Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 3, 885–899
DOI: 10.1214/11-AIHP464

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Representation formula for the entropy and functional inequalities

Joseph Lehec

CEREMADE, UMR CNRS 7534, Université Paris-Dauphine, place du maréchal de Lattre de Tassigny, 75016 Paris, France

Received 11 June 2010; revised 9 September 2011; accepted 14 November 2011

Abstract. We prove a stochastic formula for the Gaussian relative entropy in the spirit of Borell's formula for the Laplace transform. As an application, we give simple proofs of a number of functional inequalities.

Résumé. On démontre une formule stochastique pour l'entropie relative par rapport à la Gaussienne, dans le genre de la formule de Borell pour la transformée de Laplace. Cette formule donne des preuves simples d'un certain nombre d'inégalités fonctionnelles.

MSC: 39B62; 60J65

Keywords: Gaussian measure; Entropy; Functional inequalities; Girsanov's formula

1. Introduction: Borell's formula

Let γ_d be the standard Gaussian measure on $\mathbb{R}^d$:

$$\gamma_d(\mathrm{d}x) = \frac{\mathrm{e}^{-|x|^2/2}}{(2\pi)^{d/2}}\,\mathrm{d}x,$$

where $|x| = \sqrt{x \cdot x}$ denotes the Euclidean norm of x. In [5] Borell proves the following representation formula. Given a standard d-dimensional Brownian motion B and a bounded function $f : \mathbb{R}^d \to \mathbb{R}$ we have

$$\log\left(\int_{\mathbb{R}^d} \mathrm{e}^f \,\mathrm{d}\gamma_d\right) = \sup_u \left[\mathsf{E}\left(f\left(B_1 + \int_0^1 u_s \,\mathrm{d}s\right) - \frac{1}{2}\int_0^1 |u_s|^2 \,\mathrm{d}s\right)\right], \tag{1}$$

where the supremum is taken over all random processes u, say bounded and adapted to the Brownian filtration. Among other applications, he derives easily the Prékopa–Leindler inequality. The name *Borell's formula* may be unfair to Boué and Dupuis who in an earlier paper [6] obtained a stronger result, allowing the function f to depend on the whole path $(B_t)_{t\in[0,1]}$ (see Theorem 9 below for a precise statement). Anyway, Borell and Boué–Dupuis agree that representation formulas such as (1) arose much earlier in optimal control theory, particularly in Fleming and Soner's work [13], and Borell should definitely be credited for bringing these techniques in the context of functional inequalities.

The present article deals with relative entropy. Let $(\Omega, \mathcal{A}, m)$ be a measured space and μ be a probability measure. The relative entropy of μ is defined by

$$\mathrm{H}(\mu|m) = \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}m}\log\left(\frac{\mathrm{d}\mu}{\mathrm{d}m}\right)\mathrm{d}m \quad \text{if } \mu \ll m$$

and $\mathrm{H}(\mu|m) = +\infty$ otherwise. It is well known that there is a Legendre duality between relative entropy and logarithmic Laplace transform:

$$\mathrm{H}(\mu|m) = \sup_f \left(\int f \, \mathrm{d}\mu - \log\left(\int_\Omega \mathrm{e}^f \, \mathrm{d}m \right) \right). \tag{2}$$

The purpose of this article is to prove a representation formula for the Gaussian relative entropy, both in $\mathbb{R}^d$ and in the Wiener space, providing the entropy counterparts of the results mentioned above. All these formulas have a common feature: Girsanov's theorem. However, our approach is somewhat different from that of Borell and Boué–Dupuis: it draws a connection with the work of Föllmer [14,15] which makes the whole argument arguably simpler. As an application, we give new, unified and simple proofs of a number of Gaussian inequalities.

2. Representation formula for the entropy

This section contains the main results of the article. Let us recall a couple of classical facts about relative entropy, see for instance [23], Section 10, and the references therein. If $\mathcal{A}$ is the Borel σ-field of a Polish topology on Ω then it is enough to take the supremum over bounded and continuous functions in (2). In particular the map $\mu \mapsto \mathrm{H}(\mu|m)$ is lower semicontinuous with respect to the topology of weak convergence of measures. If $T : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ is a measurable map then

$$\mathrm{H}\left(\mu \circ T^{-1} | m \circ T^{-1}\right) \le \mathrm{H}(\mu|m) \tag{3}$$

and assuming that $\mathrm{H}(\mu|m) < +\infty$, equality occurs if and only if the density $\mathrm{d}\mu/\mathrm{d}m$ is a function of T. We now describe the setting of the article. Let $\mathbb{W}$ be the space of continuous paths

$$\left\{ w \in \mathcal{C}^0\left(\mathbb{R}_+, \mathbb{R}^d\right), w_0 = 0 \right\}$$

equipped with the topology of uniform convergence on compact intervals. Let $\mathcal{B}$ be the associated Borel σ-field and let γ be the Wiener measure on $(\mathbb{W}, \mathcal{B})$. Let $x_t : w \mapsto w_t$ be the coordinate process and $(\mathcal{G}_t)_{t \ge 0}$ be the natural filtration of x. It is well known that $\mathcal{B}$ coincides with the smallest σ-field containing $\bigcup_{t \ge 0} \mathcal{G}_t$. Let $\mathbb{H}$ be the Cameron–Martin space: a path U belongs to $\mathbb{H}$ if there exists $u \in \mathrm{L}^2([0, +\infty); \mathbb{R}^d)$ such that

$$U_t = \int_0^t u_s \, \mathrm{d}s, \quad t \ge 0.$$

The norm of U in $\mathbb{H}$ is then defined by

$$\|U\| = \left(\int_0^{+\infty} |u_s|^2 \, \mathrm{d}s \right)^{1/2}.$$

The Cauchy–Schwarz inequality shows that the Hilbert space $\mathbb{H}$ embeds continuously in $\mathbb{W}$. Given a probability space $(\Omega, \mathcal{A}, \mathsf{P})$ equipped with a filtration $(\mathcal{F}_t)_{t \ge 0}$ we call *drift* any adapted process U which belongs to $\mathbb{H}$ almost surely. Lastly, our Brownian motions are always d-dimensional, standard and always start from 0.

