ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE Mathématiques

Joseph Lehec

Regularization in L_1 for the Ornstein-Uhlenbeck semigroup

Tome XXV, nº 1 (2016), p. 191-204.

<http://afst.cedram.org/item?id=AFST_2016_6_25_1_191_0>

Annales de la Faculté des Sciences de Toulouse Vol. XXV, n° 1, 2016
pp. 191-204

Regularization in L_1 for the Ornstein-Uhlenbeck semigroup

Joseph Lehec[(1)]

Résumé. — Soit γ_n la mesure Gaussienne standard sur $\mathbb{R}^n$ et soit (Q_t) le semi-groupe d'Ornstein-Uhlenbeck. Eldan et Lee ont montré récemment que pour toute fonction positive f d'intégrale 1 et pour temps t la queue de distribution de $Q_t f$ vérifie

$$\gamma_n(\{Q_t f > r\}) \leqslant C_t \, \frac{(\log\log r)^4}{r\sqrt{\log r}}, \quad \forall r > 1$$

où C_t est une constante dépendant seulement de t et pas de la dimension. L'objet de cet article est de simplifier en partie leur démonstration et d'éliminer le facteur $(\log\log r)^4$.

Abstract. — Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$ and let (Q_t) be the Ornstein-Uhlenbeck semigroup. Eldan and Lee recently established that for every non-negative function f of integral 1 and any time t the following tail inequality holds true:

$$\gamma_n(\{Q_t f > r\}) \leqslant C_t \, \frac{(\log\log r)^4}{r\sqrt{\log r}}, \quad \forall r > 1$$

where C_t is a constant depending on t but not on the dimension. The purpose of the present paper is to simplify parts of their argument and to remove the $(\log\log r)^4$ factor.

1. Introduction

Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$ and let (Q_t) be the Ornstein-Uhlenbeck semigroup: for every test function f

$$Q_t f(x) = \int_{\mathbb{R}^n} f\left(\mathrm{e}^{-t}x + \sqrt{1-\mathrm{e}^{-2t}}\, y\right) \gamma_n(dy). \tag{1.1}$$

(*) Reçu le 15/05/2015, accepté le 22/07/2015

(1) CEREMADE (UMR CNRS 7534) Université Paris–Dauphine.

Article proposé par Fabrice Baudoin.

Nelson [6] established that if $p > 1$ and $t > 0$ then Q_t is a contraction from $L_p(\gamma_n)$ to $L_q(\gamma_n)$ for some $q > p$, namely for

$$q = 1 + \mathrm{e}^{2t}(p-1).$$

The semigroup (Q_t) is said to be *hypercontractive*. This turns out to be equivalent to the logarithmic Sobolev inequality (see the classical article by Gross [3]). In this paper we establish a regularity property of $Q_t f$ assuming only that f is in $L^1(\gamma_n)$.

Let f be a non-negative function satisfying

$$\int_{\mathbb{R}^n} f \, d\gamma_n = 1,$$

and let $t > 0$. Since $Q_t f \geqslant 0$ and

$$\int_{\mathbb{R}^n} Q_t f \, d\gamma_n = \int_{\mathbb{R}^n} f \, d\gamma_n = 1,$$

Markov inequality gives

$$\gamma_n \left(\{ Q_t f \geqslant r \} \right) \leqslant \frac{1}{r},$$

for all $r \geqslant 1$. Now Markov inequality is only sharp for indicator functions and $Q_t f$ cannot be an indicator function, so it may be the case that this inequality can be improved. More precisely one might conjecture that for any fixed $t > 0$ (or at least for t large enough) there exists a function α satisfying

$$\lim_{r \to +\infty} \alpha(r) = 0$$

and

$$\gamma_n \left(\{ Q_t f \geqslant r \} \right) \leqslant \frac{\alpha(r)}{r}, \tag{1.2}$$

for every $r \geqslant 1$ and for every non-negative function f of integral 1. The function α should be independent of the dimension n, just as the hypercontractivity result stated above. Such a phenomenon was actually conjectured by Talagrand in [7] in a slightly different context. He conjectured that the same inequality holds true when γ_n is replaced by the uniform measure on the discrete cube $\{-1, 1\}^n$ and the Orstein-Uhlenbeck semigroup is replaced by the semigroup associated to the random walk on the discrete cube. The Gaussian version of the conjecture would follow from Talagrand's discrete version by the central limit theorem. In this paper we will only focus on the Gaussian case.

