, we get

$$\begin{aligned} H_N(Q|P) &= \sum_{i=1}^{2^N} H\big(N(\Delta_{N,i}a, \Delta_{N,i}\beta)|N(0, 2^{-N})\big) \\ &= \sum_{i=1}^{2^N} f(2^N \Delta_{N,i}\beta) + \frac{1}{2} \sum_{i=1}^{2^N} 2^N (\Delta_{N,i}a)^2. \end{aligned}$$

Thus,

$$2^{-N} H_N(Q|P) = \int_0^1 f(\varphi_N(t))dt + \frac{1}{2} \sum_{i=1}^{2^N} (\Delta_{N,i}a)^2,$$

if we denote by φ_N the density of the finite measure q on $[0,1]$ with respect to Lebesgue measure λ on the discrete σ-field $\mathcal{B}_N$ generated by the N-th dyadic partition. Note that $\varphi_N > 0$ since β is strictly increasing. Denoting by $H_N(\lambda|q) = \int \log \varphi_N^{-1} d\lambda$ the relative entropy of λ with respect to q on $\mathcal{B}_N$, we can write

$$2^{-N} H_N(Q|P) = \frac{1}{2}\Big(q([0,1]) - 1 + H_N(\lambda|q) + \sum_{i=1}^{2^N} (\Delta_{N,i}a)^2\Big). \tag{31}$$

Since $q([0,1]) = \beta(1)$, and since $H_N(\lambda|q)$ increases to $H(\lambda|q)$, we obtain equation (27). If we assume $H(\lambda|q) < \infty$ and the existence of a finite limit $< a >_1$, then equation (27) reduces to (28) since

$$\beta(1) = q_s([0,1]) + \int_0^1 \sigma^2(t)dt.$$

In particular, we obtain the characterization (29) of a measure $Q = Q_{a,\beta} \in \mathcal{Q}_{\mathcal{D}}$. $\square$

Let us now consider the general case $Q \in \mathcal{Q}$. Thus, the coordinate process W admits a continuous Doob decomposition

$$W = M + A \tag{32}$$

under Q, where M is a continuous square-integrable martingale and A is a square-integrable, continuous and adapted process with vanishing local risk. Consider the continuous quadratic variation process $<M>$ of M and the corresponding finite random measure $q(\omega, dt)$ on $[0,1]$ with distribution function $<M>(\omega)$, and denote by

$$q(\omega, dt) = q_s(\omega, dt) + \sigma^2(\omega, t)dt \tag{33}$$

its Lebesgue decomposition into a singular and an absolutely continuous part with respect to Lebesgue measure λ on $[0,1]$; cf. [9] for an explicit construction. Our aim is to show how the specific relative entropy $h(Q|P)$ depends on the Doob decomposition (32), and in particular on the random measure $q(\cdot, dt)$.

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

$$a_{N,i} = E_Q\left[\Delta_{N,i}W|\mathcal{F}_{N,t_{i-1}}\right] = E_Q\left[\Delta_{N,i}A|\mathcal{F}_{N,t_{i-1}}\right]$$

its conditional mean, by

$$\tilde{\sigma}^2_{N,i} = E_Q\left[(\Delta_{N,i}W)^2|\mathcal{F}_{N,t_{i-1}}\right] - a^2_{N,i}$$

its conditional variance, and by

$$\sigma^2_{N,i} = E_Q\left[(\Delta_{N,i}M)^2|\mathcal{F}_{N,t_{i-1}}\right] = E_Q\left[<M>_{t_i} - <M>_{t_{i-1}} |\mathcal{F}_{N,t_{i-1}}\right] \tag{34}$$

the conditional variance of the martingale increment $\Delta_{N,i}M$.

