A dimension-free reverse logarithmic Sobolev inequality for low-complexity functions in Gaussian space

Ronen Eldan[*]
Weizmann Institute, Israel

Michel Ledoux
University of Toulouse, France

September 4, 2019

Abstract

We discuss new proofs, and new forms, of a reverse logarithmic Sobolev inequality, with respect to the standard Gaussian measure, for low-complexity functions, measured in terms of Gaussian-width. In particular, we provide a dimension-free improvement for a related result given in [5].

1 A reverse logarithmic Sobolev inequality

The recent work [5] has put forward a reverse logarithmic Sobolev inequality, with respect to the standard Gaussian measure, for low-complexity functions measured in terms of Gaussian-width. To briefly recall this inequality, we take again the notation from [5].

Let γ denote the standard Gaussian measure on $\mathbb{R}^n$, and let ν be the probability measure $d\nu = e^f d\gamma$ where $f : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable. Let

$$\mathrm{D}_{\mathrm{KL}}\big(\nu \,||\, \gamma\big) \,=\, \int_{\mathbb{R}^n} f \, d\nu \qquad \text{and} \qquad \mathcal{I}(\nu) \,=\, \int_{\mathbb{R}^n} |\nabla f|^2 d\nu$$

be respectively the Kullback-Leibler divergence (relative entropy) and the Fisher information of ν with respect to γ, assumed to be finite in the following. In particular ν has a second moment. The standard logarithmic Sobolev inequality of L. Gross (cf. e.g. [11, 3]) ensures that

$$\mathrm{D}_{\mathrm{KL}}\big(\nu \,||\, \gamma\big) \,\leq\, \frac{1}{2}\, \mathcal{I}(\nu). \tag{1}$$

Let

$$\mathcal{D}(\nu) = \mathbf{GW}(K) = \int_{\mathbb{R}^n} \sup_{t \in K} \langle y, t \rangle \, d\gamma(y)$$

be the Gaussian-width of the set $K = \{\nabla f(x) \, ; \, x \in \mathbb{R}^n\}$. Remark that the above expression is integrable if and only if f is a Lipschitz function, which is the only case in which we obtain meaningful results; Otherwise, set $\mathcal{D}(\nu) = \infty$. The quantity $\mathcal{D}(\nu)$ is a measure of the complexity of ν (rather the gradient-complexity of f). The following reverse logarithmic Sobolev inequality for measures of low-complexity has been established in [5, Theorem 4]. Set $M(\nu) = -\inf_{x \in \mathbb{R}^n} \Delta f(x)$, assumed to be finite. Then one has

$$\frac{1}{2}\mathcal{I}(\nu) \leq \mathrm{D_{KL}}(\nu \, \| \, \gamma) + \frac{1}{2} M(\nu)_+ + \mathcal{D}(\nu)^{2/3} \mathcal{I}(\nu)^{1/3}, \qquad (2)$$

where $M(\nu)_+ = \max(M(\nu), 0)$.

Our first theorem gives the following related bound.

Theorem 1. *In the preceding notation,*

$$\frac{1}{2}\mathcal{I}(\nu) \leq \mathrm{D_{KL}}\big(\nu \, \| \, \gamma\big) + M(\nu) + \mathcal{D}(\nu).$$

As discussed in [5], the inequality is sharp on the extremal functions $f(x) = \langle \alpha, x \rangle$, $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^n$, of the logarithmic Sobolev inequality which have complexity $M(\nu) = \mathcal{D}(\nu) = 0$.

To compare between Theorem 1 and the bound (2), observe that the latter trivially holds true in the case that $\mathcal{I}(\nu) \leq \mathcal{D}(\nu)$, thus we may generally assume that $\mathcal{D}(\nu) \leq \mathcal{D}(\nu)^{2/3} \mathcal{I}(\nu)^{1/3}$. Unlike inequality (2), the bound of the theorem has the feature that both sides of the inequality are additive with respect to taking products and in this sense it is dimension-free. In Section 2 below, we give a slightly different form which improves on (2), and improves on Theorem 1 when, for example, the potential f is convex.