In [1], Ball, Barthe, Bednorz, Oleszkiewicz and Wolff showed that in dimension 1 the inequality (1.2) holds with decay

$$\alpha(r) = \frac{C}{\sqrt{\log r}},$$

where the constant C depends on the time parameter t. Moreover the authors provide an example showing that the $1/\sqrt{\log r}$ decay is sharp. They also have a result in higher dimension but they loose a factor $\log\log r$ and, more importantly, their constant C then tends to $+\infty$ (actually exponentially fast) with the dimension. The deadlock was broken recently by Eldan and Lee who showed in [2] that (1.2) holds with function

$$\alpha(r) = C\,\frac{(\log\log r)^4}{\sqrt{\log r}},$$

with a constant C that is independent of the dimension. Again up to the $\log\log$ factor the result is optimal.

In this article we revisit the argument of Eldan and Lee. We shall simplify some steps of their proof and short cut some others. As a result, we are able to remove the extra $\log\log$ factor. We would like to make clear though that this note does not really contain any new idea and that the core of our argument is all Eldan and Lee's.

2. Main results

Recall that γ_n is the standard Gaussian measure and that (Q_t) is the Ornstein-Uhlenbeck semigroup, defined by (1.1). Here is our main result.

THEOREM 2.1. — *Let f be a non-negative function on $\mathbb{R}^n$ satisfying $\int_{\mathbb{R}^n} f\,d\gamma_n = 1$ and let $t > 0$. Then for every $r > 1$*

$$\gamma_n\left(\{Q_t f > r\}\right) \leqslant C\,\frac{\max(1, t^{-1})}{r\sqrt{\log r}},$$

where C is a universal constant.

As in Eldan and Lee's paper, the Ornstein-Uhlenbeck semigroup only plays a rôle through the following lemma.

LEMMA 2.2. — *Let $f\colon \mathbb{R}^n \to \mathbb{R}_+$. For every $t > 0$, we have*

$$\nabla^2 \log(Q_t f) \geqslant -\frac{1}{2t}\,\mathrm{id},$$

pointwise.

Proof. — This is really straightforward: observe that (1.1) can be rewritten as

$$Q_t f(x) = (f * g_{1-\rho})(\sqrt{\rho}\, x),$$

where $\rho = \mathrm{e}^{-2t}$ and $g_{1-\rho}$ is the density of the Gaussian measure with mean 0 and covariance $(1-\rho)\mathrm{id}$. Then differentiate twice and use the Cauchy-Schwarz inequality. Details are left to the reader. □

What we actually prove is the following, where Q_t does not appear anymore.

THEOREM 2.3. — *Let f be a positive function on $\mathbb{R}^n$ satisfying $\int_{\mathbb{R}^n} f\, d\gamma_n = 1$. Assume that f is smooth and satisfies*

$$\nabla^2 \log f \geqslant -\beta\, \mathrm{id} \tag{2.3}$$

pointwise, for some $\beta \geqslant 0$. Then for every $r > 1$

$$\gamma_n\left(\{f > r\}\right) \leqslant \frac{C\, \max(\beta, 1)}{r\sqrt{\log r}},$$

where C is a universal constant.

Obviously, Theorem 2.3 and Lemma 2.2 altogether yield Theorem 2.1.