2.1. *The upper bound*

We shall use repeatedly Girsanov's formula, see [18], Chapter 6.

Proposition 1. *Let B be a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{A}, \mathsf{P}, \mathcal{F})$ and let U be a drift. Letting μ be the law of $B + U$, we have*

$$\mathrm{H}(\mu|\gamma) \le \frac{1}{2} \mathsf{E}\|U\|^2. \tag{4}$$

Proof. Write $U_t = \int_0^t u_s \, \mathrm{d}s$ and assume for the moment that $\|U\|^2 = \int_0^\infty |u_s|^2 \, \mathrm{d}s$ is uniformly bounded. Then by Novikov's criterion

$$M_t = \exp\left(- \int_0^t u_s \cdot \mathrm{d}B_s - \frac{1}{2} \int_0^t |u_s|^2 \, \mathrm{d}s \right), \quad t \geq 0,$$

is a uniformly integrable martingale and Girsanov's formula applies. Under

$$\mathrm{dQ} = M_\infty \, \mathrm{dP}$$

the process $X := B + U$ is a Brownian motion. Therefore X has law μ and γ under P and Q, respectively. Then by (3)

$$\mathrm{H}(\mu|\gamma) \leq \mathrm{H}(\mathsf{P}|\mathsf{Q}) = -\mathsf{E}\log(M_\infty) = \frac{1}{2}\mathsf{E}\|U\|^2,$$

which concludes the proof when $\|U\|$ is bounded. In the general case, define the stopping time

$$T_n = \inf\left(t \geq 0, \int_0^t |u_s|^2 \, \mathrm{d}s \geq n \right),$$

let U_n be the stopped process $(U_n)_t = U_{t \wedge T_n}$ and μ_n be the law of $B + U_n$. With probability 1 we have $\|U\|^2 < +\infty$, thus $T_n \to +\infty$ and $U_n \to U$ in $\mathbb{H}$, hence in $\mathbb{W}$. Therefore $\mu_n \to \mu$ weakly. Also $\mathsf{E}\|U_n\|^2 \to \mathsf{E}\|U\|^2$ by monotone convergence. Thus, using the lower semicontinuity of the entropy (observe that $\mathbb{W}$ is a Polish space)

$$\begin{aligned} \mathrm{H}(\mu|\gamma) &\leq \liminf_n \mathrm{H}(\mu_n|\gamma) \\ &\leq \liminf_n \frac{1}{2}\mathsf{E}\|U_n\|^2 = \frac{1}{2}\mathsf{E}\|U\|^2. \end{aligned}$$

□

Remark. *It follows immediately that when $\mathsf{E}\|U\|^2 < +\infty$, the law of $B + U$ is absolutely continuous with respect to the Wiener measure γ. Let us point out that this is actually true for all drifts U, even if $\mathsf{E}\|U\|^2 = +\infty$, see* [18], *Chapter* 7.

2.2. *Föllmer's drift*

Let us address the question whether, given a probability measure μ on $\mathbb{W}$, equality can be achieved in (4). Recall that $(x_t)_{t \geq 0}$ is the coordinate process on Wiener space $(\mathbb{W}, \mathcal{B}, \gamma)$ and that $(\mathcal{G}_t)_{t \geq 0}$ is its natural filtration. The following is due to Föllmer [14,15].

Theorem 2. *Let μ be a measure on $(\mathbb{W}, \mathcal{B})$ having density F with respect to γ. There exists an adapted process u such that under μ the following holds.*

1. *The process $U_t = \int_0^t u_s \, \mathrm{d}s$ belongs to $\mathbb{H}$ almost surely.*
2. *The process $y = x - U$ is a Brownian motion.*
3. *The relative entropy of μ is*

$$\mathrm{H}(\mu|\gamma) = \frac{1}{2}\mathsf{E}^\mu \|U\|^2.$$

We sketch the proof for completeness.

Proof of Theorem 2. Throughout E^γ and E^μ denote expectations with respect to γ and μ respectively. On $\mathcal{G}_t$ the measure μ has density

$$F_t := \mathsf{E}^\gamma(F|\mathcal{G}_t),$$

with respect to γ. A standard martingale argument shows that

$$\mu\Big(\inf_{t\geq 0} F_t > 0\Big) = \mu(F > 0) = 1. \tag{5}$$

Since Brownian martingales can be represented as stochastic integrals there exists an adapted process v satisfying

$$\gamma\left(\int_0^{+\infty} |v_s|^2 \,\mathrm{d}s < +\infty\right) = 1 \tag{6}$$

and

$$F_t = 1 + \int_0^t v_s \cdot \mathrm{d}x_s, \quad t \geq 0.$$

Let u be the process defined by

$$u_t = \mathbf{1}_{\{F_t > 0\}} (F_t)^{-1} v_t.$$

It is adapted and (5) and (6) yield

$$\mu\left(\int_0^{\infty} |u_s|^2 \,\mathrm{d}s < +\infty\right) = 1,$$

which is the first assertion of the theorem.

The assertion 2 follows from Girsanov's formula, see [18], Theorem 6.2.

