Electron. Commun. Probab. **19** (2014), no. 72, 1–10.
DOI: 10.1214/ECP.v19-3005
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A multidimensional version of noise stability*

Joe Neeman†

Abstract

We give a multivariate generalization of Borell's noise stability theorem for Gaussian vectors. As a consequence we recover two inequalities, also due to Borell, for exit times of the Ornstein-Uhlenbeck process.

Keywords: isoperimetric inequality; noise stability; Gaussian process; exit time; hitting time.
AMS MSC 2010: 60G15.
Submitted to ECP on September 10, 2013, final version accepted on October 16, 2014.
Supersedes arXiv:1307.8082.

1 Introduction

There has been a recent flurry of activity in probability [14, 12] and computer science [9, 10, 15] around a certain paper of Borell [3] on inequalities satisfied by the Ornstein-Uhlenbeck process. In his paper, Borell proved a theorem, which is somewhat complicated to state, showing that certain quantities only decrease under Ehrhard symmetrization [7]. He then derived two simpler corollaries, about hitting times for the Ornstein-Uhlenbeck process, from this general result.

We recall that the Ornstein-Uhlenbeck process on $\mathbb{R}^n$ is the Gaussian process $\{X_t : t \in \mathbb{R}\}$ with mean zero and covariance $\mathbb{E}X_s X_t^T = e^{|t-s|} I_n$. This is a Markov process, as may be seen by the construction $X_t = e^{-t} B_{e^{2t}}$ for a Brownian motion B_t, and the stationary measure of X_t is the standard Gaussian measure, γ_n. For a measurable set $A \subset \mathbb{R}^n$, we denote its exit time under X_t by $e_A = \inf\{t \geq 0 : X_t \notin A\}$. We will assume throughout that all sets denoted by A or A_i are measurable, and we will also assume that the probability space underlying X_t is complete, i.e. that its σ-algebra includes all subsets of zero-measure sets. These hypotheses ensure that e_A is also measurable.

Although they were originally written in terms of hitting times instead of exit times, Borell's two corollaries of his general inequality may be written as follows, in which *half-space* means a set of the form $\{x \in \mathbb{R}^n : x \cdot a \leq b\}$, and half-spaces are *parallel* if they have the same normal vector a.

Theorem 1.1 (Borell)**.** *If $B \subset \mathbb{R}^n$ is a half-space with $\gamma_n(B) = \gamma_n(A)$ then e_B stochastically dominates e_A; i.e., for every $t \geq 0$,*

$$\Pr(e_A \geq t) \leq \Pr(e_B \geq t).$$

Theorem 1.2 (Borell)**.** *If B_1 and B_2 are parallel half-spaces with $\gamma_n(B_i) = \gamma_n(A_i)$ then*

$$\mathbb{E} \int_0^{t \wedge e_{A_1}} 1_{A_2}(X_s)\, ds \leq \mathbb{E} \int_0^{t \wedge e_{B_1}} 1_{B_2}(X_s)\, ds.$$

*Partially supported by DTRA grant HDTRA1-13-1-0024
†University of Texas, Austin, United States of America. E-mail: joeneeman@gmail.com

There is a third corollary of Borell's general inequality that did not appear in his original paper [3], but has nevertheless become widely applied in theoretical computer science, particularly in the study of hardness of approximation.

Theorem 1.3 (Borell)**.** *If B_1 and B_2 are parallel half-spaces with $\gamma_n(B_i) = \gamma_n(A_i)$ then for any $t > 0$,*

$$\Pr(X_0 \in A_1, X_t \in A_2) \leq \Pr(X_0 \in B_1, X_t \in B_2).$$

In the special case $A_1 = A_2$, this inequality is sometimes interpreted as showing that half-spaces are the most "noise stable" sets. Here, we think of X_t as being a noisy version of X_0, and so a set A is noise stable if the event $\{X_0 \in A\}$ tends to agree with the event $\{X_t \in A\}$. Theorem 1.3 implies that this overlap is maximized, over all sets with a fixed Gaussian measure, by half-spaces. Using an invariance principle, Mossel et al. [14] deduced from Theorem 1.3 a similar inequality on the discrete cube (although the statement on the cube is necessarily more complicated, because the direction of a half-space's normal vector becomes important); that work then laid the foundation for many applications in theoretical computer science (for a few examples, see [9, 10, 15]).

