An interpolation proof of Ehrhard's inequality

Joe Neeman   Grigoris Paouris

May 25, 2016

Abstract

We prove Ehrhard's inequality using interpolation along the Ornstein-Uhlenbeck semi-group. We also provide an improved Jensen inequality for Gaussian variables that might be of independent interest.

1   Introduction

In [10], A. Ehrhard proved the following Brunn-Minkowski like inequality for convex sets A, B in $\mathbb{R}^n$:

$$\Phi^{-1}\left(\gamma_n(\lambda A + (1 - \lambda)B)\right) \geq \lambda\Phi^{-1}(\gamma_n(A)) + (1 - \lambda)\Phi^{-1}(\gamma_n(B)), \ \lambda \in [0, 1], \quad (1)$$

where γ_n is the standard Gaussian measure in $\mathbb{R}^n$ (i.e. the measure with density $(2\pi)^{-n/2}e^{-|x|^2/2}$) and Φ is the Gaussian distribution function (i.e. $\Phi(x) = \gamma_1(-\infty, x)$).

This is a fundamental result of Gaussian space and it is known to have numerous applications (see, e.g., [14]). Ehrhard's result was extended by R. Latała [13] to the case that one of the two sets is Borel and the other is convex. Finally, C. Borell [6] proved that it holds for all pairs of Borel sets. Ehrhard's original proof for convex sets used a Gaussian symmetrization technique. Borell used the heat semi-group and a maximal inequality in his proof, which has since been further developed by Barthe and Huet [3]; very recently Ivanisvili and Volberg [12] developed this method into a general technique for proving convolution inequalities. Another proof was recently found by van Handel [17] using a stochastic variational principle.

In this work we will prove Ehrhard's inequality by constructing a quantity that is monotonic along the Ornstein-Uhlenbeck semi-group. In recent years this approach has been developed into a powerful tool to prove Gaussian inequalities such as Gaussian hypercontractivity, the log-Sobolev inequality, and isoperimetry [4]. There is no known proof of Ehrhard inequality using these techniques and the purpose of this note is to fill this gap.

An interpolation proof of the Lebesgue version of Ehrhard's inequality (the Prékopa-Leindler inequality) was presented recently in [9]. This proof uses an "improved reverse Hölder" inequality for correlated Gaussian vectors that was established in [9]. A generalization of the aforementioned inequality also appeared recently [15, 16]. This inequality, while we call an "improved Jensen inequality" for correlated Gaussian vectors, we present and actually also extend in the present note. In §2 we briefly discuss how this inequality implies several known inequalities in probability, convexity and harmonic analysis. Using a "restricted" version of this inequality (Theorem 2.2), we will present a proof of Ehrhard's inequality.

The paper is organized as follows: In §2 we introduce the notation and basic facts about the Ornstein-Uhlenbeck semi-group, and we present the proof of the restricted, improved Jensen inequality. in §3 we use Jensen inequality to provide a new proof of Prékopa-Leindler inequality. We will use the main ideas of this proof as a guideline for our proof of Ehrhard's inequality that we present in §4.

2 An "improved Jensen" inequality

Fix a positive semi-definite $D \times D$ matrix A, and let $X \sim \mathcal{N}(0, A)$. For $t \geq 0$, we define the operator P_t^A on $L_1(\mathbb{R}^D, \gamma_A)$ by

$$(P_t^A f)(x) = \mathbb{E} f(e^{-t} x + \sqrt{1 - e^{-2t}} X).$$

We will use the following well-known (and easily checked) facts:

- the measure γ_A is stationary for P_t^A;

- for any $s, t \geq 0$, $P_s^A P_t^A = P_{s+t}^A$;

- if f is a continuous function having limits at infinity then $P_s^A f$ converges uniformly to $P_t^A f$ as $s \to t$.

We will heavily use the fact that P_t^A commutes in a nice way with the composition of smooth functions: let $\Psi : \mathbb{R}^k \to \mathbb{R}$ be a bounded $\mathcal{C}^2$ function. For any bounded, measurable $f = (f_1, \ldots, f_k) : \mathbb{R}^D \to \mathbb{R}^k$, any $x \in \mathbb{R}^D$ and any $0 < s < t$, $P_{t-s}^A \Psi(P_s^A f(x))$ is differentiable in s and satisfies

$$\frac{\partial}{\partial s} P_{t-s}^A \Psi(P_s^A f) = -P_{t-s}^A \sum_{i,j=1}^k \partial_i \partial_j \Psi(f) \langle \nabla P_s^A f_i, A \nabla P_s^A f_j \rangle. \qquad (2)$$

Suppose that $D = \sum_{i=1}^k d_i$, where $d_i \geq 1$ are integers. We decompose $\mathbb{R}^D$ as $\prod_{i=1}^k \mathbb{R}^{d_i}$ and write Π_i for the projection on the ith component. Given a $k \times k$ matrix M, write $\mathcal{E}_{d_1,\ldots,d_k}(M)$ for the $D \times D$ matrix whose i, j entry is $M_{k,\ell}$ if $\sum_{a<k} d_a < i \leq \sum_{a \leq k} d_a$ and $\sum_{b<\ell} d_b < j \leq \sum_{b \leq \ell} d_b$; that is, each entry $M_{k,\ell}$ of M is expanded into a $d_k \times d_\ell$ block. We write '$\odot$' for the element-wise product of matrices, '$\succcurlyeq$' for the positive semi-definite matrix ordering, and H_J for the Hessian matrix of the function J.

Our starting point in this note is the following inequality, which may be seen as an improved Jensen inequality for correlated Gaussian variables.