Under μ, we have

$$\begin{aligned} F_t &= 1 + \int_0^t F_s u_s \cdot \mathrm{d}x_s \\ &= 1 + \int_0^t F_s u_s \cdot \mathrm{d}y_s + \int_0^t F_s |u_s|^2 \,\mathrm{d}s. \end{aligned}$$

Applying Itô's formula (recall that F is positive and y is a Brownian motion under μ) we obtain

$$\log(F) = \int_0^{+\infty} u_s \cdot \mathrm{d}y_s + \frac{1}{2}\int_0^{+\infty} |u_s|^2 \,\mathrm{d}s.$$

If $\mathsf{E}^{\mu}\|U\|^2 < +\infty$ the local martingale part in the equation above is integrable and has mean 0 so that

$$\mathrm{H}(\mu|\gamma) = \mathsf{E}^{\mu} \log(F) = \frac{1}{2}\mathsf{E}^{\mu}\|U\|^2.$$

Again, a localization argument shows that this equality remains valid when $\mathsf{E}^{\mu}\|U\|^2 = +\infty$, see [14], Lemma 2.6. □

To finish this subsection, we give a formula for Föllmer's drift when the underlying density has a Malliavin derivative, we refer to the first chapter of [19] for the (little amount of) Malliavin calculus we shall use. For suitable $F : \mathbb{W} \to \mathbb{R}$ we let $\mathrm{D}F : \mathbb{W} \to \mathbb{H}$ be the Malliavin derivative of F. The domain of D in the space $\mathrm{L}^2(\mathbb{W}, \mathcal{B}, \gamma)$ is denoted by $\mathbb{D}^2$. If $F \in \mathbb{D}^2$ then the Clark–Ocone formula asserts

$$\mathsf{E}^{\gamma}(F|\mathcal{G}_t) = 1 + \int_0^t \mathsf{E}^{\gamma}(\mathrm{D}_s F|\mathcal{G}_s) \cdot \mathrm{d}x_s, \quad t \geq 0.$$

We obtain the following result.

Lemma 3. *When $F \in \mathbb{D}^2$ the process u_t given by Theorem 2 is*

$$u_t = \frac{\mathsf{E}^\gamma(\mathrm{D}_t F|\mathcal{G}_t)}{\mathsf{E}^\gamma(F|\mathcal{G}_t)}\mathbf{1}_{\{\mathsf{E}^\gamma(F|\mathcal{G}_t)>0\}}.$$

This implies that μ-almost surely

$$u_t = \mathsf{E}^\mu\left(\frac{\mathrm{D}_t F}{F}\Big|\mathcal{G}_t\right).$$

2.3. *Optimal drift in a strong sense*

According to Theorem 2, the filtered probability space $(\mathbb{W}, \mathcal{B}, \mu, \mathcal{G})$ carries a Brownian motion y. The process $x = y + U$ has law μ and the drift U satisfies

$$\mathrm{H}(\mu|\gamma) = \frac{1}{2}\mathsf{E}^\mu\|U\|^2.$$

Still, it remains open whether *given* a probability space, a filtration and a Brownian motion, there exists a drift achieving equality in (4).

It this section, we show that this is indeed the case, under some restriction on the measure μ. The approach is taken from the article [4] in which Baudoin treats the case of Brownian bridges (see Section 2.5 below). We refer to [20] for the background on stochastic differential equations.

Theorem 4. *Let B be a Brownian motion defined on some filtered probability space $(\Omega, \mathcal{A}, \mathsf{P}, \mathcal{F})$. Let μ be a measure on $\mathbb{W}$, absolutely continuous with respect to γ and let $u_t : \mathbb{W} \to \mathbb{R}^d$ be the associated Föllmer process. If the stochastic differential equation*

$$X_t = B_t + \int_0^t u_s(X)\,\mathrm{d}s, \quad t \geq 0, \tag{7}$$

has the pathwise uniqueness property, then it has a unique strong solution. This solution X satisfies the following.

1. *The process $U_t = \int_0^t u_s(X)\,\mathrm{d}s$ belongs to $\mathbb{H}$ almost surely.*
2. *The process X has law μ.*
3. *The relative entropy of μ is given by*

$$\mathrm{H}(\mu|\gamma) = \frac{1}{2}\mathsf{E}\|U\|^2.$$

Proof. According to Theorem 2, on $(\mathbb{W}, \mathcal{B}, \mu)$ the coordinate process x satisfies

$$x_t = y_t + \int_0^t u_s(x)\,\mathrm{d}s,$$

where y is a Brownian motion. Therefore (7) has a weak solution. By Yamada and Watanabe's theorem, if pathwise uniqueness holds then (7) has a unique strong solution. Moreover, since pathwise uniqueness implies uniqueness in law, the solution X has law μ. The rest of Theorem 4 concerns the law of X, so it is contained in Theorem 2. □

We end this section by showing that for a reasonably large class of measures μ, the stochastic differential equation (7) does satisfy the pathwise uniqueness property.

Definition 5. *Let $\mathcal{S}$ be the class of probability measures on $(\mathbb{W}, \mathcal{B}, \gamma)$ having a density of the form*

$$F(w) = \Phi(w_{t_1}, \dots, w_{t_n}) \tag{8}$$

for some integer n, for some sample $0 \leq t_1 < t_2 < \cdots < t_n$ and for some function $\Phi : (\mathbb{R}^d)^n \to \mathbb{R}$ satisfying

- *Φ is Lipschitz.*
- *$\nabla\Phi$ is Lipschitz.*
- *There exists $\varepsilon > 0$ such that $\Phi \geq \varepsilon$.*

Lemma 6. *If μ belongs to $\mathcal{S}$ then the equation (7) has the pathwise uniqueness property.*

Proof. Let μ have density F given by (8). Then $F \in \mathbb{D}^2$ and

$$\mathrm{D}F(w) = \sum_{i=1}^{n} \nabla_i \Phi(w_{t_1}, \dots, w_{t_n}) \mathbf{1}_{[0,t_i]},$$

where $\nabla_i \Phi$ is the gradient of Φ in the ith variable. By Lemma 3, the process associated to μ is

$$\begin{aligned} u_t(w) &= \frac{\mathsf{E}^\gamma(\mathrm{D}_t F(w)|\mathcal{G}_t)}{\mathsf{E}^\gamma(F(w)|\mathcal{G}_t)} \\ &= \sum_{i=1}^{n} \frac{\mathsf{E}^\gamma(\nabla_i \Phi(w_{t_1}, \dots, w_{t_n})|\mathcal{G}_t)}{\mathsf{E}^\gamma(\Phi(w_{t_1}, \dots, w_{t_n})|\mathcal{G}_t)} \mathbf{1}_{[0,t_i]}(t). \end{aligned}$$