Such a reverse logarithmic Sobolev inequality is of theoretical interest in the study of approximations of partition functions and of low-complexity Gibbs measures on product spaces (cf. [5, 1]). An analogous definition of low-complexity for Boolean functions was considered in [5], where it is shown that a low-complexity condition implies that the measure can be decomposed as a mixture of approximate product measures.

In fact, it was very recently shown ([7]) that if a measure ν satisfies a reverse logarithmic Sobolev inequality, then it is close, in transportation distance, to a mixture of translated Gaussian measures. The combination of such a result with Theorem 1 gives a structure theorem for measures of low-complexity, analogous to the one given in [5], but for the Gaussian setting. We formulate this as a corollary.

Recall the quadratic Kantorovich metric $\mathrm{W}_2(\nu, \gamma)$ between ν and γ defined by

$$\mathrm{W}_2^2(\nu, \gamma) = \inf \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y)$$

where the infimum is taken over all couplings π on $\mathbb{R}^n \times \mathbb{R}^n$ with respective marginals ν and γ. A combination of Theorem 1 with [7, Theorem 5] yields the following statement.

Corollary 2. *In the preceding notation, there exists a probability measure μ such that*

$$\mathrm{W}_2^2(\nu, \gamma \star \mu) \leq 16n^{1/3}\big(M(\nu) + \mathcal{D}(\nu)\big)^{2/3}.$$

In particular, the above corollary gives a meaningful result whenever $M(\nu) + \mathcal{D}(\nu) = o(n)$. It is also conjectured that a dimension-free analogue of [7, Theorem 5] should hold true, which, combined with our bound would imply the existence of a probability measure μ such that

$$\mathrm{W}_2^2(\nu, \gamma \star \mu) \leq C\big(M(\nu) + \mathcal{D}(\nu)\big),$$

where $C > 0$ is a universal constant.

The proof of (2) strongly relies on a construction coming from stochastic control theory, of an entropy-optimal coupling of the measure ν to a Brownian motion. We will come back to it in Section 2. In contrast, our proof of Theorem 1 follows a simple and direct approach.

Proof of Theorem 1. Recall that we may assume that f is Lipschitz, since otherwise $\mathcal{D}(\nu) = \infty$. Thus we have $\nabla f(x)\frac{d\gamma}{dx} \to 0$ uniformly as $|x| \to \infty$ and therefore, by integration by parts with respect to the Gaussian measure γ,

$$
\begin{aligned}
\mathcal{I}(\nu) \;=\; \int_{\mathbb{R}^n} |\nabla f|^2 d\nu \;&=\; \int_{\mathbb{R}^n} |\nabla f|^2 e^f d\gamma \\
&=\; \int_{\mathbb{R}^n} \langle \nabla(e^f), \nabla f \rangle \, d\gamma \\
&=\; -\int_{\mathbb{R}^n} e^f \, \mathrm{L}f \, d\gamma \;=\; -\int_{\mathbb{R}^n} \mathrm{L}f \, d\nu
\end{aligned}
$$

where $\mathrm{L} = \Delta f - \langle x, \nabla f \rangle$ is the Ornstein-Uhlenbeck operator. Therefore

$$\mathcal{I}(\nu) \;=\; -\int_{\mathbb{R}^n} \Delta f \, d\nu + \int_{\mathbb{R}^n} \langle x, \nabla f \rangle \, d\nu \;\leq\; M(\nu) + \int_{\mathbb{R}^n} \langle x, \nabla f \rangle \, d\nu. \tag{3}$$

Let π be a coupling on $\mathbb{R}^n \times \mathbb{R}^n$ with respective marginals ν and γ. Then,

$$
\begin{aligned}
\int_{\mathbb{R}^n} \langle x, \nabla f(x) \rangle \, d\nu(x) \;&=\; \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, \nabla f(x) \rangle \, d\pi(x, y) \\
&=\; \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle y, \nabla f(x) \rangle \, d\pi(x, y) + \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \nabla f(x) \rangle \, d\pi(x, y).
\end{aligned}
$$