Lemma 20. *The finite-dimensional entropy $H_N(Q|P)$ can be decomposed as follows:*

$$\begin{aligned} H_N(Q|P) \quad &= \quad H_N(Q|Q_N) + E_Q\big[\sum_{i=1}^{2^N} f(2^N\sigma^2_{N,i})\big] + \frac{1}{2}2^N E_Q\big[\sum_{i=1}^{2^N} a^2_{N,i}\big] \\ &+ \quad \frac{1}{2}E_Q\big[\sum_{i=1}^{2^N}(\tilde{\sigma}^2_{N,i} - \sigma^2_{N,i})(2^N - \sigma^{-2}_{N,i})\big] \end{aligned} \tag{35}$$

where f is the function defined in Proposition 19, and where Q_N denotes the probability measure on $(\Omega, \mathcal{F}_N)$ such that the increments $\Delta_{N,i}W$ have conditional distribution $N(a_{N,i}, \sigma^2_{N,i})$ given the σ-field $\mathcal{F}_{N,t_{i-1}}$.

Proof. Since

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

and since

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

we obtain equation (35). □

Let us first look at the asymptotic behavior of the second term on the right hand side of equation (35). We denote by $Q \otimes q$ 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 $Q \otimes \lambda$, 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 $A_t \in \mathcal{F}_t$, and we denote by

$$H(Q \otimes \lambda | Q \otimes q) = E_Q[H(\lambda|q(\cdot))]$$

the relative entropy of $Q \otimes \lambda$ with respect to $Q \otimes q$ on $\mathcal{P}$.

Lemma 21.

$$\begin{aligned} &\lim_{N\uparrow\infty} 2^{-N} E_Q\Big[\sum_{i=1}^{2^N} f(2^N\sigma^2_{N,i})\Big] \\ &\quad = \frac{1}{2}\big(E_Q[q(\cdot,[0,1])] - 1 + H(Q\otimes\lambda|Q\otimes q)\big) \\ &\quad = \frac{1}{2}E_Q[q_s(\cdot,[0,1])] + E_Q\Big[\int_0^1 f(\sigma^2(\cdot,t))dt\Big]. \end{aligned} \tag{36}$$

Proof. Since

$$E_Q\Big[\int_0^1 \sigma^2_N(\cdot,t)dt\Big] = E_Q\big[< M >_1 \big] = E_Q\big[q(\cdot,[0,1])\big]$$

for any $N \geq 1$, we can write

$$\begin{aligned} 2^{-N} E_Q\big[\sum_{i=1}^{2^N} f(2^N \sigma_{N,i}^2)\big] &= E_Q\big[\int_0^1 f(\sigma_N^2(\cdot,t))dt\big] \\ &= \frac{1}{2}\big(E_Q[q(\cdot,[0,1])] - 1 - E_Q[\int_0^1 \log \sigma_N^2(\cdot,t)dt]\big) \\ &= \frac{1}{2}\big(E_Q[q(\cdot,[0,1])] - 1 + H_N(Q\otimes\lambda|Q\otimes q)\big), \end{aligned}$$

where $H_N(Q\otimes\lambda|Q\otimes q)$ denotes the relative entropy of $Q\otimes\lambda$ with respect to $Q\otimes q$ on $\mathcal{P}_N$. Since $\mathcal{P}_N$ increases to $\mathcal{P}$, these entropies increase to

$$H(Q\otimes\lambda|Q\otimes q) = E_Q\big[\int_0^1 \log(\sigma_N^{-2}(\cdot,t))dt\big],$$

and this yields equation (36). □

For a martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$ with absolutely continuous quadratic variation, the following proposition is due to N. Gantert in [11]. Here we extend it to the case where the quadration variation may have a singular component; see also [9].

Proposition 22. *For a martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$,*

$$\begin{aligned} h(Q|P) &\geq \frac{1}{2}\big(E_Q[q(\cdot,[0,1])] - 1 + H(Q\otimes\lambda|Q\otimes q)\big) \\ &= \frac{1}{2}E_Q\big[q_s(\omega,[0,1])\big] + E_Q\big[\int_0^1 f\big(\sigma^2(\omega,t)\big)dt\big]. \end{aligned} \tag{37}$$