It is enough to prove that there is a constant C such that

$$\left|u_t(w) - u_t(\tilde{w})\right| \leq C \sup_{0 \leq s \leq t} |w_s - \tilde{w}_s| \tag{9}$$

for all $t \geq 0$ and for all $w, \tilde{w} \in \mathbb{W}$. Fix $t \geq 0$ and assume that $t_k \leq t < t_{k+1}$ for some $k \in \{0, \dots, n-1\}$. By the Markov property of the Brownian motion

$$\mathsf{E}\big(\Phi(w_{t_1}, \dots, w_{t_n})|\mathcal{G}_t\big) = \Psi(w_{t_1}, \dots, w_{t_k}, w_t),$$

where $\Psi(x_1, \dots, x_k, x)$ equals

$$\int_{\mathbb{W}} \Phi(x_1, \dots, x_k, x + w_{t_{k+1}-t}, \dots, x + w_{t_n - t}) \gamma(\mathrm{d}w).$$

Then observe that $\|\Psi\|_{\mathrm{lip}} \leq \|\Phi\|_{\mathrm{lip}}$. We have a similar property when $0 \leq t < t_1$ and when $t_n \leq t$. The argument applies also to $\nabla_i \Phi$. The inequality (9) follows easily. $\square$

To sum up, we have the following representation formula.

Theorem 7. *Let $(\Omega, \mathcal{A}, \mathsf{P}, \mathcal{F})$ be a filtered probability space and let $B : \Omega \to \mathbb{W}$ be a Brownian motion. For all $\mu \in \mathcal{S}$ we have*

$$\mathrm{H}(\mu|\gamma) = \min_U \left(\frac{1}{2}\mathsf{E}\|U\|^2\right),$$

where the minimum is on all drifts U such that $B + U$ has law μ.

2.4. *The Boué and Dupuis formula*

In this subsection the previous results are translated in terms of log-Laplace using the following lemma.

Lemma 8. *Let $f : \mathbb{W} \to \mathbb{R}$ bounded from below. For every positive ε there exists $\mu \in \mathcal{S}$ such that*

$$\log\left(\int_{\mathbb{W}} \mathrm{e}^f \, \mathrm{d}\gamma\right) \leq \int_{\mathbb{W}} f \, \mathrm{d}\mu - \mathrm{H}(\mu|\gamma) + \varepsilon. \tag{10}$$

Proof. By monotone convergence we can assume that f is also bounded from above, and that $\int e^f \, d\gamma = 1$. Set $F = e^f$ and let μ be a probability measure on $\mathbb{W}$ having density G with respect to γ. Then

$$\mathrm{H}(\mu|\gamma) - \int f \, d\mu = \int \frac{G}{F} \log\left(\frac{G}{F}\right) F \, d\gamma.$$

Using $t\log(t) \le |t-1| + |t-1|^2/2$ we get

$$\begin{aligned}\mathrm{H}(\mu|\gamma) - \int f \, d\mu &\le \int \left|\frac{G}{F} - 1\right| F \, d\gamma + \frac{1}{2}\int \left|\frac{G}{F} - 1\right|^2 F \, d\gamma \\ &\le \|F - G\|_{\mathrm{L}^1(\gamma)} + C\|F - G\|^2_{\mathrm{L}^2(\gamma)}\end{aligned}$$

for some constant C (recall that f is bounded below). Therefore, it is enough to show that there exists $\mu \in \mathcal{S}$ whose density G is arbitrarily close to F in $\mathrm{L}^2(\gamma)$. This is a standard fact. □

Here is the Boué and Dupuis formula.

Theorem 9. *For every function $f : \mathbb{W} \to \mathbb{R}$ measurable and bounded from below, we have*

$$\log\left(\int_{\mathbb{W}} e^f \, d\gamma\right) = \sup_U \left[\mathsf{E}\left(f(B+U) - \frac{1}{2}\|U\|^2\right)\right],$$

where the supremum is taken over all drifts U.

This is actually slightly more general than the result in [6], which concerns the space $\mathcal{C}([0,T],\mathbb{R}^d)$ for some finite time horizon T.

Proof of Theorem 9. Let U be a drift and μ be the law of $B+U$. By Proposition 1 and the entropy/log-Lapace duality

$$\mathsf{E}\left(f(B+U) - \frac{1}{2}\|U\|^2\right) \le \int f \, d\mu - \mathrm{H}(\mu|\gamma) \le \log\left(\int_{\mathbb{W}} e^f \, d\gamma\right).$$

On the other hand, given $\varepsilon > 0$, there exists a probability measure $\mu \in \mathcal{S}$ satisfying (10). Since $\mu \in \mathcal{S}$, Theorem 7 asserts that there exists a drift U such that $B+U$ has law μ and satisfying

$$\mathrm{H}(\mu|\gamma) = \frac{1}{2}\mathsf{E}\|U\|^2.$$

Then (10) becomes

$$\log\left(\int_{\mathbb{W}} e^f \, d\gamma\right) \le \mathsf{E}\left(f(B+U) - \frac{1}{2}\|U\|^2\right) + \varepsilon,$$

which concludes the proof. □

2.5. *Brownian bridges*

A measure μ on $\mathbb{W}$ satisfying

$$\mu(dw) = \rho(w_1)\gamma(dw), \tag{11}$$

where ρ is some density on $(\mathbb{R}^d, \gamma_d)$ is said to be a Brownian bridge. It can be seen as the law of a Brownian motion conditioned to have law $\rho(x)\gamma_d(dx)$ at time 1.