Let us comment on the optimality of Theorem 2.1 and Theorem 2.3. In dimension 1, consider the function

$$f_\alpha(x) = \mathrm{e}^{\alpha x - \alpha^2/2}.$$

Observe that $f_\alpha \geqslant 0$ and that $\int_{\mathbb{R}} f_\alpha\, d\gamma_1 = 1$. Note that for every $t \geqslant 1$ we have

$$\gamma_1\left([t, +\infty)\right) \geqslant \frac{c\, \mathrm{e}^{-t^2/2}}{t},$$

where c is a universal constant. So if $\alpha > 0$ and $r \geqslant \mathrm{e}$ then

$$\gamma_1\left(\{f_\alpha \geqslant r\}\right) \geqslant \frac{c\, \exp\left(-\frac{1}{2}\left(\frac{\log r}{\alpha} + \frac{\alpha}{2}\right)^2\right)}{\frac{\log r}{\alpha} + \frac{\alpha}{2}}.$$

Choosing $\alpha = \sqrt{2\log r}$ we get

$$\gamma_1\left(\{f_\alpha \geqslant r\}\right) \geqslant \frac{c'}{r\sqrt{\log r}}.$$

Since $(\log f_\alpha)'' = 0$ this shows that the dependence in r in Theorem 2.3 is sharp. Actually this example also shows that the dependence in r in Theorem 2.1 is sharp. Indeed, it is easily seen that

$$Q_t f_\alpha = f_{\alpha \mathrm{e}^{-t}},$$

for every $\alpha \in \mathbb{R}$ and $t > 0$. This implies that f_α always belongs to the image Q_t. Of course, this example also works in higher dimension: just replace f_α by

$$f_u(x) = \mathrm{e}^{\langle u, x\rangle - |u|^2/2}$$

where u belongs to $\mathbb{R}^n$.

THEOREM 2.4. — *Let X be a random vector having density f with respect to the Gaussian measure, and assume that f satisfies* (2.3)*. Then for every $r > 1$*

$$\mathsf{P}\left(f(X) \in (r, \mathrm{e}\, r]\right) \leqslant C\, \frac{\max(\beta, 1)}{\sqrt{\log r}}.$$

Theorem 2.4 easily yields Theorem 2.3.

Proof of Theorem 2.3. Let G be standard Gaussian vector on R^n and let X be a random vector having density f with respect to γ_n. Then using Theorem 2.4

$$\begin{aligned}
\mathsf{P}[f(G) > r] &= \sum_{k=0}^{+\infty} \mathsf{P}\left(f(G) \in (e^k r, e^{k+1} r]\right) \\
&\leqslant \sum_{k=0}^{+\infty} (\mathrm{e}^k r)^{-1} \mathsf{E}\left[f(G)\, \mathbf{1}_{\{f(G) \in (e^k r, e^{k+1} r]\}}\right] \\
&= \sum_{k=0}^{+\infty} (\mathrm{e}^k r)^{-1} \mathsf{P}\left(f(X) \in (e^k r, e^{k+1} r]\right) \\
&\leqslant \sum_{k=0}^{+\infty} (\mathrm{e}^k r)^{-1}\, C\, \frac{\max(\beta, 1)}{\sqrt{\log(\mathrm{e}^k r)}} \\
&\leqslant C\, \frac{\mathrm{e}}{\mathrm{e} - 1}\, \frac{1}{r}\, \frac{\max(\beta, 1)}{\sqrt{\log r}},
\end{aligned}$$

which is the result. □

The rest of the note is devoted to the proof of Theorem 2.4.

3. Preliminaries: the stochastic construction

Let μ be a probability measure on $\mathbb{R}^n$ having density f with respect to the Gaussian measure. We shall assume that f is bounded away from 0, that f is $\mathcal{C}^2$ and that ∇f and $\nabla^2 f$ are bounded. A simple density argument shows that we do not lose generality by adding these technical assumptions.