If $h(Q|P) < \infty$ then we have $H(Q\otimes\lambda|Q\otimes q) < \infty$, and in particular

$$\sigma^2(\cdot,\cdot) > 0 \qquad Q\otimes\lambda - a.s.. \tag{38}$$

Moreover, equality holds in (37) *if and only if Q is "almost locally normal" in the sense that*

$$\lim_{N\uparrow\infty} 2^{-N} H_N(Q|Q_N) = 0. \tag{39}$$

Proof. For $Q \in \mathcal{Q}_{\mathcal{M}}$ we have $A = 0$, hence $a_{N,i} = 0$ and $\tilde{\sigma}_{N,i}^2 = \sigma_{N,i}^2$. Thus, equation (35) implies

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

and so inequality (37) as well as the condition for equality follow from Lemma 21. Due to (37), $h(Q|P) < \infty$ implies $H(Q\otimes\lambda|Q\otimes q)\big) < \infty$, hence $Q\otimes\lambda \ll Q\otimes q$, and in particular (38). □

Let us denote by

$$\tilde{\alpha}^2_{N,i} = E_Q[(\Delta_{N,i}A)^2|\mathcal{F}_{N,t_{i-1}}] - a^2_{N,i} \tag{40}$$

the conditional variance of $\Delta_{N,i}A$ with respect to $\mathcal{F}_{N,t_{i-1}} \subset \mathcal{F}_{t_{i-1}}$, and recall the definition of $\alpha^2_{N,i}$ in (11). Since $\tilde{\alpha}^2_{N,i}$ is defined with respect to the smaller σ-field, we have

$$E_Q[\alpha^2_{N,i}] \leq E_Q[\tilde{\alpha}^2_{N,i}] \leq E_Q[(\Delta A_{N,i})^2]. \tag{41}$$

This shows that the condition

$$\lim_{N\uparrow\infty} E_Q\big[\sum_{i=1}^{2^N} \tilde{\alpha}^2_{N,i}\big] = 0 \tag{42}$$

strengthens our assumption on $Q \in \mathcal{Q}$ that A has vanishing local risk, and that it is satisfied as soon as A has energy 0 under Q.

Remark 23. *Note that we have $\tilde{\alpha}^2_{N,i} = \alpha^2_{N,i}$ as soon as Q is Markovian, that is, if W is a Markov process under Q. Thus, condition (42) is satisfied for any Markovian $Q \in \mathcal{Q}$.*

Definition 24. *We denote by $\mathcal{Q}_{\mathcal{E}}$ the class of all probability measures $Q \in \mathcal{Q}$ such that*

i) Q satisfies condition (42),

ii) the process A has "finite energy"under Q, that is,

$$\exists < A >_1 := \lim_{N\uparrow\infty} \sum_{i=1}^{2^N} (\Delta_{N,i}A)^2 \quad \text{in } L^1(Q). \tag{43}$$

Thus, the chain of inclusions in (18) can be extended as follows:

$$\mathcal{Q}_{\mathcal{M}} \subset \mathcal{Q}_{\mathcal{H}} \subset \mathcal{Q}_{\mathcal{S}} \subset \mathcal{Q}_{\mathcal{D}} \subset \mathcal{Q}_{\mathcal{E}} \subset \mathcal{Q}.$$

In [9], Proposition 22 for a martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$ is extended to a large class of semimartingale measures $Q \in \mathcal{Q}_{\mathcal{S}}$. Here we go two steps further and consider the case $Q \in \mathcal{Q}_{\mathcal{E}}$, and in particular the case $Q \in \mathcal{Q}_{\mathcal{D}}$ where W is a Dirichlet process under Q.

Theorem 25. *Let $Q \in \mathcal{Q}_{\mathcal{E}}$ be such that the variance $\sigma^2(\cdot,\cdot)$ in (33) is bounded away from 0. Then*

$$\begin{aligned} h(Q|P) &\geq \frac{1}{2}\big(E_Q[q(\cdot,[0,1])] - 1 + E_Q\big[H\big(\lambda|q(\cdot)\big)\big] + \frac{1}{2}E_Q[< A >_1] \\ &= \frac{1}{2}E_Q\big[q_s(\cdot,[0,1])\big] + E_Q\big[\int_0^1 f\big(\sigma^2(\cdot,t)\big)dt\big] + \frac{1}{2}E_Q[< A >_1] \end{aligned} \tag{44}$$