Lemma 10. *Let ν have density ρ with respect to γ_d, we have*

$$\mathrm{H}(\nu|\gamma_d) = \inf_{\mu}\big(\mathrm{H}(\mu|\gamma)\big),$$

where the infimum is on all probability measures satisfying $\mu \circ (x_1)^{-1} = \nu$. The infimum is attained when μ is the bridge (11).

In other words, among all processes having law ν at time 1, the bridge minimizes the relative entropy. This is essentially a particular case of (3), see also [4] and [16], p. 161.

Assume that ρ is differentiable and that $\nabla\rho \in \mathrm{L}^2(\gamma_d)$. Then $F(w) = \rho(w_1)$ belongs to $\mathbb{D}^2$ and has Malliavin derivative

$$\mathrm{D}F(w) = \nabla\rho(w_1)\mathbf{1}_{[0,1]}.$$

By Lemma 3 the Föllmer process of the bridge μ is such that

$$u_t = \mathsf{E}^{\mu}\big(\nabla\log(\rho)(w_1)|\mathcal{G}_t\big)\mathbf{1}_{[0,1]}(t), \quad \mu\text{-a.s.}$$

We obtain the following result.

Lemma 11. *Under μ, the process $(u_t)_{t\in[0,1]}$ is a martingale. In particular*

$$\mathsf{E}^{\mu}(u_t) = \mathsf{E}^{\mu}\nabla\log(\rho)(w_1) = \mathsf{E}^{\gamma}\nabla\rho(w_1) = \int_{\mathbb{R}^d} x\,\nu(\mathrm{d}x).$$

Now assume that ρ and $\nabla\rho$ are Lipschitz and that $\rho \geq \varepsilon$, so that the bridge μ belongs to $\mathcal{S}$. It is easily seen that u_t can also be written as

$$u_t(w) = \nabla\log P_{1-t}\rho(w_t)\mathbf{1}_{[0,1]}(t),$$

where P_t denotes the heat semigroup on $\mathbb{R}^d$:

$$\partial_t P_t = \frac{1}{2}\Delta P_t.$$

The stochastic differential equation (7) becomes

$$X_t = B_t + \int_0^{t\wedge 1} \nabla\log(P_{1-s}\rho)(X_s)\,\mathrm{d}s, \quad t \geq 0. \tag{12}$$

By Lemma 6, there is a unique strong solution. Combining Lemma 10 with Theorem 4 we obtain the following dual formulation of Borell's result (1).

Theorem 12. *Let ν and ρ be as above. Then*

$$\mathrm{H}(\nu|\gamma_d) = \inf_{U}\left(\frac{1}{2}\mathsf{E}\|U\|^2\right),$$

where the infimum is taken on all drifts U satisfying $B_1 + U_1 = \nu$ in law. The infimum is attained by the drift

$$U_t = \int_0^{t\wedge 1} \nabla\log(P_{1-s}\rho)(X_s)\,\mathrm{d}s,$$

where X is the unique solution of (12).

3. Applications

Following Borell, we now derive functional inequalities from the representation formula. Let us point out that in all but one applications we use Proposition 1 and Theorem 2 rather than Theorem 7.

3.1. *Transportation cost inequality*

Let T_2 be the transportation cost for the Euclidean distance squared: given two probability measures μ and ν on $\mathbb{R}^d$

$$\mathrm{T}_2(\mu,\nu)=\inf\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\,\mathrm{d}\pi(x,y)\right),\tag{13}$$

where the infimum is taken over all couplings π of μ and ν, namely all probability measures on the product space $\mathbb{R}^d\times\mathbb{R}^d$ having marginals μ and ν. There is a huge literature about this optimization problem, usually referred to as Monge–Kantorovitch problem, see Villani's book [24]. Talagrand's inequality [22] asserts that

$$\mathrm{T}_2(\nu,\gamma_d)\le 2\mathrm{H}(\nu|\gamma_d)$$

for every probability measure ν on $\mathbb{R}^d$. The purpose of this subsection is to prove a Wiener space version of this inequality.

On Wiener space the natural definition of T_2 involves the norm of the Cameron–Martin space $\mathbb{H}$: given two probability measures μ,ν on $(\mathbb{W},\mathcal{B})$

$$\mathrm{T}_2(\mu,\nu)=\inf\left(\int_{\mathbb{W}\times\mathbb{W}}\left\|w-w'\right\|^2\pi\left(\mathrm{d}w,\mathrm{d}w'\right)\right),$$

where the infimum is taken over all couplings π of μ and ν such that $w-w'\in\mathbb{H}$ for π-almost all (w,w').

Theorem 13. *Let μ be a probability measure on* $(\mathbb{W},\mathcal{B})$. *Then*

$$\mathrm{T}_2(\mu,\gamma)\le 2\mathrm{H}(\mu|\gamma).$$

Here is a short proof based of Theorem 2. Fair enough, Feyel and Üstünel [12] have a very similar argument.

Proof of Theorem 13. Assume that μ is absolutely continuous with respect to γ (otherwise $\mathrm{H}(\mu|\gamma)=+\infty$). According to Theorem 2 there exists a Brownian motion B and a drift U such that $B+U$ has law μ and

$$\mathrm{H}(\mu|\gamma)=\frac{1}{2}\mathsf{E}\|U\|^2.$$

Then $(B,B+U)$ is a coupling of (γ,μ) and by definition of T_2

$$\mathrm{T}_2(\mu,\gamma)\le\mathsf{E}\|U\|^2=2\mathrm{H}(\mu|\gamma).\qquad\square$$

Let us point out that Talagrand's inequality can be recovered easily from this theorem, applying it to a Brownian bridge. Details are left to the reader.