Now, on the one hand,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle y, \nabla f(x) \rangle \, d\pi(x, y) \;\leq\; \int_{\mathbb{R}^n \times \mathbb{R}^n} \sup_{t \in K} \langle y, t \rangle \, d\pi(x, y) \;=\; \int_{\mathbb{R}^n} \sup_{t \in K} \langle y, t \rangle \, d\gamma(y).$$

On the other hand, by the standard quadratic inequality,

$$
\begin{aligned}
\int_{\mathbb{R}^n \times \mathbb{R}^n} &\langle x - y, \nabla f(x) \rangle \, d\pi(x, y) \\
&\leq\; \frac{1}{2} \int_{\mathbb{R}^n \times \mathbb{R}^n} \big|\nabla f(x)\big|^2 d\pi(x, y) + \frac{1}{2} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y) \\
&=\; \frac{1}{2} \int_{\mathbb{R}^n} \big|\nabla f(x)\big|^2 d\nu(x) + \frac{1}{2} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 d\pi(x, y).
\end{aligned}
$$

3

Taking the infimum over all couplings π with respective marginals ν and γ, it holds true that

$$\int_{\mathbb{R}^n} \langle x, \nabla f \rangle \, d\nu \leq \int_{\mathbb{R}^n} \sup_{t \in K} \langle y, t \rangle \, d\gamma(y) + \frac{1}{2} \int_{\mathbb{R}^n} |\nabla f(x)|^2 d\nu(x) + \frac{1}{2} \, W_2^2(\nu, \gamma).$$

Therefore, together with (3),

$$\frac{1}{2} \mathcal{I}(\nu) \leq M(\nu) + \mathcal{D}(\nu) + \frac{1}{2} \, W_2^2(\nu, \gamma). \tag{4}$$

It remains to recall the quadratic transportation cost inequality by M. Talagrand (cf. [11, 3])

$$\frac{1}{2} \, W_2^2(\nu, \gamma) \leq D_{KL}(\nu \,\|\, \gamma) \tag{5}$$

and the proof is complete. $\qquad\square$

Together with the logarithmic Sobolev inequality (1) $D_{KL}(\nu\|\gamma) \leq \frac{1}{2} \mathcal{I}(\nu)$, the step (4) of the preceding proof actually also yields a reverse transportation cost inequality

$$D_{KL}(\nu \,\|\, \gamma) \leq \frac{1}{2} \, W_2^2(\nu, \gamma) + M(\nu) + \mathcal{D}(\nu). \tag{6}$$

Theorem 1 may also be deduced from a classical integrability result for the supremum of a Gaussian process. Given a bounded set $K \subset \mathbb{R}^n$, observe that $x \mapsto \sup_{t \in K} \langle x, t \rangle$ is integrable with respect to γ. Setting $Z = Z(x) = \sup_{t \in K} \left[\langle x, t \rangle - \frac{1}{2}|t|^2 \right]$, $x \in \mathbb{R}^n$, it holds true that

$$\int_{\mathbb{R}^n} e^Z d\gamma \leq e^{\int_{\mathbb{R}^n} \sup_{t \in K} \langle x, t \rangle d\gamma}. \tag{7}$$

This inequality was originally put forward in [10, 12] in the context of concentration properties of suprema of Gaussian processes. Now, the classical entropic inequality (Gibbs variational principle) expresses that

$$\int_{\mathbb{R}^n} Z \, d\nu \leq D_{KL}(\nu \,\|\, \gamma) + \log \int_{\mathbb{R}^n} e^Z d\gamma.$$

With $K = \{\nabla f(x) \,;\, x \in \mathbb{R}^n\}$, it therefore follows that

$$\int_{\mathbb{R}^n} \left[\langle x, \nabla f \rangle - \frac{1}{2} |\nabla f|^2 \right] d\nu \leq \int_{\mathbb{R}^n} Z \, d\nu \leq D_{KL}(\nu \,\|\, \gamma) + \int_{\mathbb{R}^n} \sup_{t \in K} \langle x, t \rangle \, d\gamma.$$

Again, together with (3), this yields the conclusion of the theorem.