Theorem 2.1. *Let $\Omega_1, \ldots, \Omega_k$ be open intervals, and let $\Omega = \prod_{i=1}^{k} \Omega_i$. Take $X \sim \gamma_A$ and write $X_i = \Pi_i X$. For a bounded, $\mathcal{C}^2$ function $J : \Omega \to \mathbb{R}$, the following are equivalent:*

(2.1.a) for every $x \in \Omega$, $A \odot \mathcal{E}_{d_1, \ldots, d_k}(H_J(x)) \succcurlyeq 0$

(2.1.b) for every k-tuple of measurable functions $f_i : \mathbb{R}^{d_i} \to \Omega_i$,

$$\mathbb{E}J(f_1(X_1), \ldots, f_k(X_k)) \geq J(\mathbb{E}f_1(X_1), \ldots, \mathbb{E}f_k(X_k)). \tag{3}$$

We remark that the restriction that J be bounded can often be lifted. For example, if J is a continuous but unbounded function then one can still apply Theorem 2.1 on bounded domains $\Omega_i' \subset \Omega_i$. If J is sufficiently nice (e.g. monotonic, or bounded above) then one can take a limit as Ω_i' exhausts Ω_i (e.g. using the monotone convergence theorem, or Fatou's lemma).

As we have already mentioned, Theorem 2.1 is known to have many consequences. However, we do not know how to obtain Ehrhard's inequality using only Theorem 2.1; we will first need to extend Theorem 2.1 in a few ways. To motivate our first extension, note that the usual Jensen inequality on $\mathbb{R}$ extends easily to the case where some function is convex only on a sub-level set. To be more precise, take a function $\psi : \mathbb{R} \to \mathbb{R}$ and the set $B = \{x \in \mathbb{R}^d : \psi(x) < 0\}$. If B is connected and ψ is convex when restricted to B, then one easily finds that $\mathbb{E}\psi(X) \geq \psi(\mathbb{E}X)$ for any random vector supported on B. A similar modification may be made to Theorem 2.1.

Theorem 2.2. *Take the notation and assumptions of Theorem 2.1, and assume in addition that $\{x \in \Omega : J(x) < 0\}$ is connected. Then the following are equivalent:*

(2.2.a) for every $x \in \Omega$ such that $J(x) < 0$, $A \odot \mathcal{E}_{d_1, \ldots, d_k}(H_J(x)) \succcurlyeq 0$

(2.2.b) for every k-tuple of measurable functions $f_i : \mathbb{R}^{d_i} \to \Omega_i$ that γ_A-a.s. satisfy $J(f_1, \ldots, f_k) < 0$,

$$\mathbb{E}J(f_1(X_1), \ldots, f_k(X_k)) \geq J(\mathbb{E}f_1(X_1), \ldots, \mathbb{E}f_k(X_k)).$$

Note that the threshold of zero in the conditions $J(x) < 0$ and $J(f_1, \ldots, f_k) \leq 0$ is arbitrary, since we may apply the theorem to the function $J(\cdot) - a$ for any $a \in \mathbb{R}$. Of course, taking a sufficiently large recovers Theorem 2.1.

Proof. Suppose that (2.2.a) holds. By standard approximation arguments, it suffices to prove (2.2.b) for a more restricted class of functions f. Indeed, let F be the set of measurable $f = (f_1, \ldots, f_k)$ satisfying $J(f) < 0$ γ_A-a.s. and let $F_\epsilon \subset F$ be

those functions that are continuous, vanish at infinity, and satisfy $J(f) \leq -\epsilon \, \gamma_A$-a.s. Now, every $f \in F$ can be approximated pointwise by a sequence $f^{(n)} \in F_{1/n}$ (here we are using the fact that $\{x : J(x) < 0\}$ is connected); hence, it suffices to prove (2.2.b) for $f \in F_\epsilon$, where $\epsilon > 0$ is arbitrarily small. From now on, fix $\epsilon > 0$ and fix $f = (f_1, \ldots, f_k) \in F_\epsilon$.

Recalling that $\Pi_i : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \to \mathbb{R}^{d_i}$ is the projection onto the ith block of coordinates, define $g_i = f_i \circ \Pi_i$ and $G_{s,t}(x) = P_{t-s}^A J(P_s^A g(x))$. Since $f \in F_\epsilon$, we have $G_{0,0}(x) \leq -\epsilon$ for every $x \in \mathbb{R}^D$. Moreover, since f is continuous and vanishes at infinity, $P_s^A g \to g$ uniformly as $s \to 0$. Since g is bounded, J is uniformly continuous on the range of g and so there exists $\delta > 0$ such that $|G_{s,s}(x) - G_{r,r}(x)| < \epsilon$ for every $x \in R^D$ and every $|s - r| \leq \delta$.

Now, fix $r \geq 0$ and assume that $G_{r,r} \leq -\epsilon$ pointwise; by the previous paragraph, $G_{s,s} < 0$ pointwise for every $r \leq s \leq r + \delta$. Now we apply the commutation formula (2): with $B_s = B_s(x) = A \odot \mathcal{E}_{d_1,\ldots,d_k}(H_J(P_s^A g))$, we have

$$\frac{\partial}{\partial s} G_{s,t} = -P_{t-s}^A \sum_{i,j=1}^k \langle \nabla P_s^A g_i, B \nabla P_s^A g_j \rangle$$

(here, we have used the observation that $P_s^A g_i(x)$ depends only on $\Pi_i x$, and so $\nabla P_s^A g_i$ is zero outside the ith block of coordinates). The assumption (2.2.a) implies that B_s is positive semi-definite whenever $G_{s,s} < 0$; since $G_{s,s} < 0$ for every $s \in [r, r+\delta]$, we see that for such s, $\frac{\partial}{\partial s} G_{s,r+\delta} \leq 0$ pointwise. Since $G_{s,r+\delta}$ is continuous in s and $G_{r,r} \leq -\epsilon$, it follows that $G_{s,s} \leq -\epsilon$ pointwise for all $s \in [r, r + \delta]$.

Next, note that $r = 0$ satisfies the assumption $G_{r,r} \leq -\epsilon$ of the previous paragraph. By induction, it follows that $G_{r,r} \leq -\epsilon$ pointwise for all $r \geq 0$. Hence, the matrix B_s is positive semi-definite for all $s \geq 0$ and $x \in \mathbb{R}^D$, which implies that $G_{s,t}(x)$ is non-increasing in s for all $t \geq s$ and $x \in \mathbb{R}^D$. Hence,

$$\mathbb{E}J(f_1(X_1), \ldots, f_k(X_k)) = \lim_{t \to \infty} G_{0,t}(0) \geq \lim_{t \to \infty} G_{t,t}(0) = J(\mathbb{E}f_1, \ldots, \mathbb{E}f_k).$$

This completes the proof of (2.2.b).