Eldan and Lee's argument is based on a stochastic construction which we describe now. Let (B_t) be a standard n-dimensional Brownian motion and let (P_t) be the associated semigroup:

$$P_t h(x) = \mathsf{E}[h(x + B_t)],$$

for all test functions h. Note that (P_t) is the heat semigroup, not the Ornstein-Uhlenbeck semigroup. Consider the stochastic differential equation

$$\begin{cases} X_0 = 0 \\ dX_t = dB_t + \nabla \log(P_{1-t} f)(X_t)\, dt, \quad t \in [0,1]. \end{cases} \tag{3.1}$$

The technical assumptions made on f insure that the map

$$x \mapsto \nabla \log P_{1-t} f(x)$$

is Lipschitz, with a Lipschitz norm that does not depend on $t \in [0,1]$. So the equation (3.1) has a strong solution (X_t). In our previous work [4] we study the process (X_t) in details and we give some applications to functional inequalities. Let us recap here some of these properties and refer to [4, section 2.5] for proofs. Recall that if μ_1, μ_2 are two probability measures, the relative entropy of μ_1 with respect to μ_2 is defined by

$$\mathrm{H}(\mu_1 \mid \mu_2) = \int \log\left(\frac{d\mu_1}{d\mu_2}\right) d\mu_1,$$

if μ_1 is absolutely continuous with respect to μ_2 (and $\mathrm{H}(\mu_1 \mid \mu_2) = +\infty$ otherwise). Also in the sequel we call *drift* any process (u_t) taking values in $\mathbb{R}^n$ which is adapted to the natural filtration of (B_t) (this means that u_t depends only on $(B_s)_{s \leqslant t}$) and satisfies

$$\int_0^1 |u_t|^2\, ds < +\infty,$$

almost surely. Let (v_t) be the drift

$$v_t = \nabla \log P_{1-t} f(X_t).$$

Using Itô's formula it is easily seen that

$$d \log P_{1-t} f(X_t) = \langle v_t, dB_t \rangle + \frac{1}{2} |v_t|^2\, dt.$$

Therefore, for every $t \in [0,1]$

$$P_{1-t} f(X_t) = \exp\left(\int_0^t \langle v_s, dB_s \rangle + \frac{1}{2} \int_0^t |v_s|^2 \, ds \right). \tag{3.2}$$

Combining this with the Girsanov change of measure theorem one can show that the random vector X_1 has law μ (again we refer to [4] for details). Moreover we have the equality

$$\mathrm{H}(\mu \mid \gamma_n) = \frac{1}{2} \mathsf{E} \left[\int_0^1 |v_s|^2 \, ds \right]. \tag{3.3}$$

Also, if (u_t) is any drift and if ν is the law of

$$B_1 + \int_0^1 u_t \, dt,$$

then

$$\mathrm{H}(\nu \mid \gamma_n) \leqslant \frac{1}{2} \mathsf{E} \left[\int_0^1 |u_s|^2 \, ds \right]. \tag{3.4}$$

So the drift (v_t) is in some sense optimal. Lastly, and this will play a crucial rôle in the sequel, the process (v_t) is a martingale.

Eldan and Lee introduce a perturbed version of the process (X_t), which we now describe. From now on we fix $r > 1$ and we let

$$T = \inf\{ t \in [0,1], \ P_{1-t} f(X_t) > r \} \ \wedge 1$$

be the first time the process $(P_{1-t} f(X_t))$ hits the value r, with the convention that $T = 1$ if it does not ever reach r (note that our definition of T differs from the one of Eldan and Lee). Now given $\delta > 0$ we let (X_t^δ) be the process defined by

$$X_t^\delta = X_t + \delta \int_0^{T \wedge t} v_s \, ds.$$

Since T is a stopping time, this perturbed process is still of the form Brownian motion plus drift:

$$X_t^\delta = B_t + \int_0^t (1 + \delta \mathbf{1}_{\{s \leqslant T\}}) v_s \, ds.$$

So letting μ^δ be the law of X_1^δ and using (3.4) we get

$$\begin{aligned} \mathrm{H}(\mu^\delta \mid \gamma_n) &\leqslant \frac{1}{2} \mathsf{E} \left[\int_0^1 (1 + \delta \, \mathbf{1}_{\{s \leqslant T\}})^2 |v_s|^2 \, ds \right] \\ &= \frac{1}{2} \mathsf{E} \left[\int_0^1 |v_s|^2 \, ds \right] + \left(\delta + \frac{\delta^2}{2} \right) \mathsf{E} \left[\int_0^T |v_s|^2 \, ds \right]. \end{aligned} \tag{3.5}$$

4. Proof of the main result

The proof can be decomposed into two steps. Recall that r is fixed from the beginning and that X_1^δ actually depends on r through the stopping time T.