If $h(Q|P) < \infty$ then equality holds if and only if Q satisfies condition (39). In that case,

$$Q \in \mathcal{Q}_{\mathcal{D}} \iff h(Q|P) = \frac{1}{2}E_Q\big[q_s(\omega,[0,1])\big] + E_Q\big[\int_0^1 f\big(\sigma^2(\omega,t)\big)dt\big]. \tag{45}$$

Proof. Equation (35) can be written as

$$
\begin{aligned}
2^{-N}H_N(Q|P) &= 2^{-N}H_N(Q|Q_N) + E_Q\big[\int_0^1 f(\sigma_N^2(\cdot,t)dt)\big] + \frac{1}{2}I_N \\
&+ \frac{1}{2}E_Q\big[\sum_{i=1}^{2^N}\delta_{N,i}(1-2^{-N}\sigma_{N,i}^{-2})\big], \qquad (46)
\end{aligned}
$$

where

$$I_N := E_Q\big[\sum_{i=1}^{2^N} a_{N,i}^2\big] \qquad \text{and} \qquad \delta_{N,i} := \tilde{\sigma}_{N,i}^2 - \sigma_{N,i}^2.$$

Note that

$$I_N = E_Q\big[\sum_{i=1}^{2^N}(\Delta_{N,i}A)^2\big] - J_N,$$

where

$$J_N := E_Q\big[\sum_{i=1}^{2^N}\tilde{\alpha}_{N,i}^2\big].$$

Since $Q \in \mathcal{Q}_{\mathcal{E}}$, we obtain

$$\lim_{N\uparrow\infty} I_N = E_Q[< A >_1],$$

Let us now show that the last term in (46) converges to 0 as $N \uparrow \infty$. Since

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

satisfies

$$|\delta_{N,i}| \leq \tilde{\alpha}_{N,i}^2 + 2\sigma_{N,i}\tilde{\alpha}_{N,i}, \qquad (47)$$

we get

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

hence

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

due to condition (42). Moreover, if $\sigma^2(\cdot,\cdot) \geq c \quad Q\otimes\lambda$-a.s. for some $c > 0$ then

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

and this implies

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

Thus, the last two terms in equation (46) converge to 0. In view of Lemma 21, this completes the proof. □

Remark 26. *The proof shows that, instead of requiring that $\sigma^2(\cdot,\cdot)$ is bounded away from 0, it is enough to assume that the conditional variances $\sigma_{N,i}^2$ and $\tilde{\alpha}_{N,i}^2$ of M and A under $Q \in \mathcal{Q}_{\mathcal{E}}$ satisfy the condition*

$$\lim_{N\uparrow\infty} E_Q\big[2^{-N}\sum_{i=1}^{2^N}\tilde{\alpha}_{N,i}^2\sigma_{N,i}^{-2}\big] = 0. \tag{50}$$

This includes the case of a martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$, and also the case where the process A is locally deterministic in the sense that $\tilde{\alpha}_{N,i}^2 = 0$ for large enough N.

Theorem 25 allows us to prove an extension of Talagrand's inequality on Wiener space beyond the absolutely continuous case $Q \ll P$. For $Q \in \mathcal{Q}_{\mathcal{S}}$ we refer to [9] for an extension that covers Talagrand's inequality (2) as a special case. Here we focus on the case $Q \in \mathcal{Q}_{\mathcal{D}}$ and consider the following Wasserstein distance $W_{\mathcal{D}}(Q,P)$, where the cost function is defined in terms of quadratic variation.

Definition 27. *The* Wasserstein distance $W_{\mathcal{D}}(Q,P)$ *between $Q \in \mathcal{Q}_{\mathcal{D}}$ and Wiener measure P is defined as*

$$W_{\mathcal{D}}(Q,P) = \inf\big(\tilde{E}\big[<\tilde{Y}-\tilde{X}>_1\big]\big)^{1/2}, \tag{51}$$

where the infimum is taken over all adaptive couplings $(\tilde{Y},\tilde{X})$ of Q and P on some filtered probability space $(\tilde{\Omega},\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{0\le t\le 1},\tilde{P})$ such that $\tilde{Y}$ is a Dirichlet process on $(\tilde{\Omega},\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{0\le t\le 1},\tilde{P})$.