Now suppose that (2.2.b) holds. Choose some $v \in \mathbb{R}^D$ and some $y \in \Omega$ with $J(y) < 0$; to prove (2.2.a), it is enough to show that

$$v^T (A \odot \mathcal{E}_{d_1,\ldots,d_k}(H_J(y)))v \geq 0. \tag{4}$$

Since Ω is open, there is some $\delta > 0$ such that $y + z \in \Omega$ whenever $\max_i |z_i| \leq \delta$. For this δ, define $\psi : \mathbb{R} \to \mathbb{R}$ by

$$\psi(t) = \max\{-\delta, \min\{\delta, t\}\}.$$

For $\epsilon > 0$, define $f_{i,\epsilon} : \mathbb{R}^{d_i} \to \Omega_i$ by

$$f_{i,\epsilon}(x) = y_i + \psi(\epsilon \langle x, \Pi_i v \rangle).$$

By (2.2.b),

$$\mathbb{E}J(f_{1,\epsilon}(X_1), \ldots, f_{k,\epsilon}(X_k)) \geq J(\mathbb{E}f_{1,\epsilon}(X_1), \ldots, \mathbb{E}f_{k,\epsilon}(X_k)).$$

4

Since ψ is odd, $\mathbb{E}f_{i,\epsilon}(X_i) = y_i$ for all $\epsilon > 0$; hence,

$$\mathbb{E}J(f_{1,\epsilon}(X_1), \ldots, f_{k,\epsilon}(X_k)) \geq J(y). \tag{5}$$

Taylor's theorem implies that for any z with $y + z \in \Omega$,

$$J(y+z) = J(y) + \sum_{i=1}^{k} \frac{\partial J(y)}{\partial y_i} z_i + \sum_{i,j=1}^{k} \frac{\partial^2 J(y)}{\partial y_i \partial y_j} z_i z_j + \rho(|z|),$$

where ρ is some function satisfying $\epsilon^{-2}\rho(\epsilon) \to 0$ as $\epsilon \to 0$. Now consider what happens when we replace z_i above with $Z_i = \psi(\epsilon \langle X_i, \Pi_i v \rangle)$ and take expectations. One easily checks that $\mathbb{E}Z_i = 0$, $\mathbb{E}\rho(|Z|) = o(\epsilon^2)$, and

$$\mathbb{E}Z_i Z_j = \epsilon^2 (\Pi_i v)^T \mathbb{E}[X_i X_j](\Pi_i v) + o(\epsilon^2);$$

hence,

$$\mathbb{E}J(y+Z) = J(y) + \epsilon^2 \sum_{i,j=1}^{k} \frac{\partial^2 J(y)}{\partial y_i \partial y_j}(\Pi_i v)^T \mathbb{E}[X_i X_j](\Pi_i v) + o(\epsilon^2)$$

$$= J(y) + \epsilon^2 v^T (A \odot \mathcal{E}_{d_1, \ldots, d_k}(H_J(y)))v + o(\epsilon^2).$$

On the other hand, $\mathbb{E}J(y+Z) = \mathbb{E}J(f_{1,\epsilon}(X_1), \ldots, f_{k,\epsilon}(X_k))$, which is at least $J(y)$ according to (5). Taking $\epsilon \to 0$ proves (4). $\qquad\square$

3 A short proof of Prékopa-Leindler inequality

The Prékopa-Leindler inequality states that if $f, g, h : \mathbb{R}^d \to [0, \infty)$ satisfy

$$h(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}$$

for all $x, y \in \mathbb{R}^d$ and some $\lambda \in (0, 1)$ then

$$\mathbb{E}h \geq (\mathbb{E}f)^\lambda (\mathbb{E}g)^{1-\lambda},$$

where expectations are taken with respect to the standard Gaussian measure on $\mathbb{R}^d$. By applying a linear transformation, the standard Gaussian measure may be replaced by any Gaussian measure; by taking a limit over Gaussian measures with large covariances, the expectations may also be replaced by integrals with respect to the Lebesgue measure.

As M. Ledoux brought to our attention, the Prékopa-Leindler inequality may be seen as a consequence of Theorem 2.1; we will present only the case $d = 1$, but the case for general d may be done in a similar way. Alternatively, one may prove the Prékopa-Leindler inequality for $d = 1$ first and then extend to arbitrary d using induction and Fubini's theorem.

Fix $\lambda \in (0, 1)$, let $(X, Y) \sim \mathcal{N}\big(0, \big(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\big)\big)$ and let $Z = \lambda X + (1 - \lambda)Y$. Let $\sigma^2 = \sigma^2(\rho, \lambda)$ be the variance of Z and let $A = A(\rho, \lambda)$ be the covariance of

(X, Y, Z). Note that A is a rank-two matrix, and that it may be decomposed as $A = uu^T + vv^T$ where u and v are both orthogonal to $(\lambda, 1 - \lambda, -1)^T$.

For $\alpha, R \in \mathbb{R}_+$, define $J_{\alpha,R} : \mathbb{R}_+^3 \to \mathbb{R}$ by

$$J_{\alpha,R}(x, y, z) = (x^\lambda y^{1-\lambda} z^{-\alpha})^R.$$

Lemma 3.1. *For any λ and ρ, and for any $\alpha < \sigma^2$, there exists $R \in \mathbb{R}_+$ such that $A \odot H_{J_{\alpha,R}} \succneqq 0$.*

To see how the Prékopa-Leindler inequality follows from Theorem 2.1 and Lemma 3.1, suppose that $h(\lambda x + (1-\lambda)y) \geq f^\lambda(x) + g^{1-\lambda}(y)$ for all $x, y \in \mathbb{R}$. Then $J_{\alpha,R}(f(X), g(Y), h^{1/\alpha}(Z)) \leq 1$ with probability one (because $Z = \lambda X + (1 - \lambda)Y$ with probability one). By Theorem 2.1, with the R from Lemma 3.1 we have

$$
\begin{aligned}
1 &\geq \mathbb{E} J_{\alpha,R}(f(X), g(Y), h(Z)) \\
&\geq J_{\alpha,R}(\mathbb{E}f(X), \mathbb{E}g(Y), \mathbb{E}h(Z)) \\
&= \left(\frac{(\mathbb{E}f(X))^\lambda (\mathbb{E}g(Y))^{1-\lambda}}{(\mathbb{E}h^{1/\alpha}(Z))^\alpha} \right)^R.
\end{aligned}
$$