The first step is to prove that if δ is small then μ and μ^δ are not too different.

PROPOSITION 4.1. — *Assuming* (2.3)*, we have*

$$d_{TV}(\mu, \mu^\delta) \leqslant \delta \sqrt{(\beta+1)\log r},$$

for every $\delta > 0$*, where* d_{TV} *denotes the total variation distance.*

The second step is to argue that $f(X_1^\delta)$ tends to be bigger than $f(X_1)$. An intuition for this property is that the difference between X_1^δ and X_1 is somehow in the direction of $\nabla f(X_1)$.

PROPOSITION 4.2. — *Assuming* (2.3)*, we have*

$$\mathsf{P}\left(f(X_1^\delta) \leqslant r^{1+2\delta} \mathrm{e}^{-4}\right) \leqslant \mathsf{P}(f(X_1) \leqslant r) + (\beta+4)\delta^2 \log(r),$$

for all $\delta > 0$.

Remark.— Note that both propositions use the convexity hypothesis (2.3).

It is now very easy to prove Theorem 2.4. Since X_1 has law μ, all we need to prove is

$$\mathsf{P}(f(X_1) \in (r, \mathrm{e}r]) \leqslant C \, \frac{\max(\beta, 1)}{\sqrt{\log r}}.$$

We choose

$$\delta = \frac{5}{2\log r}.$$

For this value of δ, Proposition 4.2 gives

$$\mathsf{P}(f(X_1^\delta) \leqslant \mathrm{e}\, r) \leqslant \mathsf{P}(f(X_1) \leqslant r) + \frac{25}{4} \, \frac{\beta+4}{\log r},$$

whereas Proposition 4.1 yields

$$\begin{aligned} \mathsf{P}(f(X_1) \leqslant \mathrm{e}r) &\leqslant \mathsf{P}(f(X_1^\delta) \leqslant \mathrm{e}r) + d_{TV}(\mu, \mu^\delta) \\ &\leqslant \mathsf{P}(f(X_1^\delta) \leqslant \mathrm{e}r) + \frac{5}{2} \left(\frac{\beta+1}{\log r} \right)^{1/2}. \end{aligned}$$

Combining the two inequalities we obtain

$$\mathsf{P}(f(X_1) \leqslant \mathrm{e}\, r) \leqslant \mathsf{P}(f(X_1) \leqslant r) + C\, \frac{\max(\beta, 1)}{\sqrt{\log r}},$$

which is the result.

Remark. We actually prove the slightly stronger statement:

$$\mathsf{P}(f(X_1) \in (r, \mathrm{e}r]) \leqslant C \max\left(\frac{\max(\beta,1)}{\log r}, \left(\frac{\max(\beta,1)}{\log r} \right)^{1/2} \right).$$

5. Proof of the total variation estimate

We actually bound the relative entropy of μ^δ with respect to μ. Recall that $\log f$ is assumed to be weakly convex: there exists $\beta \geqslant 0$ such that

$$\nabla^2 \log f \geqslant -\beta\, \mathrm{id}, \tag{5.6}$$

pointwise.

Proposition 5.1. — *Assuming* (5.6), *we have*

$$\mathrm{H}(\mu^\delta \mid \mu) \leqslant \delta^2 (\beta + 1) \log r,$$

for all $\delta > 0$.

This yields Proposition 4.1 by Pinsker's inequality.