3.2. *Logarithmic Sobolev inequality*

In this section we prove the logarithmic Sobolev inequality for the Wiener measure, which extends the classical log-Sobolev inequality for the Gaussian measure, due to Gross [17]. When μ is a measure on $(\mathbb{W},\mathcal{B},\gamma)$ with density F such that $\mathrm{D}F$ is well defined, the Fisher information of μ is

$$\mathrm{I}(\mu|\gamma)=\int_{\mathbb{W}}\frac{\|\mathrm{D}F\|^2}{F}\,\mathrm{d}\gamma=\int_{\mathbb{W}}\left\|\frac{\mathrm{D}F}{F}\right\|^2\mathrm{d}\mu.$$

Theorem 14. *Let μ have density F with respect to γ and assume that $F \in \mathbb{D}^2$. Then*

$$\mathrm{H}(\mu|\gamma) \leq \frac{1}{2}\mathrm{I}(\mu|\gamma). \tag{14}$$

Proof. We consider the probability space $(\mathbb{W}, \mathcal{B}, \mu)$. Recall that $(\mathcal{G}_t)_{t\geq 0}$ is the filtration of the coordinate process. By Theorem 2 and Lemma 3, letting

$$u_t = \mathsf{E}^\mu\left(\frac{\mathrm{D}_t F}{F}\middle|\mathcal{G}_t\right)$$

we have

$$\mathrm{H}(\mu|\gamma) = \frac{1}{2}\mathsf{E}^\mu\int_0^\infty |u_t|^2\,\mathrm{d}t.$$

By Jensen's inequality

$$\mathsf{E}^\mu|u_t|^2 \leq \mathsf{E}^\mu\left|\frac{\mathrm{D}_t F}{F}\right|^2$$

so that

$$\mathrm{H}(\mu|\gamma) \leq \frac{1}{2}\mathsf{E}^\mu\left\|\frac{\mathrm{D}F}{F}\right\|^2$$

which is the result. □

This may not be the most straightforward proof, see [8]. Let us emphasize that applying (14) to a Brownian bridge yields the usual log-Sobolev inequality. More precisely, let ν be a probability measure on $\mathbb{R}^d$ having a smooth density ρ with respect to γ_d and let μ be the measure on $\mathbb{W}$ given by

$$\mu(\mathrm{d}w) = \rho(w_1)\gamma(\mathrm{d}w).$$

Then $\mathrm{H}(\nu|\gamma_d) = \mathrm{H}(\mu|\gamma)$. On the other hand letting $F(w) = \rho(w_1)$ we have

$$\mathrm{D}F(w) = \nabla\rho(w_1)\mathbf{1}_{[0,1]},$$

which implies easily that $\mathrm{I}(\nu|\gamma_d) = \mathrm{I}(\mu|\gamma)$. Thus (14) becomes

$$\mathrm{H}(\nu|\gamma_d) \leq \frac{1}{2}\mathrm{I}(\nu|\gamma_d).$$

3.3. *Shannon's inequality*

Given a random vector η on $\mathbb{R}^d$ having density ρ with respect to the Lebesgue measure, Shannon's entropy is defined as

$$\mathrm{S}(\eta) = -\int_{\mathbb{R}^d}\rho\log(\rho)\,\mathrm{d}x.$$

In other words $\mathrm{S}(\eta) = -\mathrm{H}(\nu|\lambda_d)$ where ν is the law of η and λ_d is the Lebesgue measure on $\mathbb{R}^d$.

Theorem 15. *Let η, ξ be independent random vectors on $\mathbb{R}^d$ and $\theta \in [0, \pi/2]$*

$$\mathrm{S}\big(\cos(\theta)\eta + \sin(\theta)\xi\big) \geq \cos(\theta)^2\mathrm{S}(\eta) + \sin(\theta)^2\mathrm{S}(\xi). \tag{15}$$

This inequality plays a central role in information theory, see [11] for an overview on the topic.

Proof of Theorem 15. Let ν_θ be the law of $\cos(\theta)\eta + \sin(\theta)\xi$. By Theorem 2, Lemmas 10 and 11 there exists a Brownian motion X and a drift U such that

- $X_1 + U_1$ has law ν_0.
- $\mathrm{H}(\nu_0|\gamma_d) = \mathsf{E}\|U\|^2/2$.
- $\mathsf{E}(U) = \mathsf{E}(\eta)\mathbf{1}_{[0,1]}$.

Similarly, there exists a Brownian motion Y and a drift V satisfying the corresponding properties for $\nu_{\pi/2}$. Besides, we can clearly assume that Y is independent of X. Then $\cos(\theta)X + \sin(\theta)Y$ is a Brownian motion and

$$\cos(\theta)X_1 + \sin(\theta)Y_1 + \cos(\theta)U_1 + \sin(\theta)V_1$$

has law ν_θ. By Proposition 1 and Lemma 10

$$\mathrm{H}(\nu_\theta|\gamma_d) \le \frac{1}{2}\mathsf{E}\left\|\cos(\theta)U + \sin(\theta)V\right\|^2.$$

Denoting the inner product in $\mathbb{H}$ by $\langle\cdot,\cdot\rangle$ we have

$$\mathsf{E}\langle U, V\rangle = \langle \mathsf{E}U, \mathsf{E}V\rangle = (\mathsf{E}\eta)\cdot(\mathsf{E}\xi),$$

so that

$$\begin{aligned}\mathrm{H}(\nu_\theta|\gamma_d) \le{}& \cos(\theta)^2\mathrm{H}(\nu_0|\gamma_d) + \sin(\theta)^2\mathrm{H}(\nu_{\pi/2}|\gamma_d)\\ &+ \cos(\theta)\sin(\theta)(\mathsf{E}\eta)\cdot(\mathsf{E}\xi).\end{aligned}$$

This is easily seen to be equivalent to (15). $\square$

3.4. *Brascamp–Lieb inequality*

Let us focus on a family of inequalities dating back to Brascamp and Lieb's article [7] on optimal constants in Young's inequality. Since then a number of nice alternate proofs have been discovered, see [3,9,10] and the survey article [1]. This subsection is inspired by the (unpublished) proof of Maurey relying on Borell's formula.