In other words, $(\mathbb{E}h^{1/\alpha}(Z))^\alpha \geq (\mathbb{E}f)^\lambda (\mathbb{E}g)^{1-\lambda}$. This holds for any ρ and any $\alpha < \sigma^2$. By sending $\rho \to 1$, we send $\sigma^2 \to 1$ and so we may take $\alpha \to 1$ also. Finally, note that in this limit Z converges in distribution to $\mathcal{N}(0, 1)$. Hence, we recover the Prékopa-Leindler inequality for the standard Gaussian measure.

Proof of Lemma 3.1. By a computation,

$$
H_{J_{\alpha,R}} = J_{\alpha,R}(x, y, z) \begin{pmatrix} \frac{\lambda R(\lambda R - 1)}{x^2} & \frac{\lambda R(1-\lambda)R}{xy} & -\frac{\lambda \alpha R^2}{xz} \\ \frac{\lambda R(1-\lambda)R}{xy} & \frac{(1-\lambda)R((1-\lambda)R-1)}{y^2} & -\frac{(1-\lambda)\alpha R^2}{yz} \\ -\frac{\lambda \alpha R^2}{xz} & -\frac{(1-\lambda)\alpha R^2}{yz} & \frac{\alpha R(\alpha R + 1)}{z^2} \end{pmatrix}.
$$

We would like to show that $A \odot H_J \succneqq 0$; since elementwise multiplication commutes with multiplication by diagonal matrices, it is enough to show that

$$
A \odot \left(\begin{pmatrix} \lambda \\ 1 - \lambda \\ -\alpha \end{pmatrix}^{\otimes 2} - \frac{1}{R} \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & -\alpha \end{pmatrix} \right) \geq 0. \tag{6}
$$

Let $\theta = (\lambda, 1 - \lambda, -\alpha)^T$ and recall that $A = uu^T + vv^T$, where u and v are both orthogonal to $(\lambda, 1 - \lambda - 1)^T$. Then

$$A \odot (\theta\theta^T) = (u \odot \theta)(u \odot \theta)^T + (v \odot \theta)(v \odot \theta)^T,$$

where $u \odot \theta$ and $v \odot \theta$ are both orthogonal to $(1, 1, \frac{1}{\alpha})^T$ (call this w). In particular, $A \odot (\theta\theta^T)$ is a rank-two, positive semi-definite matrix whose null space is the span of w.

On the other hand, $A \odot \mathrm{diag}(\lambda, 1 - \lambda, -\alpha) = \mathrm{diag}(\lambda, 1 - \lambda, -\alpha\sigma^2)$ (call this D). Then $w^T D w = 1 - \sigma^2/\alpha < 0$. As a consequence of the following Lemma,

$$A \circ (\theta\theta^T) - \frac{1}{R} D \geq 0$$

for all sufficiently large R. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma 3.2. *Let A be a positive semi-definite matrix and let B be a symmetric matrix. If $u^T B u \geq \delta|u|^2$ for all $u \in \ker(A)$ and $v^T A v \geq \delta|v|^2$ for all $v \in \ker(A)^\perp$ then $A + \epsilon B \succcurlyeq 0$ for all $0 \leq \epsilon \leq \frac{\delta^2}{\|B\|^2 + \delta\|B\|}$, where $\|B\|$ is the operator norm of B.*

Proof. Any vector w may be decomposed as $w = u + v$ with $u \in \ker(A)$ and $v \in \ker(A)^\perp$. Then

$$w^T(A + \epsilon B)w = u^T A u + \epsilon u^T B u - 2\epsilon u^T B v + \epsilon v^T B v$$
$$\geq \delta|u|^2 - \epsilon\|B\||u|^2 - 2\epsilon\|B\||u||v| + \epsilon\delta|v|^2.$$

Considering the above expression as a quadratic polynomial in $|u|$ and $|v|$, we see that it is non-negative whenever $(\delta - \epsilon\|B\|)\delta \geq \epsilon\|B\|^2$. $\qquad\qquad$ $\square$

We remark that the preceding proof of the Prékopa-Leindler inequality may be extended in an analogous way to prove Barthe's inequality [2].

4   Proof of Ehrhard's inequality

The parallels between the Prékopa-Leindler and Ehrhard inequalities become obvious when they are both written in the following form. The version of Prékopa-Leindler that we proved above may be restated to say that

$$\left.\begin{array}{c} \exp(R(\lambda \log f(X) + (1 - \lambda) \log g(Y) - \alpha \log h(Z))) \leq 0 \text{ a.s.} \\ \text{implies} \\ \exp(R(\lambda \log \mathbb{E}f(X) + (1 - \lambda) \log \mathbb{E}g(Y) - \alpha \log \mathbb{E}h(Z))) \leq 0. \end{array}\right\} \qquad (7)$$

On the other hand, here we will prove that

$$\left.\begin{array}{c} \Phi\left(R(\lambda\Phi^{-1}(f(X)) + (1 - \lambda)\Phi^{-1}(g(Y)) - \sigma\Phi^{-1}(h(Z)))\right) \leq 0 \text{ a.s.} \\ \text{implies} \\ \Phi\left(R(\lambda\Phi^{-1}(\mathbb{E}f(X)) + (1 - \lambda)\Phi^{-1}(\mathbb{E}g(Y)) - \sigma\Phi^{-1}(\mathbb{E}h(Z)))\right) \leq 0. \end{array}\right\} \qquad (8)$$

(It may not yet be clear why the α in (7) has become σ in (8); this turns out to be the right choice, as will become clear from the example in Section 4.1.) This implies Ehrhard's inequality in the same way that (7) implies the Prékopa-Leindler inequality. In particular, our proof of (7) suggests a strategy for attacking (8): define the function

$$J_R(x, y, z) = \Phi\left(R(\lambda\Phi^{-1}(x) + (1 - \lambda)\Phi^{-1}(y) - \sigma\Phi^{-1}(z))\right).$$

7

(We will drop the parameter R when it can be inferred from the context.) In analogy with our proof of Prékopa-Leindler, we might then try to show that for sufficiently large R, $A \odot H_{J_R} \succcurlyeq 0$. Unfortunately, this is false.