Proof. — Observe that

$$\mathrm{H}(\mu^\delta \mid \mu) = \mathrm{H}(\mu^\delta \mid \gamma_n) - \int_{\mathbb{R}^n} \log(f)\, d\mu^\delta. \tag{5.7}$$

Now (5.6) gives

$$\begin{aligned} \log(f)(X_1^\delta) &\geqslant \log f(X_1) + \langle \nabla \log f(X_1), X_1^\delta - X_1 \rangle - \frac{\beta}{2} |X_1^\delta - X_1|^2, \\ &\geqslant \log f(X_1) + \delta \int_0^T \langle v_1, v_s \rangle\, ds - \frac{\beta \delta^2}{2} \int_0^T |v_s|^2\, ds, \end{aligned} \tag{5.8}$$

almost surely. We shall use this inequality several times in the sequel. Recall that X_1 has law μ and that X_1^δ has law μ^δ. Taking expectation in the previous inequality and using (5.7) we get

$$\begin{aligned} \mathrm{H}(\mu^\delta \mid \mu) &\leqslant \mathrm{H}(\mu^\delta \mid \gamma_n) - \mathrm{H}(\mu \mid \gamma_n) \\ &\quad - \delta\, \mathsf{E}\left[\int_0^T \langle v_1, v_s \rangle\, ds \right] + \frac{\beta\delta^2}{2}\, \mathsf{E}\left[\int_0^T |v_s|^2\, ds \right]. \end{aligned}$$

Together with (3.3) and (3.5) we obtain

$$\mathrm{H}(\mu^\delta \mid \mu) \leqslant -\delta\, \mathsf{E}\left[\int_0^T \langle v_1 - v_s, v_s\rangle\, ds\right] + \frac{(1+\beta)\delta^2}{2}\, \mathsf{E}\left[\int_0^T |v_s|^2\, ds\right].$$

Now since (v_t) is a martingale and T a stopping time we have

$$\mathsf{E}\left[\langle v_1, v_s\rangle\, \mathbf{1}_{\{s \leqslant T\}}\right] = \mathsf{E}\left[|v_s|^2 \mathbf{1}_{\{s\leqslant T\}}\right]$$

for all time $s \leqslant 1$. This shows that the first term in the previous inequality is 0. To bound the second term, observe that the definition of T and the equality (3.2) imply that

$$\int_0^T \langle v_s, dB_s\rangle + \frac{1}{2}\int_0^T |v_s|^2\, ds \leqslant \log r,$$

almost surely. Since (v_t) is a bounded drift, the process $(\int_0^t \langle v_s, dB_s\rangle)$ is a martingale. Now T is a bounded stopping time, so by the optional stopping theorem

$$\mathsf{E}\left[\int_0^T \langle v_s, dB_s\rangle\right] = 0.$$

Therefore, taking expectation in the previous inequality yields

$$\mathsf{E}\left[\int_0^T |v_s|^2\, ds\right] \leqslant 2\log r,$$

which concludes the proof. □

6. Proof of Proposition 4.2

The goal is to prove that

$$\mathsf{P}\left(f(X_1^\delta) \leqslant r^{1+2\delta}\mathrm{e}^{-4}\right) \leqslant \mathsf{P}(f(X_1) \leqslant r) + \delta^2(\beta+4)\log r.$$

Obviously

$$\mathsf{P}\left(f(X_1^\delta) \leqslant r^{1+2\delta}\mathrm{e}^{-4}\right) \leqslant \mathsf{P}(f(X_1) \leqslant r) + \mathsf{P}\left(f(X_1^\delta) \leqslant r^{1+2\delta}\mathrm{e}^{-4};\ f(X_1) > r\right)$$

Now recall the inequality (5.8) coming for the weak convexity of $\log f$ and rewrite it as

$$\log f(X_1^\delta) \geqslant K_1 + 2\delta K_T + Y$$

where (K_t) is the process defined by

$$K_t = \log(P_{1-t})(f)(X_t) = \int_0^t \langle v_s, dB_s\rangle + \frac{1}{2}\int_0^t |v_s|^2\, ds,$$

and Y is the random variable

$$Y = -2\delta \int_0^T \langle v_s, dB_s \rangle + \delta \int_0^T \langle v_1 - v_s, v_s \rangle \, ds - \frac{\beta \delta^2}{2} \int_0^T |v_s|^2 \, ds.$$