At the same time, the integrability inequality (7) may be seen as a consequence of the transportation cost inequality (5) and the Kantorovich duality. The argument actually works for any probability μ on the Borel sets of $\mathbb{R}^n$ satisfying the transportation cost inequality

$$\frac{1}{2C} \, W_2^2(\mu', \mu) \leq D_{KL}(\mu' \,\|\, \mu) \tag{8}$$

4

for some constant $C > 0$ and every $\mu' \ll \mu$ ($C = 1$ for $\mu = \gamma$). Namely, the Kantorovich duality (cf. [11]) expresses that

$$\frac{1}{2} W_2^2(\mu', \mu) = \sup \left[\int_{\mathbb{R}^n} \varphi \, d\mu' + \int_{\mathbb{R}^n} \psi \, d\mu \right]$$

where the supremum runs over the set of measurable functions $(\varphi, \psi) \in L^1(\mu') \times L^1(\mu)$ satisfying

$$\varphi(x) + \psi(y) \le \frac{1}{2} |x - y|^2 \tag{9}$$

for $d\mu'$-almost all $x \in \mathbb{R}^n$ and $d\mu$-almost all $y \in \mathbb{R}^n$. Given then a couple of functions (φ, ψ) satisfying (9), the choice in (8) of $\frac{d\mu'}{d\mu} = \frac{e^g}{\int_{\mathbb{R}^n} e^g d\mu}$ where $g = \frac{1}{C}[\varphi + \int_{\mathbb{R}^n} \psi \, d\mu]$ yields that $\log \int_{\mathbb{R}^n} e^g d\mu \le 0$, that is

$$\int_{\mathbb{R}^n} e^{\frac{1}{C}\varphi} d\mu \le e^{-\frac{1}{C}\int_{\mathbb{R}^n} \psi d\mu}.$$

For every $x, y \in \mathbb{R}^n$, and $t \in K$,

$$\langle x, t \rangle - \frac{1}{2}|t|^2 = \langle y, t \rangle + \langle x - y, t \rangle - \frac{1}{2}|t|^2 \le \langle y, t \rangle + \frac{1}{2}|x - y|^2.$$

Therefore, if $\varphi(x) = \sup_{t \in K} \left[\langle x, t \rangle - \frac{1}{2}|t|^2 \right]$, $x \in \mathbb{R}^n$, then $\psi(y) = -\sup_{t \in K} \langle y, t \rangle$ is a valid candidate for (9). Hence

$$\int_{\mathbb{R}^n} e^{\frac{1}{C}Z} d\mu \le e^{\frac{1}{C}\int_{\mathbb{R}^n} \sup_{t \in K} \langle x, t \rangle d\mu}$$

which amounts to (7) when $\mu = \gamma$.

2 Stochastic calculus and the Föllmer process

As mentioned above, the proof of (2) developed in [5] uses tools from stochastic control theory, and in particular the so-called Föllmer process [8] to achieve an entropy-optimal coupling of the measure ν to a Brownian motion. This argument has already been proved useful in the study of various functional inequalities [4, 9, 6].