4.1 An example

Recall from the proof of Theorem 2.2 that if $A \odot H_J \succcurlyeq 0$ then

$$G_{s,t,R}(x,y) := P_{t-s}^A J_R(P_s^1 f(x), P_s^1 g(y), P_s^{\sigma^2} h(\lambda x + (1-\lambda)y))$$

is non-increasing in s for every x and y. We will give an example in which $G_{s,t,R}$ may be computed explicitly and it clearly fails to be non-increasing.

From now on, define $f_s = P_s^1 f$, $g_s = P_s^1 g$ and $h_s = P_s^{\sigma^2} h$. Let $f(x) = 1_{\{x \leq a\}}$, $g(y) = 1_{\{y \leq b\}}$ and $h(z) = 1_{\{z \leq c\}}$, where $c \geq \lambda a + (1-\lambda)b$. A direct computation yields

$$f_s(x) = \Phi\left(\frac{a - e^{-s}x}{\sqrt{1 - e^{-2s}}}\right)$$

$$g_s(y) = \Phi\left(\frac{b - e^{-s}y}{\sqrt{1 - e^{-2s}}}\right)$$

$$h_s(z) = \Phi\left(\frac{c - e^{-s}z}{\sigma\sqrt{1 - e^{-2s}}}\right).$$

Hence,

$$J(f_s(x), g_s(y), h_s(\lambda x + (1-\lambda y))) = \Phi\left(R\frac{\lambda a + (1-\lambda)b - c}{\sqrt{1 - e^{-2s}}}\right).$$

If $c > \lambda a + (1-\lambda)b$ then the above quantity is increasing in s. Since it is also independent of x and y, it remains unchanged when applying P_{t-s}^A. That is,

$$G_{s,t,R} = \Phi\left(R\frac{\lambda a + (1-\lambda)b - c}{\sqrt{1 - e^{-2s}}}\right)$$

is increasing in s. On the bright side, in this example $G_{s,r,R\sqrt{1-e^{-2s}}}$ is constant. Since Theorem 2.1 was not built to consider such behavior, we will adapt it so that the function J is allowed to depend on s.

4.2 Allowing J to depend on t

Recalling the notation of §2, we assume from now on that $\Omega_i \subseteq [0,1]$ for each i. Then A is a $k \times k$ matrix; let $\sigma_1^2, \ldots, \sigma_k^2$ be its diagonal elements. We will consider functions of the form $J : \Omega \times [0, \infty] \to \mathbb{R}$. We write H_J for the Hessian matrix of J with respect to the variables in Ω, and $\frac{\partial J}{\partial t}$ for the partial derivative of J with respect to the variable in $[0, \infty]$. Let $I : [0,1] \to \mathbb{R}$ be the function $I(x) = \phi(\Phi^{-1}(x))$.

8

Lemma 4.1. *With the notation above, suppose that $J : \Omega \times [0, \infty] \to \mathbb{R}$ is bounded and $\mathcal{C}^2$, and take $(X_1, \ldots, X_k) \sim \gamma_A$. Let $\lambda_1, \ldots, \lambda_k$ be non-negative numbers with $\sum_i \lambda_i = 1$, let $D(x)$ be the $k \times k$ diagonal matrix with $\lambda_i \sigma_i^2 / I^2(x_i)$ in position i, and take some $\epsilon \geq 0$. If $\frac{\partial J}{\partial t}(x, t) \leq 0$ and*

$$A \odot H_J(x, t) - (e^{2(t+\epsilon)} - 1) \frac{\partial J(x, t)}{\partial t} D^2 \succcurlyeq 0 \tag{9}$$

for every $x \in \Omega$ and $t > 0$ then for every k-tuple of measurable functions $f_i : \mathbb{R} \to \Omega_i$,

$$\mathbb{E}J(P_\epsilon^{\sigma_1} f_1(X_1), \ldots, P_\epsilon^{\sigma_k} f_k(X_k), 0) \geq J(\mathbb{E}f_1(X_1), \ldots, \mathbb{E}f_k(X_k), \infty). \tag{10}$$

Note that Lemma 4.1 has an extra parameter $\epsilon \geq 0$ compared to our previous versions of Jensen's inequality. This is for convenience when applying Lemma 4.1: when $\epsilon > 0$ then the function $e^{2(t+\epsilon)} - 1$ is bounded away from zero, which makes (9) easier to check.

Proof. Write $f_{i,s}$ for $P_{s+\epsilon}^{\sigma_i^2} f_i$ and $f_s = (f_{1,s}, \ldots, f_{k,s})$. Define

$$G_{s,t} = P_{t-s-\epsilon}^A J(f_{1,s}, \ldots, f_{k,s}, s).$$

We differentiate in s, using the commutation formula (2). Compared to the proof of Theorem 2.2, an extra term appears because the function J itself depends on s:

$$-\frac{\partial}{\partial s} G_{s,t} = P_{t-s-\epsilon} \sum_{i,j=1}^k \partial_i \partial_j J(f_s, s) A_{ij} f_{i,s}' f_{j,s}' - P_{t-s-\epsilon} \frac{\partial J}{\partial s}(f_s, s)$$

$$= P_{t-s-\epsilon} v_s^T (A \odot H_J(f_s, s)) v_s - P_{t-s-\epsilon} \frac{\partial J}{\partial s}(f_s, s),$$

where $v_s = \nabla f_s$. Bakry and Ledoux [1] proved that $|v_{i,s}| \leq \sigma_i^{-1} (e^{2(s+\epsilon)} - 1)^{-1/2} I(f_{i,s})$. Hence,

$$v_s^T D(f_s) v_s = \sum_{i=1}^k \lambda_i \left(\frac{\sigma_i |v_{i,s}'|}{I(f_{i,s})} \right)^2 \leq (e^{2(s+\epsilon)} - 1)^{-1},$$

and so

$$-\frac{\partial}{\partial s} G_{s,t} \geq P_{t-s} \left(v_s^T (A \odot H_J(f_s, s)) v_s - (e^{2(s+\epsilon)} - 1) \frac{\partial J}{\partial s}(f_s, s) v_s^T D(f_s) v_s \right).$$