For a martingale measure $Q \in \mathcal{Q}_{\mathcal{M}}$, the following corollary is proved in [9]. For $Q \in \mathcal{Q}_{\mathcal{D}}$, the proof is essentially the same, and so we just sketch the argument and refer to [9] for further details.

Corollary 28. *For a probability measure $Q \in \mathcal{Q}_{\mathcal{D}}$ that satisfies condition* (50)*,*

$$W_{\mathcal{D}}(Q,P) \le \sqrt{2\,h(Q|P)}. \tag{52}$$

Proof. We may assume $h(Q|P) < \infty$. As shown in [9], this implies that there is a Wiener process $W^Q = (W_t^Q)_{0\le t\le 1}$, defined on the filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{0\le t\le 1},Q)$, such that the coupling (W^Q,W) of P and Q is optimal for the Wasserstein distance $W_{\mathcal{D}}$, that is,

$$W_{\mathcal{D}}^2(Q,P) = E_Q\big[<W-W^Q>_1\big]. \tag{53}$$

Moreover,

$$E_Q\big[<W - W^Q>_1\big] = E_Q\Big[\int_0^1 \big(\sigma(\cdot,s)-1\big)^2 ds + q_s(\cdot,(0,1])\Big].$$

Since

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

we obtain

$$\begin{aligned} E_Q\big[<W - W^Q>_1\big] &\le 2\,E_Q\Big[\int_0^1 f\big(\sigma^2(\cdot,s)\big)dt + \frac{1}{2}q_s(\cdot,(0,1])\Big] \\ &\le 2\,h(Q|P), \end{aligned} \tag{54}$$

where the second inequality follows from Theorem 25, and so we have shown inequality (52). □

Acknowledgments: It is a great pleasure to contribute to this volume in honor of Masatoshi Fukushima's Beiju. His deep insight into the interplay between Markov processes and Dirichlet spaces has also inspired some of my own work, including the present paper. Since our first encounter during the 6th Berkeley Symposium, more than 50 years ago, our paths have crossed many times, including mutual visits in Hanover, Bonn, Osaka, Zürich and Berlin. I want to express my warmest thanks for the rewarding experience that each encounter has turned out to be, and my best wishes for the years to come.

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] J. Bertoin. Les Processus de Dirichlet en tant qu'Espaces de Banach. *Stochastics*, 18:155–188, 1986.

[3] F. Coquet, A. Jakubowski, J. Mémin, and L. Slominski. Natural decomposition of processes and weak dirichlet processes. In *In memoriam Paul-André Meyer: Séminaire de Probabilités XXXIX*, volume 1874 of *Lecture Notes in Math.*, pages 81–116. Springer, Berlin, 2006.

[4] C. Dellacherie and P.-A. Meyer. *Probabilités et Potentiel, Ch. V -VIII.* Hermann, Paris, 1980.

[5] 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.

[6] H. Föllmer. The exit measure of a supermartingale. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 21:154–166, 1972.

[7] H. Föllmer. Dirichlet processes. In D. Williams, editor, *Stochastic Integrals*, volume 851 of *Lecture Notes in Mathematics*, pages 476–478. Springer, 1981.

[8] 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.

[9] H. Föllmer. Optimal Coupling on Wiener Space and an Extension of Talagrand's Transport Inequality. In G. Yin and Th. Zariphopoulou, editors, *Stochastic Analysis, Stochastic Control, and Stochastic Optimization - A Commemorative Volume to Honor Professor Mark H. Davis's Contributions*. Springer, 2022.

[10] M. Fukushima. *Dirichlet Forms and Markov Processes*, volume 23 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, 1980.

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

[12] 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.

[13] F. Gozzi and F. Russo. Weak dirichlet processes with a stochastic control perspective. *Stochastic Processes and their Applications*, 116(11):1563–1583, 2006.

[14] S.E. Graversen and M. Rao. Quadratic variation and energy. *Nagoya Math. J.*, 100:163–180, 1985.

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

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

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

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