To summarize a few facts from [9, 5], let $(B_t)_{t \ge 0}$ be standard Brownian motion in $\mathbb{R}^n$ (starting from the origin) adapted to a filtration $(\mathcal{F}_t)_{t \ge 0}$. Set $v(t, x) = \nabla \log Z(t, x)$, $t \in [0, 1]$, $x \in \mathbb{R}^n$, where

$$Z(t, x) = \mathbb{E}\big([e^f](x + B_{1-t})\big).$$

The Föllmer process $(X_t)_{t \in [0,1]}$ solves the stochastic differential equation

$$X_0 = 0, \quad dX_t = dB_t + v_t dt$$

where $v_t = v(t, X_t)$. Amongst its relevant properties, the random variable X_1 has distribution ν, $(v_t)_{t \in [0,1]}$ is a martingale, and

$$\mathbb{E}\left(\int_0^1 |v_t|^2 dt \right) = 2 \, D_{KL}(\nu \,||\, \gamma).$$

The arguments developed in [5] thus make use of these properties towards a proof of the inequality (2). Now, actually, a small variation in the same spirit allows for the following inequality.

Theorem 3. *In the notation of Section 1, assume that ν has a finite second moment. Then*

$$\int_{\mathbb{R}^n} |x|^2 d\nu - \int_{\mathbb{R}^n} |x|^2 d\gamma \leq 2\,\mathrm{D}_{\mathrm{KL}}(\nu\,||\,\gamma) + \mathcal{D}(\nu).$$

For the sake of intuition, let us consider an equivalent form of the bound provided by this theorem. Denote

$$\mathrm{H}(\nu) = -\int_{\mathbb{R}^n} \log\left(\frac{d\nu}{dx}\right) d\nu,$$

the differential entropy of ν. It is then straightforward to check that the inequality of Theorem 3 is equivalent to

$$\mathrm{H}(\nu) - \mathrm{H}(\gamma) \leq \frac{1}{2}\,\mathcal{D}(\nu). \tag{10}$$

Note that, in the special case that ν has the form $\log\frac{d\nu}{dx} = \sup_{t \in K}\left[\langle x, t\rangle - \frac{1}{2}|t|^2\right] + \mathrm{const}$, this bound becomes somewhat similar to the bound (7).

Proof of Theorem 3. Observe first that by integration by parts

$$\int_{\mathbb{R}^n} |x|^2 d\nu - \int_{\mathbb{R}^n} |x|^2 d\gamma = \int_{\mathbb{R}^n} |x|^2 d\nu - n = \int_{\mathbb{R}^n} \langle x, \nabla f\rangle\, d\nu$$

where the boundary term vanishes due to the legitimate assumption that f is Lipschitz (otherwise $\mathcal{D}(\nu) = \infty$ and there is nothing to prove). The inequality of the theorem amounts to

$$\int_{\mathbb{R}^n} \langle x, \nabla f\rangle\, d\nu \leq 2\,\mathrm{D}_{\mathrm{KL}}(\nu\,||\,\gamma) + \mathcal{D}(\nu). \tag{11}$$

Recall $K = \{\nabla f(x)\,;\, x \in \mathbb{R}^n\}$. Arguing as for the proof of Theorem 1, for any coupling π with respective marginals ν and γ,

$$\int_{\mathbb{R}^n} \langle x, \nabla f\rangle\, d\nu = \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x, \nabla f(x)\rangle\, d\pi(x, y)$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \nabla f(x)\rangle\, d\pi(x, y) + \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle y, \nabla f(x)\rangle\, d\pi(x, y)$$

$$\leq \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \nabla f(x)\rangle\, d\pi(x, y) + \int_{\mathbb{R}^n \times \mathbb{R}^n} \sup_{t \in K} \langle y, t\rangle\, d\pi(x, y)$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \nabla f(x)\rangle\, d\pi(x, y) + \mathcal{D}(\nu).$$

The inequality (11) would then follow if for some coupling π,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x - y, \nabla f(x)\rangle\, d\pi(x, y) \leq 2\,\mathrm{D}_{\mathrm{KL}}(\nu\,||\,\gamma).$$

But the point is that the Föllmer process actually produces an exact coupling for this identity to hold. Namely, by the definition and properties of $(X_t)_{t \in [0,1]}$, X_1 has law ν, B_1 has law γ and

$$d\langle X_t - B_t, v_t \rangle = |v_t|^2 dt + \langle X_t - B_t, dv_t \rangle.$$