Clearly, the argument of P_{t-s} is non-negative pointwise if

$$A \odot H_{J_R}(x, s) \succcurlyeq (e^{2(s+\epsilon)} - 1) \frac{\partial J_R}{\partial s}(x, s) D(x)$$

for all x, s. In this case, $G_{s,t}$ is non-increasing in s and we conclude as in the proof of Theorem 2.2. $\square$

By combining the ideas of Theorem 2.2 and Lemma 4.1, we obtain the following combined version.

Corollary 4.2. *With the notation and assumptions of Lemma 4.1, suppose in addition that $\{x \in \Omega : J(x, 0) < a\}$ is connected, that $\frac{\partial J(x,t)}{\partial t} \leq 0$ whenever $J(x, t) < a$, and that*

$$A \odot H_J(x, t) - (e^{2(t+\epsilon)} - 1)\frac{\partial J(x,t)}{\partial t}D^2 \succcurlyeq 0$$

for every $t \geq 0$ and every x such that $J(x, t) < a$. Then for every k-tuple of measurable functions $f_i : \mathbb{R} \to \Omega_i$ satisfying $J(P_\epsilon^{\sigma_1} f_1, \ldots, P_\epsilon^{\sigma_k} f_k) < 0$,

$$\mathbb{E}J(P_\epsilon^{\sigma_1} f_1(X_1), \ldots, P_\epsilon^{\sigma_k} f_k(X_k), 0) \geq J(\mathbb{E}f_1(X_1), \ldots, \mathbb{E}f_k(X_k), \infty).$$

4.3 The Hessian of J

Define $J_R : (0, 1)^3 \to 0$ by

$$J_R(x, y, z) = \Phi\left(R\left(\lambda\Phi^{-1}(x) + (1 - \lambda)\Phi^{-1}(y) - \sigma\Phi^{-1}(z)\right)\right).$$

Let $H_J = H_J(x, y, z)$ denote the 3×3 Hessian matrix of J; let A be the 3×3 covariance matrix of (X, Y, Z). In order to apply Corollary 4.2, we will compute the matrix $A \odot H_J$. First, we define some abbreviations: set

$$u = \Phi^{-1}(x) \qquad\qquad \Xi = \lambda u + (1 - \lambda)v - \sigma w$$
$$v = \Phi^{-1}(y) \qquad\qquad \theta = (\lambda, 1 - \lambda, -\sigma)^T$$
$$w = \Phi^{-1}(z) \qquad\qquad \mathcal{I} = \mathrm{diag}(\phi(u), \phi(v), \phi(w))$$

We will use a subscript s to denote that any of the above quantities is evaluated at (f_s, g_s, h_s) instead of (x, y, z). That is $u_s = \Phi^{-1}(f_s)$, $\Xi_s = \lambda u_s + (1 - \lambda)v_s - \sigma w_s$, and so on.

Lemma 4.3. $H_J = \phi(R\Xi)\mathcal{I}^{-1}\left(R\,\mathrm{diag}(\lambda u, (1 - \lambda)v, -\sigma w) - R^3\Xi\theta\theta^T\right)\mathcal{I}^{-1}.$

Proof. Noting that $\frac{du}{dx} = 1/\phi(u)$, the chain rule gives

$$\frac{d}{dx}\Phi(R\Xi) = R\lambda\frac{\phi(R\Xi)}{\phi(u)} = R\lambda\exp\left(-\frac{R^2\Xi^2 - u^2}{2}\right).$$

Differentiating again,

$$\frac{d^2}{dx^2}\Phi(R\Xi) = R\lambda(u - R^2\Xi\lambda)\frac{\phi(R\Xi)}{\phi^2(u)}.$$

For cross-derivatives,

$$\frac{d^2}{dxdy}\Phi(R\Xi) = -R^3\Xi\lambda(1 - \lambda)\frac{\phi(R\Xi)}{\phi(u)\phi(v)}.$$

Putting these together with the analogous terms involving differentiation by z,

$$\frac{H_J}{\phi(R\Xi)} = -R^3\Xi \begin{pmatrix} \frac{\lambda^2}{\phi^2(u)} & \frac{\lambda(1-\lambda)}{\phi(u)\phi(v)} & -\frac{\lambda\sigma}{\phi(u)\phi(w)} \\ \frac{\lambda(1-\lambda)}{\phi(u)\phi(v)} & \frac{(1-\lambda)^2}{\phi^2(v)} & -\frac{(1-\lambda)\sigma}{\phi(v)\phi(w)} \\ -\frac{\lambda\sigma}{\phi(u)\phi(w)} & -\frac{(1-\lambda)\sigma}{\phi(u)\phi(v)} & \frac{\sigma^2}{\phi^2(w)} \end{pmatrix}$$

$$+ R \begin{pmatrix} \frac{\lambda u}{\phi^2(u)} & 0 & 0 \\ 0 & \frac{(1-\lambda)v}{\phi^2(v)} & 0 \\ 0 & 0 & -\frac{\sigma w}{\phi^2(w)} \end{pmatrix}.$$

Recalling the definition of $\mathcal{I}$ and θ, this may be rearranged into the claimed form. $\square$

Having computed H_J, we need to examine $A \odot H_J$. Recall that A is a rank-two matrix and so it may be decomposed as $A = aa^T + bb^T$. Moreover, the fact that $Z = \lambda X + (1-\lambda)Y$ means that a and b are both orthogonal to $(\lambda, 1-\lambda, -1)^T$. Recalling the definition of θ, this implies that $a \odot \theta$ and $b \odot \theta$ are both orthogonal to $(1, 1, \sigma^{-1})^T$. This observation allows us to deal with the $\theta\theta^T$ term in Lemma 4.3:

$$A \odot \theta\theta^T = (aa^T) \odot (\theta\theta^T) + (bb^T) \odot (\theta\theta^T) = (a \odot \theta)^{\otimes 2} + (b \odot \theta)^{\otimes 2}.$$