Note that in Theorem 1.3, the joint distribution of (X_0, X_t) has mean zero and covariance $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right) \otimes I_n$, where $\rho = e^{-t}$. Our main result is a multivariate generalization of Theorem 1.3, which allows for more than two Gaussian vectors and a more general covariance structure than that endowed by the Ornstein-Uhlenbeck process. Specifically, we consider a collection $(X_1, \ldots, X_k)$ of n-dimensional Gaussian vectors, where each X_i has covariance I_n and the pair X_i, X_j has covariance $m_{ij} I_n$ for some $m_{ij} \geq 0$. Our generalization is also strong enough to recover Theorems 1.1 and 1.2; we thank M. Ledoux for pointing this out.

Theorem 1.4. *For some $k \geq 2$ and $n \geq 1$, let $M = (m_{ij})$ be a $k \times k$ positive semidefinite matrix with $m_{ij} \geq 0$, and let $X = (X_1, \ldots, X_k)$ be a kn-dimensional Gaussian vector with covariance $M \otimes I_n$. For any measurable $A_1, \ldots, A_k \subset \mathbb{R}^n$,*

$$\Pr(X_i \in A_i \text{ for all } i) \leq \Pr(X_i \in B_i \text{ for all } i) \tag{1.1}$$

whenever B_i is a collection of parallel half-spaces with $\gamma_n(B_i) = \gamma_n(A_i)$.

Note that Theorem 1.4 characterizes the matrices M for which (1.1) holds, in the sense that if some $m_{ij} < 0$ then there exist sets $A_1, \ldots A_k$ for which (1.1) fails. In particular, one can take A_i and A_j to be anti-parallel half-spaces (meaning that their normals are negative scalar multiples of one another) and take $A_\ell = \mathbb{R}^n$ for $\ell \notin \{i, j\}$.

By setting $k = 2$, Theorem 1.4 recovers Theorem 1.3. We should remark that a weaker version of Theorem 1.4 was obtained by Isaksson and Mossel [8], who showed that the inequality (1.1) also holds under the hypothesis that the off-diagonal elements of M^{-1} are non-positive. We acknowledge A. Sen for pointing out that their condition on M implies the one in Theorem 1.4. In any case, the condition of Isaksson and Mossel would still suffice for recovering Theorems 1.1 and 1.2.

1.1 The equality cases of Theorem 1.4

Next, we consider the configurations $A_1, \ldots, A_k$ that achieve equality in (1.1). If $A_1, \ldots, A_k$ are parallel half-spaces then equality is clearly attained. However, there are other configurations that also attain equality. For example, suppose that $X_0, \ldots, X_i$ are independent of $X_{i+1}, \ldots X_k$. Then any configuration where $A_1, \ldots A_i$ are parallel half-spaces and $A_{i+1}, \ldots, A_k$ are parallel half-spaces achieves equality in (1.1) (but A_1 and A_{i+1} need not be parallel).

It turns out that if M is strictly positive definite then the preceding example is essentially the only one. We say that the matrix M is *reducible* if there are disjoint sets

$I, J \subset \{1, \dots, k\}$ such that $m_{ij} = 0$ for all $i \in I$, $j \in J$. Of course, this implies that $\{X_i\}_{i \in I}$ and $\{X_j\}_{j \in J}$ are independent. We say that M is *irreducible* if it is not reducible. In this case, a decomposition into independent collections of variables is not possible and we show that the only equality cases of (1.1) are families of parallel half-spaces:

Theorem 1.5. *Suppose that M is strictly positive definite and irreducible. Then the configuration $A_1, \dots, A_k$ attains equality in* (1.1) *if and only if there exists a collection $(B_1, \dots, B_k)$ of parallel half-spaces such that $A_i = B_i$ up to sets of Lebesgue measure zero.*

By applying Theorem 1.5 to each irreducible component of M, one can also characterize the equality cases of (1.1) when M is strictly positive definite and reducible.