Let E be a Euclidean space, let $E_1,\dots,E_m$ be subspaces and for all i let P_i be the orthogonal projection with range E_i. The crucial hypothesis is the so-called frame condition: there exist $c_1,\dots,c_m$ in $\mathbb{R}_+$ such that

$$\sum_{i=1}^m c_i P_i = \mathrm{id}_E. \tag{16}$$

Let $x \in E$, we then have $|x|^2 = (\sum c_i P_i x)\cdot x$ and since P_i is an orthogonal projection

$$|x|^2 = \sum_{i=1}^m c_i |P_i x|^2. \tag{17}$$

From now on $\mathbb{W}$ denotes the space of continuous paths taking values in E and starting from 0 and γ denotes the Wiener measure on $\mathbb{W}$. The spaces $\mathbb{W}_i$ and measures γ_i are defined similarly.

Theorem 16. *Under the frame condition, for every probability measure μ on $\mathbb{W}$ we have*

$$\mathrm{H}(\mu|\gamma) \ge \sum_{i=1}^m c_i \mathrm{H}(\mu_i|\gamma_i),$$

where $\mu_i = \mu \circ P_i^{-1}$ is the push-forward of μ by the projection P_i.

Proof. According to Theorem 2 there exists a standard Brownian motion B on E and a drift U such that $B+U$ has law μ and

$$\mathrm{H}(\mu|\gamma)=\frac{1}{2}\mathsf{E}\|U\|^2.$$

Since P_i is an orthogonal projection, the process P_iB is a standard Brownian motion on E_i. Also P_iB+P_iU has law $\mu\circ P_i^{-1}=\mu_i$. By Proposition 1

$$\mathrm{H}(\mu_i|\gamma_i)\le\frac{1}{2}\mathsf{E}\|P_iU\|^2,\quad i=1,\dots,m.$$

On the other hand, the frame condition (17) implies easily that

$$\|U\|^2=\sum_{i=1}^n c_i\|P_iU\|^2$$

pointwise. Taking expectation yields the result. □

As observed by Carlen and Cordero [9], this super-additivity property of the relative entropy is equivalent to the following Brascamp–Lieb inequality.

Corollary 17. *Under the frame condition, given m functions $F_i:\mathbb{W}_i\to\mathbb{R}_+$, we have*

$$\int_{\mathbb{W}}\prod_{i=1}^m(F_i\circ P_i)^{c_i}\,\mathrm{d}\gamma\le\prod_{i=1}^m\left(\int_{\mathbb{W}_i}F_i\,\mathrm{d}\gamma_i\right)^{c_i}.$$

When the functions F_i depend only on the point w_1 rather than on the whole path w we recover the usual Brascamp–Lieb inequality for the Gaussian measure.

3.5. *Reversed Brascamp–Lieb inequality*

Again E is a Euclidean space and $E_1,\dots,E_m$ are subspaces satisfying the frame condition (16). Observe that if $x_1,\dots,x_m$ belong to $E_1,\dots,E_m$ respectively, then for any $y\in E$, the Cauchy–Schwarz inequality and (17) yield

$$\begin{aligned}\left(\sum_{i=1}^m c_ix_i\right)\cdot y&=\sum_{i=1}^m c_i(x_i\cdot P_iy)\\&\le\left(\sum_{i=1}^m c_i|x_i|^2\right)^{1/2}\left(\sum_{i=1}^m c_i|P_iy|^2\right)^{1/2}\\&=\left(\sum_{i=1}^m c_i|x_i|^2\right)^{1/2}|y|.\end{aligned}$$

Hence

$$\left|\sum_{i=1}^m c_ix_i\right|^2\le\sum_{i=1}^m c_i|x_i|^2.\tag{18}$$

Let $\mathcal{S}_i$ be the class of probability measures on E_i which satisfy the conditions of Definition 5, replacing $\mathbb{R}^d$ by E_i. Here is the reversed version of Theorem 16.

Theorem 18. *Given m probability measures $\mu_1, \dots, \mu_m$ belonging to $\mathcal{S}_1, \dots, \mathcal{S}_m$ respectively, there exist m processes $X_1, \dots, X_m$ (defined on the same probability space) such that*

1. *X_i has law μ_i for all $i = 1, \dots, m$.*
2. *Letting μ be the law of $\sum c_i X_i$ we have*

$$\mathrm{H}(\mu|\gamma) \le \sum_{i=1}^m c_i \mathrm{H}(\mu_i|\gamma_i).$$

Proof. Again let B be a standard Brownian motion on E. For $i = 1, \dots, m$, the process $P_i B$ is a standard Brownian motion on E_i. Since $\mu_i \in \mathcal{S}_i$ there exists a drift U_i such that the process $X_i = P_i B + U_i$ has law μ_i and

$$\mathrm{H}(\mu_i|\gamma_i) = \frac{1}{2}\mathsf{E}\|U_i\|^2.$$

Let $X = \sum c_i X_i$ and let μ be the law of X. Since $\sum c_i P_i$ is the identity of E

$$X = B + \sum_{i=1}^m c_i U_i.$$

By Proposition 1, we get

$$\mathrm{H}(\mu|\gamma) \le \frac{1}{2}\mathsf{E}\left\|\sum_{i=1}^m c_i U_i\right\|^2.$$

On the other hand (18) easily implies that

$$\left\|\sum_{i=1}^m c_i U_i\right\|^2 \le \sum_{i=1}^m c_i \|U_i\|^2,$$

pointwise. Taking expectation we get the result. □

This sub-additivity property of the entropy is a multi-marginal version of the *displacement convexity* property put forward by Sturm [21]. By duality, we obtain the following reversed Brascamp–Lieb inequality.