To summarize:

Lemma 4.4. *The matrix $B := A \odot \theta\theta^T$ is positive semidefinite and has rank two. Its kernel is the span of $(1, 1, \frac{1}{\sigma})^T$.*

On the other hand, the diagonal entries of A are $1, 1$, and σ^2; hence,

$$A \odot \mathrm{diag}(\lambda u, (1-\lambda)v, -\sigma w) = \mathrm{diag}(\lambda u, (1-\lambda)v, -\sigma^3 w) =: D.$$

Combining this with Lemma 4.3, we have

$$A \odot H_J = R\phi(R\Xi)\mathcal{I}^{-1}(D - R^2\Xi B)\mathcal{I}^{-1}. \tag{11}$$

Consider the expression above in the light of our earlier proof of Prékopa-Leindler. Again, we have a sum of two matrices (D and $-R^2\Xi B$), one of which is multiplied by a factor (R^2) that we may take to be large. There are two important differences. The first is that the matrix D (whose analogue was constant in the proof of Prékopa-Leindler) cannot be controlled pointwise in terms of B. This difference is closely related to the example in Section 4.1; we will solve it by making J depend on t in the right way; the $\frac{dJ}{dt}$ term in Corollary 4.2 will then cancel out part of D's contribution.

The second difference is that in (11), the term that is multiplied by a large factor (namely, $-\Xi B$) is not everywhere positive semi-definite because there exist $(x, y, z) \in \mathbb{R}^3$ such that $\Xi(x, y, z) > 0$. This is the reason that we consider the "restricted" formulation of Jensen's inequality in Theorem 2.2 and Corollary 4.2.

11

4.4 Adding the dependence on t

Recall that X and Y have variance 1 and covariance ρ, that $Z = \lambda X + (1-\lambda)Y$, and that A is the covariance of (X, Y, Z). Recall also the notations u, v, w, Ξ, and their subscripted variants. For $R > 0$, define $r(t) = R\sqrt{1 - e^{-2t-\epsilon}}$ and

$$J_R(x, y, z, t) = \Phi\left(r(t)\left(\lambda\Phi^{-1}(x) + (1-\lambda)\Phi^{-1}(y) - \sigma\Phi^{-1}(z)\right)\right)$$
$$= \Phi(r(t)\Xi). \tag{12}$$

Let $E = \mathrm{diag}(\lambda, 1-\lambda, \sigma)/(1 + \sigma^{-1})$.

Lemma 4.5. *Define $\Omega_\epsilon = [\Phi(-1/\epsilon), \Phi(1/\epsilon)]^3$. For every ρ, λ, and ϵ, there exists $R > 0$ such that*

$$A \odot H_J - (e^{2(t+\epsilon)} - 1)\frac{\partial J}{\partial t}\mathcal{I}^{-1}E\mathcal{I}^{-1} \succcurlyeq 0$$

on $\{(x, t) \in \Omega_\epsilon \times [0, \infty) : \Xi(x) \leq -\epsilon\}$.

Proof. We computed $A \odot H_J$ in (11) already; applying that formula and noting that $\mathcal{I}^{-1} \succcurlyeq 0$, it suffices to show that

$$r(t)\phi(r(t)\Xi)(D - r^2(t)\Xi B) - (e^{2(t+\epsilon)} - 1)\frac{\partial J}{\partial t}E \succcurlyeq 0$$

whenever $\Xi \leq -\epsilon$. (Recall that $D = \mathrm{diag}(\lambda u, (1-\lambda)v, -\sigma^3 w)$, and that B is a rank-two positive semidefinite matrix that depends only on ρ and λ, and whose kernel is the span of $(1, 1, \sigma^{-1})^T$.) We compute

$$\frac{\partial J}{\partial t} = r'(t)\Xi\phi(r(t)\Xi) = \frac{r(t)}{e^{2t+\epsilon} - 1}\Xi\phi(r(t)\Xi).$$

Now, there is some $\delta = \delta(\epsilon) > 0$ such that

$$\frac{e^{2(t+\epsilon)} - 1}{e^{2t+\epsilon} - 1} \geq 1 + \delta$$

for all $t \geq 0$. For this δ,

$$r(t)\phi(r(t)\Xi)(D - r^2(t)\Xi B) - (e^{2(t+\epsilon)} - 1)\frac{\partial J}{\partial t}E$$
$$\succcurlyeq r(t)\phi(r(t)\Xi)(D - (1+\delta)\Xi E - r^2(t)\Xi B);$$

Hence, it suffices to show that $D - (1+\delta)\Xi E - r^2(t)\Xi B \succcurlyeq 0$. Since $\Xi \leq -\epsilon$, it suffices to show that $r^2(t)\epsilon B + D - (1+\delta)\Xi E \succcurlyeq 0$. Now, B is a rank-two positive semi-definite matrix depending only on λ and ρ. Its kernel is spanned by $\theta = (1, 1, \sigma^{-1})^T$. Note that $\theta^T D\theta = \Xi$ and $\theta^T E\theta = 1$. Hence,

$$\theta^T(D - (1+\delta)\Xi E)\theta = -\delta\Xi \geq \delta\epsilon > 0.$$

12

Next, note that we can bound the norm of $D - (1 + \delta)\Xi E$ uniformly: on Ω_ϵ, $\|D\| \le 1/\epsilon$ and $|\Xi| \le 2/\epsilon$. All together, if we assume (as we may) that $\delta \le 1$ then $\|D + (1 + \delta)\Xi E\| \le 5/\epsilon$. By Lemma 3.2, if $\eta > 0$ is sufficiently small then

$$\epsilon B + \eta(D - (1 + \delta)\Xi E) \succcurlyeq 0.$$

To complete the proof, choose R large enough so that $R^2(1 - e^\epsilon) \ge 1/\eta$; then $r^2(t) \ge 1/\eta$ for all t. $\qquad\square$

Finally, we complete the proof of (8) by a series of simple approximations. First, let C_a denote the set of continuous functions $\mathbb{R} \to [0, 1]$ that converge to a at $\pm\infty$, and note that it suffices to prove (8) in the case that $f, g \in C_0$ and $g \in C_1$. Indeed, any measurable $f, g : \mathbb{R} \to [0, 1]$ may be approximated (pointwise at γ_1-almost every point) from below by functions in C_0, and any measurable $h : \mathbb{R} \to [0, 1]$ may be approximated from above by functions in C_1. If we can prove (8) for these approximations, then it follows (by the dominated convergence theorem) for the original f, g, and h.