To see why we require the non-degeneracy of M in Theorem 1.5, take $n = 1$ and suppose that the X_i almost surely satisfy $X_1 = \sum_{i \geq 2} a_i X_i$ for some non-negative coefficients a_i. Then $\{X_i \leq b_i \text{ for all } i \geq 2\}$ implies $X_1 \leq \sum a_i b_i$. Therefore, the quantity

$$\Pr(X_1 \in A_1, X_i \leq b_i \text{ for all } i \geq 2)$$

does not depend on A_1 as long as A_1 contains $\{x \in \mathbb{R} : x \leq \sum a_i b_i\}$; in particular, one can maximize the above quantity without requiring A_1 to be a half-space. Now, the columns of M are positively independent (in the sense that no column can be written as a non-negative linear combination of the others) then an example like this does not exist. We ask, therefore, whether the Theorem 1.5's assumption that M be positive definite could be replaced by the assumption that M has positively independent columns.

1.2 From noise stability to exit times

Next, we will show how Theorem 1.4 may be used to recover Theorem 1.1. This reduction is quite similar to one by Burchard and Schmuckenschläger [4], who were studying exit times of Brownian motion on manifolds. (In that case, the study of exit times has a fairly long history; see [4] for references.) As we will see, though, our approach to Theorem 1.4 is quite different to that of Burchard and Schmuckenschläger, who studied two-point symmetrizations.

Let X_t be the Ornstein-Uhlenbeck process, and consider the finite dimensional marginal $(X_{t_1}, \dots, X_{t_k})$ for a sequence of times $t_1 < \cdots < t_k$. This is a mean-zero Gaussian vector with covariance $M \otimes I_n$, where $m_{ij} = e^{-|t_i - t_j|}$. Clearly, then, M satisfies the hypothesis of Theorem 1.4 and in particular, we have

$$\Pr(X_{t_i} \in A \text{ for all } i) \leq \Pr(X_{t_i} \in B \text{ for all } i) \tag{1.2}$$

when B is a half-space with $\gamma_n(A) = \gamma_n(B)$. This is essentially a discrete version of Theorem 1.1, since the event $\{X_{t_i} \in A \text{ for all } i\}$ is a discretization of $\{e_A \geq t_k\}$.

To complete the proof, we need to show that one can take limits. Setting $t_i = it/k$ in (1.2), we have

$$\begin{aligned} \Pr(e_A \geq \tau) &= \Pr(X_t \in A \text{ for all } 0 \leq t < \tau) \\ &\leq \Pr(X_{i\tau/k} \in A \text{ for all } i = 1, \dots, k) \\ &\leq \Pr(X_{i\tau/k} \in B \text{ for all } i = 1, \dots, k), \end{aligned}$$

where the last inequality follows from (1.2). Next, we send $k \to \infty$. Recall that X_t is uniformly continuous on $[0, \tau]$ with probability 1. In particular, for any $\epsilon, \delta > 0$, we may take $k = k(\epsilon, \delta)$ large enough so that with probability $1 - \delta$, $0 \leq s, t \leq \tau$ and $|s - t| \leq 1/k$ imply that $|X_s - X_t| \leq \epsilon$; for this k,

$$\begin{aligned} \Pr(X_{i\tau/k} \in B \text{ for all } i = 1, \dots, k) &\leq \Pr(X_t \in B_\epsilon \text{ for all } 0 \leq t < \tau) + \delta \\ &= \Pr(e_{B_\epsilon} \geq \tau) + \delta, \end{aligned}$$

where B_ϵ is the ϵ-enlargement of B: $B_\epsilon = \{x \in \mathbb{R}^n : d(x, B) \leq \epsilon\}$. Since $\delta > 0$ is arbitrary, we have shown that for any $\epsilon > 0$, $\Pr(e_A \geq \tau) \leq \Pr(e_{B_\epsilon} \geq \tau)$. It only remains to show, then, that $\Pr(e_{B_\epsilon} \geq \tau)$ converges to $\Pr(e_B \geq \tau)$ as $\epsilon \to 0$.