Recall that the stopping time T is the first time the process (K_t) exceeds the value $\log r$ if it ever does, and $T = 1$ otherwise. In particular, if

$$K_1 = \log f(X_1) > \log r$$

then $K_T = \log r$. So if $f(X_1) > r$ then

$$f(X_1^\delta) > r^{1+2\delta} \, \mathrm{e}^Y.$$

Therefore

$$\mathsf{P}\left(f(X_1^\delta) \leqslant r^{1+2\delta} \mathrm{e}^{-4}; \; f(X_1) > r \right) \leqslant \mathsf{P}(Y \leqslant -4).$$

So we are done if we can prove that

$$\mathsf{P}(Y \leqslant -4) \leqslant (\beta + 4) \delta^2 \log r. \tag{6.1}$$

There are three terms in the definition of Y. The problematic one is

$$\delta \int_0^T \langle v_1 - v_s, v_s \rangle \, ds.$$

We know from the previous section that it has expectation 0. A natural way to get a deviation bound would be to estimate its second moment but it is not clear to us how to do this. Instead we make a complicated detour.

Lemma 6.1. — *Let Z be an integrable random variable satisfying $\mathsf{E}[\mathrm{e}^Z] \leqslant 1$. Then*

$$\mathsf{P}(Z \leqslant -2) \leqslant -\mathsf{E}[Z].$$

Remark. — Note that $\mathsf{E}[Z] \leqslant 0$ by Jensen's inequality.

Proof. — Simply write

$$\begin{aligned} \mathsf{E}\left[\mathrm{e}^Z\right] &\geqslant \mathsf{E}\left[\mathrm{e}^Z \, \mathbf{1}_{\{Z > -2\}}\right] \\ &\geqslant \mathsf{E}\left[(Z+1) \, \mathbf{1}_{\{Z > -2\}}\right] \\ &= \mathsf{E}[Z] - \mathsf{E}\left[Z \, \mathbf{1}_{\{Z \leqslant -2\}}\right] + 1 - \mathsf{P}(Z \leqslant -2) \\ &\geqslant \mathsf{E}[Z] + \mathsf{P}(Z \leqslant -2) + 1. \end{aligned}$$

So if $\mathsf{E}[\mathrm{e}^Z] \leqslant 1$ then $\mathsf{P}(Z \leqslant -2) \leqslant -\mathsf{E}[Z]$. □

LEMMA 6.2. — *Let Z be the variable*

$$Z = -\delta \int_0^T \langle v_s, dB_s \rangle + \delta \int_0^T \langle v_1 - v_s, v_s \rangle \, ds - \frac{(\beta+1)\delta^2}{2} \int_0^T |v_s|^2 \, ds.$$

Then

$$\mathsf{P}(Z \leqslant -2) \leqslant \delta^2(\beta+1) \log r.$$

Proof. — As we have seen before the first two terms in the definition of Z have expectation 0 and

$$\mathsf{E}[Z] = -\frac{(\beta+1)\delta^2}{2} \, \mathsf{E}\left[\int_0^T |v_s|^2 \, ds \right] \geqslant -\delta^2(\beta+1) \log r.$$

By Lemma 6.1 it is enough to show that $\mathsf{E}[\mathrm{e}^Z] \leqslant 1$. To do so, we use the Girsanov change of measure formula. The process (X_t^δ) is of the form Brownian motion plus drift:

$$\begin{aligned} X_t^\delta &= X_t + \delta \int_0^{T \wedge t} v_s \, ds \\ &= B_t + \int_0^t (1 + \delta \mathbf{1}_{\{s \leqslant T\}}) v_s \, ds, \end{aligned}$$