Corollary 19. *Assuming the frame condition, given m functions $F_i : \mathbb{W}_i \to \mathbb{R}_+$ bounded away from 0, and a function $G : \mathbb{W} \to \mathbb{R}_+$ satisfying*

$$\prod_{i=1}^m F_i(w_i)^{c_i} \le G\left(\sum_{i=1}^m c_i w_i\right) \tag{19}$$

for all $(w_1, \dots, w_m) \in \mathbb{W}_1 \times \cdots \times \mathbb{W}_m$, we have

$$\prod_{i=1}^m \left(\int_{\mathbb{W}_i} F_i \,\mathrm{d}\gamma_i\right)^{c_i} \le \int_{\mathbb{W}} G \,\mathrm{d}\gamma.$$

Proof. By Lemma 8, for every i, there exists a measure $\mu_i \in \mathcal{S}_i$ such that

$$\log\left(\int_{\mathbb{W}_i} F_i \,\mathrm{d}\gamma_i\right) \le \int_{\mathbb{W}_i} \log(F_i)\,\mathrm{d}\mu_i - \mathrm{H}(\mu_i|\gamma_i) + \varepsilon.$$

Let $X_1, \dots, X_m$ be the random processes given by the previous theorem, let $X = \sum c_i X_i$ and let μ be the law of X. Then by duality and the hypothesis (19) we get

$$\log\left(\int_{\mathbb{W}} G \, \mathrm{d}\gamma\right) \geq \mathsf{E} \log(G)(X) - \mathrm{H}(\mu|\gamma)$$
$$\geq \mathsf{E}\left(\sum_{i=1}^{m} c_i \log(F_i)(X_i)\right) - \mathrm{H}(\mu|\gamma).$$

Since $\mathrm{H}(\mu|\gamma) \leq \sum c_i \mathrm{H}(\mu_i|\gamma_i)$, this is at least

$$\sum c_i \left(\log\left(\int_{\mathbb{W}_i} F_i \, \mathrm{d}\gamma_i\right) - \varepsilon\right).$$

Letting ε tend to 0 yields the result. □

Again when the functions depend only on the value of the path at time 1, we recover the reversed Brascamp–Lieb inequality for the Gaussian measure, which is due to Barthe [2].

Acknowledgements

The author is grateful to Patrick Cattiaux and Massimiliano Gubinelli for communicating references and to Christian Léonard, Bernard Maurey and Patrick Cattiaux again for valuable discussions.

References

[1] K. Ball. Convex geometry and functional analysis. In *Handbook of the Geometry of Banach Spaces*, Vol. 1 161–194. W. B. Johnson and J. Lindenstrauss (Eds). North-Holland, Amsterdam, 2001. MR1863692
[2] F. Barthe. On a reverse form of the Brascamp–Lieb inequality. *Invent. Math.* **134** (1998) 335–361. MR1650312
[3] F. Barthe and N. Huet. On Gaussian Brunn–Minkowski inequalities. *Studia Math.* **191** (2009) 283–304. MR2481898
[4] F. Baudoin. Conditioned stochastic differential equations: Theory, examples and application to finance. *Stochastic Process. Appl.* **100** (2002) 109–145. MR1919610
[5] C. Borell. Diffusion equations and geometric inequalities. *Potential Anal.* **12** (2000) 49–71. MR1745333
[6] M. Boué and P. Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.* **26** (1998) 1641–1659. MR1675051
[7] H. J. Brascamp and E. H. Lieb. Best constants in Young's inequality, its converse and its generalization to more than three functions. *Adv. Math.* **20** (1976) 151–173. MR0412366
[8] M. Capitaine, E. P. Hsu and M. Ledoux. Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces. *Electron. Commun. Probab.* **2** (1997) 71–81. MR1484557
[9] E. Carlen and D. Cordero-Erausquin. Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geom. Funct. Anal.* **19** (2009) 373–405. MR2545242
[10] D. Cordero-Erausquin and M. Ledoux. The geometry of Euclidean convolution inequalities and entropy. *Proc. Amer. Math. Soc.* **138** (2010) 2755–2769. MR2644890
[11] A. Dembo, T. M. Cover and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Inform. Theory* **37** (1991) 1501–1518. MR1134291
[12] D. Feyel and A. S. Üstünel. Measure transport on Wiener space and the Girsanov theorem. *C. R. Math. Acad. Sci. Paris* **334** (2002) 1025–1028. MR1913729
[13] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions*, 2nd edition. *Stochastic Modelling and Applied Probability* **25**. Springer, New York, 2006. MR2179357
[14] H. Föllmer. An entropy approach to the time reversal of diffusion processes. In *Stochastic Differential Systems (Marseille-Luminy, 1984)* 156–163. *Lecture Notes in Control and Inform. Sci.* **69**. Springer, Berlin, 1985. MR0798318
[15] H. Föllmer. Time reversal on Wiener space. In *Stochastic Processes – Mathematics and Physics (Bielefeld, 1984)* 119–129. *Lecture Notes in Math.* **1158**. Springer, Berlin, 1986. MR0838561
[16] H. Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87* 101–203. *Lecture Notes in Math.* **1362**. Springer, Berlin, 1988. MR0983373
[17] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** (1975) 1061–1083. MR0420249
[18] R. S. Liptser and A. N. Shiryayev. *Statistics of Random Processes, Vol. 1: General Theory*. *Applications of Mathematics* **5**. Springer, New York, 1977. MR0474486

[19] D. Nualart. *The Malliavin Calculus and Related Topics*, 2nd edition. *Probability and Its Applications*. Springer, Berlin, 2006. MR2200233
[20] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales, Vol. 2: Itô Calculus*. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge, 2000. MR1780932
[21] K. T. Sturm. On the geometry of metric measure spaces. I. *Acta Math*. **196** (2006) 65–131. MR2237206
[22] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal*. **6** (1996) 587–600. MR1392331
[23] S. R. S. Varadhan. *Large Deviations and Applications*. *CBMS–NSF Regional Conference Series in Applied Mathematics* **46**. SIAM, Philadelphia, 1984. MR0758258
[24] C. Villani. *Optimal Transport: Old and New*. *Grundlehren der Mathematischen Wissenschaften* **338**. Springer, Berlin, 2009. MR2459454