Consider instead the equivalent statement that $\Pr(e_{B_\epsilon} < \tau)$ converges to $\Pr(e_B < \tau)$. Since B is closed and X_t has continuous paths, $e_B < \tau$ implies that there is some $\epsilon > 0$ with $e_{B_\epsilon} < \tau$. That is, the function $1_{\{e_{B_\epsilon} < \tau\}}$ converges pointwise (and upwards) to $1_{\{e_B < \tau\}}$ as $\epsilon \to 0$. By the monotone convergence theorem, it follows that $\Pr(e_{B_\epsilon} < \tau)$ converges to $\Pr(e_B < \tau)$ as $\epsilon \to 0$. Hence

$$\Pr(e_A \geq \tau) \leq \lim_{\epsilon \to 0} \Pr(e_{B_\epsilon} \geq \tau) = \Pr(e_B \geq \tau)$$

and so we have recovered Theorem 1.1.

We have mentioned already that it is also possible to recover Theorem 1.1 from the result in [8]. Indeed, the matrix M with entries $m_{ij} = e^{-|t_i - t_j|}$ does satisfy the hypothesis in [8] (namely that the off-diagonal entries of its inverse are non-positive), although this is certainly less obvious then the fact that M satisfies the conditions of Theorem 1.4.

Let us also indicate how Theorem 1.2 is recovered. We want to show that

$$\mathbb{E} \int_0^{t \wedge e_{A_1}} 1_{A_2}(X_s)\, ds \tag{1.3}$$

is only increased when A is replaced by B (recall that B_1 and B_2 are parallel half-spaces satisfying $\gamma_n(B_i) = \gamma_n(A_i)$. We may suppose that $A_2 \subset A_1$, since if not then (1.3) may be trivially made larger by moving some of A_2's mass inside A_1; if this is impossible because $\gamma_n(A_2) > \gamma_n(A_1)$ then (1.3) is trivially bounded by $t \wedge e_{A_1}$, which, by Theorem 1.1, is stochastically dominated by $t \wedge e_{B_1}$, and this in turn is equal to the right hand side of (1.3) with B replacing A.

Now that we have reduced to the case $A_2 \subset A_1$, we may discretize (1.3) as

$$\frac{1}{k} \sum_{i=1}^{k} \Pr(X_{t_j} \in A_1 \text{ for all } j < i \text{ and } X_{t_i} \in A_2).$$

By (1.2), this is only increased when A is replaced by B. To recover Theorem 1.2 from here, it suffices to take a limit in much the same manner as before; we omit the details.

2 The Ornstein-Uhlenbeck semigroup

We will prove Theorem 1.4 by differentiating a particular functional under the Ornstein-Uhlenbeck semigroup. This proof method has a long history, beginning with Varopoulos' work [16] connecting the heat semigroup with Sobolev inequalties. More recently, and more apropos of this work, Bakry and Ledoux [1] proved the Gaussian isoperimetric inequality by differentiating Bobkov's functional [2] under the Ornstein-Uhlenbeck semigroup. We will follow quite a similar approach here, using a generalization of a functional that was introduced by Mossel and the author [13] to prove Theorem 1.4 in the case $k = 2$.

We define the Ornstein-Uhlenbeck semigroup $\{P_t : t \geq 0\}$, which acts on bounded, measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ by

$$(P_t f)(x) = \int_{\mathbb{R}^d} f(e^{-t}x + \sqrt{1 - e^{-2t}}y)\, d\gamma_d(y),$$

where γ_d is the standard Gaussian measure on $\mathbb{R}^d$. Note that the dependence of this definition on d is implicit; we will sometimes take $d = n$ and sometimes take $d = kn$, but

in either case the dimension should be clear from the context. Equivalently, $(P_t f)(X_0) = \mathbb{E}(f(X_t) \mid X_0)$, where $\{X_t : t \geq 0\}$ is the Ornstein-Uhlenbeck process from the previous section. From either definition, one can easily see that P_0 is the identity operator, while $P_t f \to \mathbb{E}f$ as $t \to \infty$.

One remarkable property of the Ornstein-Uhlenbeck semigroup is that it has very nice formulas for its commutation with smooth functions. In particular, for any bounded, measurable $F = (f_1, \ldots, f_k) : \mathbb{R}^d \to \mathbb{R}^k$, for any smooth $\Psi : \mathbb{R}^k \to \mathbb{R}$, and for any $0 < s < t$, there is the formula (see, eg., [11])

$$\frac{d}{ds} P_s \Psi(P_{t-s}F) = P_s \sum_{i,j=1}^{k} \partial_{ij}^2 \Psi(P_{t-s}F) \langle \nabla P_{t-s} f_i, \nabla P_{t-s} f_j \rangle. \tag{2.1}$$

We begin with an observation that comes, essentially, from applying (2.1) to the function that is Ψ composed with a linear operator. In the following, $\odot$ denotes the Hadamard (elementwise) product between two matrices and H_Ψ denotes the Hessian matrix of Ψ.