Note also that the drift term is bounded. Therefore, Girsanov's formula applies, see for instance [5, chapter 6] (beware that the authors oddly use the letter M to denote expectation). The process (D_t^δ) defined by

$$D_t^\delta = \exp\left(-\int_0^t (1 + \delta \, \mathbf{1}_{\{s \leqslant T\}}) \langle v_s, dB_s \rangle - \frac{1}{2} \int_0^t \left| (1 + \delta \, \mathbf{1}_{\{s \leqslant T\}}) v_s \right|^2 \, ds \right)$$

is a non-negative martingale of expectation 1 and under the measure Q^δ defined by

$$d\mathsf{Q}^\delta = D_1^\delta \, d\mathsf{P}$$

the process (X_t^δ) is a standard Brownian motion. In particular

$$\mathsf{E}[f(X_1^\delta) D_1^\delta] = \mathsf{E}[f(B_1)] = 1.$$

Now use inequality (5.8) once again and combine it with equality (3.2) written at time $t = 1$. This gives exactly

$$f(X_1^\delta) \, D_1^\delta \geqslant \mathrm{e}^Z.$$

Therefore $\mathsf{E}[\mathrm{e}^Z] \leqslant 1$, which concludes the proof. □

We now prove inequality (6.1). The idea being that the annoying term in Y is handled by the previous lemma. Observe that

$$Y = Z - \delta \int_0^T \langle v_s, dB_s \rangle - \frac{\delta^2}{2} \int_0^T |v_s|^2 \, ds.$$

So

$$\mathsf{P}(Y \leqslant -4) \leqslant \mathsf{P}(Z \leqslant -2) + \mathsf{P}\left(\delta \int_0^T \langle v_s, dB_s \rangle \geqslant 1 \right) + \mathsf{P}\left(\frac{\delta^2}{2} \int_0^T |v_s|^2 \, ds \geqslant 1 \right).$$

Recall that $\int_0^T \langle v_s, dB_s \rangle$ has mean 0 and observe that

$$\mathsf{E}\left[\left(\delta \int_0^T \langle v_s, dB_s \rangle \right)^2 \right] = \delta^2 \mathsf{E}\left[\int_0^T |v_s|^2 \, ds \right] \leqslant 2\delta^2 \log r.$$

So by Tchebychev inequality

$$\mathsf{P}\left(\delta \int_0^T \langle v_s, dB_s \rangle \geqslant 1 \right) \leqslant 2\delta^2 \log r.$$

Similarly by Markov inequality

$$\mathsf{P}\left(\frac{\delta^2}{2} \int_0^T |v_s|^2 \, ds \geqslant 1 \right) \leqslant \delta^2 \log r.$$

Putting everything together we get (6.1), which concludes the proof.

Acknowledgments

We would like to thank Ronen Eldan and James Lee for introducing us to the problem and communicating their work to us. We are also grateful to Michel Ledoux for some fruitful discussions at the final stage of this work.

Bibliography

[1] Ball (K.), Barthe (F.), Bednorz (W.), Oleszkiewicz (K.), Wolff (P.). L_1-smoothing for the Ornstein-Uhlenbeck semigroup, Mathematika 59, no. 1, p. 160-168 (2013).

[2] Eldan (R.), Lee (J.). — Regularization under diffusion and anti-concentration of temperature, arXiv:1410.3887.

[3] Gross (L.). — Logarithmic Sobolev inequalities, Amer. J. Math. 97, no. 4, p. 1061-1083 (1975).

[4] Lehec (J.). — Representation formula for the entropy and functional inequalities, Ann. Inst. Henri Poincaré Probab. Stat. 49, no. 3, p. 885-899 (2013).

[5] Liptser (R.), Shiryaev (A.). — Statistics of random processes., Vol I, general theory. 2nd edition, Stochastic Modelling and Applied Probability, Springer-Verlag, Berlin (2001).

[6] Nelson (E.). — The free Markoff field. J. Functional Analysis 12, p. 211-227 (1973).

[7] Talagrand (M.). — A conjecture on convolution operators, and a non-Dunford-Pettis operator on L_1, Israel J. Math. 68, no. 1, p. 82-88 (1989).