Since $(v_t)_{t \in [0,1]}$ is a martingale,

$$\mathbb{E}\big(\langle X_1 - B_1, v_1 \rangle\big) = \mathbb{E}\left(\int_0^1 |v_t|^2 dt \right) = 2\,\mathrm{D_{KL}}\big(\nu \,||\, \gamma\big)$$

from which the claim follows since $v_1 = \nabla f(X_1)$. □

It remains to connect Theorem 3, or rather inequality (11), to Theorem 1. By the definition of the Fisher information (cf. (3))

$$\mathcal{I}(\nu) = - \int_{\mathbb{R}^n} \Delta f \, d\nu + \int_{\mathbb{R}^n} \langle x, \nabla f \rangle \, d\nu,$$

so that (11) expresses that

$$\frac{1}{2} \mathcal{I}(\nu) \leq \mathrm{D_{KL}}\big(\nu \,||\, \gamma\big) - \frac{1}{2} \int_{\mathbb{R}^n} \Delta f \, d\nu + \frac{1}{2} \mathcal{D}(\nu). \tag{12}$$

As a first minor observation, note that, according to the discussion following Theorem 1, the above inequality improves upon (2) with the correct factor $\frac{1}{2}$ in front of $M(\nu)_+$ (unlike Theorem 1).

While presented and established with the quantity $M(\nu) = -\inf_{x \in \mathbb{R}^n} \Delta f(x)$, the proof of Theorem 1 shows in the same way that

$$\frac{1}{2} \mathcal{I}(\nu) \leq \mathrm{D_{KL}}\big(\nu \,||\, \gamma\big) - \int_{\mathbb{R}^n} \Delta f \, d\nu + \mathcal{D}(\nu). \tag{13}$$

Hence, if $\mathcal{D}(\nu) \geq \int_{\mathbb{R}^n} \Delta f \, d\nu$ (which holds in particular for convex potentials), the inequality (12) improves upon (13). On the other hand, it does not seem possible to reach (12) as simply as (13), and in any case, the inequality of Theorem 3, even up to a constant, may not be deduced from Theorem 1.

References

[1] T. Austin. The structure of low-complexity Gibbs measures on product spaces (2018).

[2] T. Austin. Multi-variate correlation and mixtures of product measures (2018).

[3] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften 348. Springer (2014).

[4] C. Borell. Isoperimetry, log-concavity, and elasticity of option prices. In *New directions in Mathematical Finance,* 73–91, Wiley (2002).

[5] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* 28, 1548–1596 (2018).

[6] R. Eldan, J. R. Lee. Regularization under diffusion and anti-concentration of the information content. *Duke Math. J.* 167, 969–993 (2018).

[7] R. Eldan, J. Lehec, Y. Shenfeld. The logarithmic Sobolev inequality via the Föllmer process (2018).

[8] H. Föllmer. An entropy approach to the time reversal of diffusion processes. In *Stochastic differential systems (Marseille-Luminy, 1984), Lecture Notes in Control and Inform. Sci.* 69, 156–163 (1985). Springer.

[9] J. Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* 49, 885–899 (2013).

[10] B. S. Tsirel'son. A geometric approach to maximum likelihood estimation for an infinite-dimensional Gaussian location. I. *Teor. Veroyatnost. i Primenen.* 27, 388–395 (1982).

[11] C. Villani. *Topics in optimal transportation.* Studies in Mathematics 58. American Mathematical Society (2003).

[12] R. Vitale. The Wills functional and Gaussian processes. *Ann. Probab.* 24, 2172–2178 (1996).

Department of Mathematics

Weizmann Institute of Science, 76100 Rehovot, Israel

roneneldan@gmail.com

Institut de Mathématiques de Toulouse

Université de Toulouse – Paul-Sabatier, F-31062 Toulouse, France

& Institut Universitaire de France

ledoux@math.univ-toulouse.fr