Proposition 2.1. *Let M be a $k \times k$ positive semi-definite matrix with $m_{ii} = 1$ and let $X = (X_1, \ldots, X_k)$ be a kn-dimensional Gaussian vector with mean zero and covariance $M \otimes I_n$. If $\Psi : [0,1]^k \to \mathbb{R}$ satisfies $M \odot H_\Psi \leq 0$ then for all bounded, measurable $F(x_1, \ldots, x_k) = (f_1(x_1), \ldots, f_k(x_k)) : \mathbb{R}^{kn} \to [0,1]^k$,*

$$\mathbb{E}\Psi(F(X)) \leq \Psi(\mathbb{E}F).$$

We remark that the assumption $m_{ii} = 1$ in Proposition 2.1 is not necessary, but it makes our notation simpler. Anyway, reducing to the case $m_{ii} = 1$ is simply a matter of rescaling X_i.

In a private communication, R. O'Donnell pointed out that the converse of Proposition 2.1 is also true: if $\mathbb{E}\Psi(F(X)) \leq \Psi(\mathbb{E}F)$ for every F then $M \odot H_\Psi \leq 0$. This may be seen by setting $f_i(x) = \max\{0, \min\{1, a_i \cdot x + \epsilon b_i\}\}$. By considering the second-order Taylor expansion of Ψ and taking $\epsilon \to 0$, one sees that $b^T(M \odot H_\Psi(a))b \leq 0$.

Before proving Proposition 2.1, we introduce some notation that will be useful in what follows: for any $f : \mathbb{R}^n \to \mathbb{R}$ and any $n \times m$ matrix M, denote the function $f \circ M : \mathbb{R}^m \to \mathbb{R}$ by f^M.

Proof. Let Q be the positive semi-definite square root of $M \otimes I_n$, and for $i = 1, \ldots, k$, let Q_i be the $n \times kn$ matrix consisting of rows $(i-1)n + 1$ through in of Q. Let Z be a standard Gaussian vector in $\mathbb{R}^{kn}$, and note that $QZ = (Q_1 Z, \ldots, Q_k Z)$ is a kn-dimensional Gaussian vector with mean 0 and covariance $M \otimes I_n$ (i.e., QZ has the same distribution as X). We consider the quantity

$$G(s,t,z) = \left(P_s \Psi(P_{t-s} f_1^{Q_1}, \ldots, P_{t-s} f_k^{Q_k}) \right)(z)$$

for $s, t \in [0, \infty)$ and $z \in \mathbb{R}^{kn}$. First, let us check how P_t commutes with linear transformations. Since Q is the square root of $M \otimes I_n$, we have $Q_i Q_i^T = I_n$ and so

$$\begin{aligned}
(P_t f_i^{Q_i})(x) &= \int_{\mathbb{R}^{kn}} f_i(e^{-t} Q_i x + \sqrt{1 - e^{-2t}} Q_i y)\, d\gamma_{kn}(y) \\
&= \int_{\mathbb{R}^n} f_i(e^{-t} Q_i x + \sqrt{(1 - e^{-2t})} y)\, d\gamma_n(y) \\
&= (P_t f_i)^{Q_i}(x).
\end{aligned} \tag{2.2}$$

Now, the gradient commutes with linear transformations as $\nabla f^A = A^T(\nabla f)^A$. Combining this with (2.2),

$$\nabla P_{t-s} f_i^{Q_i} = \nabla (P_{t-s} f_i)^{Q_i} = Q_i^T (\nabla P_{t-s} f_i)^{Q_i}.$$