Now consider $f, g \in C_0$ and $h \in C_1$ satisfying $\Xi(f, g, h) \le 0$ pointwise. For $\delta > 0$, define

$$f_\delta = \Phi(-1/\delta) \vee f \wedge \Phi(1/(3\delta))$$
$$g_\delta = \Phi(-1/\delta) \vee g \wedge \Phi(1/(3\delta))$$
$$h_\delta = \Phi\left(-\frac{1}{3\delta} \vee (\Phi^{-1}(h) + \delta) \wedge \frac{1}{\delta}\right).$$

If $\delta > 0$ is sufficiently small then $\Xi(f_\delta, g_\delta, h_\delta) \le -\delta$ pointwise; moreover, f_δ, g_δ, and h_δ all take values in $[\Phi(-1/\delta), \Phi(1/\delta)]$, are continuous, and have limits at $\pm\infty$. Since $f_\delta \to f$ as $\delta \to 0$ (and similarly for g and h), it suffices to show that

$$\lambda\Phi^{-1}(\mathbb{E}f_\delta) + (1 - \lambda)\Phi^{-1}(\mathbb{E}g_\delta) \le \sigma\Phi^{-1}(\mathbb{E}h_\delta) \tag{13}$$

for all sufficiently small $\delta > 0$.

Since f_δ has limits at $\pm\infty$, it follows that $P_\epsilon f_\delta \to f_\delta$ uniformly as $\epsilon \to 0$ (similarly for g_δ and h_δ). By taking ϵ small enough (at least as small as $\delta/2$), we can ensure that $\Xi(P_\epsilon^1 f_\delta, P_\epsilon^1 g_\delta, P_\epsilon^{\sigma^2} h_\delta) < -\epsilon$ pointwise. Now we apply Corollary 4.2 with $\Omega_i = [\Phi(-1/\epsilon), \Phi(1/\epsilon)]$, the function J defined in (12), $a = \frac{1}{2}$, and with $(\lambda_1, \lambda_2, \lambda_3) = (\lambda, 1 - \lambda, \sigma^{-1})/(1 + \sigma^{-1})$. Lemma 4.5 implies that the condition of Corollary 4.2 is satisfied. We conclude that

$$\frac{1}{2} \ge J_R(\mathbb{E}f_\delta, \mathbb{E}g_\delta, \mathbb{E}h_\delta, \infty)$$
$$= \Phi\left(R\left(\lambda\Phi^{-1}(\mathbb{E}f_\delta) + (1 - \lambda)\Phi^{-1}(\mathbb{E}g_\delta) - \sigma\Phi^{-1}(\mathbb{E}h_\delta)\right)\right),$$

which implies (13) and completes the proof of (8).

Acknowledgements

References

[1] D. Bakry and M. Ledoux, *Levy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator*, Invent. Math. 123, (1996), no. 1, 259–281.

[2] F. Barthe, *On a reverse form of the Brascamp-Lieb inequality*, Invent. Math. 134 (1998), 335–361.

[3] F. Barthe and N. Huet, *On Gaussian Brunn-Minkowski inequalities*, Studia Math. 191 (2009), no. 3, 283–304.

[4] D. Bakry, I. Gentil and M. Ledoux, *Analysis and geometry of Markov diffusion operators*. Springer, 2013.

[5] C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. 30, 2 (1975), 207–216.

[6] C. Borell, *The Ehrhard inequality*, C.R. Math. Acad. Sci. Paris 337 (2003), no. 10, 663–666.

[7] C. Borell, *Minkowski sums and Brownian exit times*, Ann. Fac. sci. Toulouse: Math., 16, no. 1 (2007), pp. 37–47.

[8] E. A. Carlen, E. H. Lieb and M. Loss, *A sharp analog of Young's inequality on S^N and related entropy inequalities*, J. Geom. Anal., 14, no. 3, pp. 487–520.

[9] W-K. Chen, N. Dafnis and G. Paouris, *Improved Holder and reverse Holder inequalities for Gaussian random vectors*, Adv. Math. 280 (2015), 643–689.

[10] A. Ehrhard, *Symétrisation dans l'espace de Gauss*, Math. Scand. 53 (1983) 281–301.

[11] R. Gardner, *The Brunn-Minkowski inequality*, Bull. Amer. Math. Soc. 39, (2002) pp. 355–405.

[12] P. Ivanisvili and A. Volberg, *Bellman partial differential equation and the hill property for classical isoperimetric problems*, ArXiv preprint `arXiv:1506.03409`.

[13] R. Latala, *A note on the Ehrhard inequality*, Studia Math. 118 (1996), 169–174.

[14] R. Latala, *On some inequalities for Gaussian measures*, Proc. ICM (2002), vol. II, 813–822.

[15] M. Ledoux, *Remarks on Gaussian noise stability, Brascamp-Lieb and Slepian inequalities*, Geom. Aspects Func. Anal. (2014), 309–333.

[16] J. Neeman, *A multi-dimensional version of noise stability*, Electron. Commun. Probab. 19 (2014), no. 72, 1–10.

[17] R. van Handel, *The Borell-Ehrhard Game*, ArXiv preprint `arXiv:1605.00285`.

Joe Neeman
Institute of Applied Mathematics
University of Bonn
Bonn, 53113 Germany
E-mail: `joeneeman@gmail.com`

Grigoris Paouris
Department of Mathematics
Texas A & M University
College Station, TX 77843 U.S.A.
E-mail: `grigorios.paouris@gmail.com`, `grigoris@math.tamu.edu`