In particular,

$$\begin{aligned}\langle \nabla P_{t-s} f_i^{Q_i}, \nabla P_{t-s} f_j^{Q_j} \rangle &= \langle Q_i^T (\nabla P_{t-s} f_i)^{Q_i}, Q_j^T (\nabla P_{t-s} f_j)^{Q_j} \rangle \\ &= m_{ij} \langle (\nabla P_{t-s} f_i)^{Q_i}, (\nabla P_{t-s} f_j)^{Q_j} \rangle. \end{aligned} \tag{2.3}$$

For brevity, let $v_i = (\nabla P_{t-s} f_i)^{Q_i}$. Then, by (2.1) and (2.3),

$$\begin{aligned}\frac{\partial G(s,t,\cdot)}{\partial s} &= P_s \sum_{i,j=1}^k \langle \nabla P_{t-s} f_i^{Q_i}, \nabla P_{t-s} f_j^{Q_j} \rangle \partial_{ij}^2 \Psi(P_{t-s} F^Q) \\ &= P_s \sum_{i,j=1}^k m_{ij} \langle v_i, v_j \rangle \partial_{ij}^2 \Psi(P_{t-s} F^Q). \end{aligned} \tag{2.4}$$

Note that if $v^T = (v_1^T \dots v_k^T)$ then the last line may be rewritten as

$$P_s \left(v^T ((M \odot H_\Psi) \otimes I_n) v \right). \tag{2.5}$$

In particular, if $M \odot H_\Psi \le 0$ then $\frac{\partial G(s,t,z)}{\partial s} \le 0$ for every s, t and z. Hence,

$$\lim_{t\to\infty} G(t,t,Z) \le \lim_{t\to\infty} G(0,t,Z).$$

But since $(Q_1 Z, \dots, Q_k Z)$ has the same distribution as $(X_1, \dots, X_k)$, $\mathbb{E}G(t,t,Z)$ converges to $\mathbb{E}\Psi(F(X))$ as $t \to \infty$, while $\mathbb{E}G(0,t,Z)$ converges to $\Psi(\mathbb{E}F(X))$. □

With hardly any extra effort, the proof of Proposition 2.1 also allows us to characterize its equality cases. Indeed, if $\mathbb{E}\Psi(F) = \Psi(\mathbb{E}F)$ then we must have $\frac{\partial F(s,t,z)}{\partial s} = 0$ for every s, t, and z. Going back to (2.4) and (2.5), we see that $P_s \left(v^T ((M \odot H_\Psi) \otimes I_n) v \right)$ must be identically zero, and hence $((M \odot H_\Psi) \otimes I_n) v = 0$. In other words, we have the following corollary:

Corollary 2.2. *Under the hypothesis of Proposition 2.1, if $\mathbb{E}\Psi(F) = \Psi(\mathbb{E}F)$ then for every $t > 0$,*

$$\left((M \odot H_\Psi(P_t F^Q)) \otimes I_n \right) \begin{pmatrix} (\nabla P_t f_1)^{Q_1} \\ \vdots \\ (\nabla P_t f_k)^{Q_k} \end{pmatrix} = 0 \text{ a.s.}$$

3 Proof of Theorem 1.4

Before proving Theorem 1.4, note that by translating X, A_i, and B_i, it suffices to consider the case in which X has mean zero. Moreover, by scaling each X_i, A_i, and B_i, we may assume that $m_{ii} = 1$ for each i (here and in the previous sentence we are using the fact that a collection of parallel half-spaces remains one under translation and scaling).

Let $\Phi(y) = (2\pi)^{-1/2} \int_{-\infty}^y e^{-z^2/2}\, dz$ denote the Gaussian cumulative distribution function, and let $\Phi^{-1} : (0,1) \to \mathbb{R}$ be its inverse, which we will extend to $[0,1]$ by setting $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = \infty$. Consider the function

$$J(x_1, \dots, x_k; M) = \Pr\left(X_{i,1} \le \Phi^{-1}(x_i) \text{ for all } i \right), \tag{3.1}$$

where $X_{i,1}$ denotes the first coordinate of $X_i \in \mathbb{R}^n$. Observe that $\Pr(X_{i,1} \leq \Phi^{-1}(x_i)) = x_i$; in particular, the collection $(B_1, \ldots, B_k)$ with $B_i = \{y \in \mathbb{R}^n : y_1 \leq \Phi^{-1}(x_i)\}$ is a set of parallel half-spaces with $\Pr(X_i \in B_i) = x_i$. Since every such set of parallel half-spaces may be obtained by applying a fixed rotation to each B_i, Theorem 1.4 is equivalent to the statement

$$\Pr(X_i \in A_i \text{ for all } i) \leq J(\gamma_n(A_1), \ldots, \gamma_n(A_k); M).$$

Next, note that if $x_1, \ldots, x_k \in \{0, 1\}$ then $J(x_1, \ldots, x_k)$ is 1 if all the x_i are 1, and 0 otherwise. In particular,

$$\Pr(X_i \in A_i \text{ for all } i) = \mathbb{E}J(1_{A_1}(X_1), \ldots, 1_{A_k}(X_k); M)$$

and hence (1.1) is equivalent to the statement

$$\mathbb{E}J(f_1(X_1), \ldots, f_k(X_k); M) \leq J(\mathbb{E}f_1, \ldots, \mathbb{E}f_k; M) \tag{3.2}$$

in the special case $f_i = 1_{A_i}$. In fact, we will prove (3.2) for general measurable functions $f_i : \mathbb{R}^n \to [0, 1]$. We remark, however, that (3.2) is no stronger than (1.1), since it may be deduced from the $(n+1)$-dimensional case of (1.1) (see [13]).

Unsurprisingly, the proof of (3.2) goes through Proposition 2.1. The main task left, therefore, is to compute the Hessian of J and check that it satisfies the hypothesis of Proposition 2.1.

Proposition 3.1. *If $m_{ii} = 1$ and $m_{ij} \geq 0$ then J satisfies $M \odot H_J \leq 0$.*

Next, we will compute H_J and show that $m_{ij} \geq 0$ implies that $M \odot H_J \leq 0$. For a vector v, let $v_{\hat{i}}$ denote v without the ith entry, and for a square matrix M, let $M_{\hat{i}}$ denote M without the ith row and column. For a square matrix M, let $\overline{M_i}$ denote the Schur complement of the element (i, i) in M. In other words, $\overline{M_i}$ is defined by

$$\overline{M_i} = M_{\hat{i}} - M_{i\hat{i}}^T M_{i\hat{i}}/m_{ii} = M_{\hat{i}} - M_{i\hat{i}}^T M_{i\hat{i}},$$

where M_i denotes the ith row of M, so $M_{i\hat{i}}$ is the ith row of M with its ith element removed. A well-known formula for conditional distributions of Gaussian vectors (see, e.g., [6]) states if X has mean zero and covariance M satisfying $m_{ii} = 1$, then conditioned on $X_i = x_i$, $X_{\hat{i}}$ has mean $x_i M_{i\hat{i}}$ and covariance $\overline{M_i}$.

To compute the first derivatives of J, let

$$K(y_1, \ldots, y_k; M) = \Pr\left(X_i \leq y_i \text{ for all } i\right),$$

and note that $J(x_1, \ldots, x_k) = K(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_k))$. Now, for any i we may write K as

$$K(y; M) = \int_{-\infty}^{y_i} \phi(z) \Pr(X_j \leq y_j \text{ for all } j \neq i \mid X_i = z)\, dz,$$

where $\phi(z) = (2\pi)^{-1/2} e^{-z^2/2}$ is the standard Gaussian density on $\mathbb{R}$. Hence,

$$\partial_i K(y; M) = \phi(y_i) \Pr(X_j \leq y_j \text{ for all } j \neq i \mid X_i = y_i).$$

Now, given that $X_i = y_i$, $X_{\hat{i}}$ has mean $M_{i\hat{i}} y_i$ and covariance $\overline{M_i}$; hence,

$$\Pr(X_j \leq y_j \text{ for all } j \neq i \mid X_i = y_i) = K(y_{\hat{i}} - M_{i\hat{i}} y_i; \overline{M_i}),$$

and so we have the formula

$$\partial_i K(y; M) = \phi(y_i) K(y_{\hat{i}} - M_{i\hat{i}} y_i; \overline{M_i})$$

(bear in mind that this formula is only valid under the assumption $m_{ii} = 1$; if not then m_{ii} makes an appearance in the formula also). Applying the chain rule and the identity $\frac{d}{dx}\Phi^{-1}(x) = 1/\phi(\Phi^{-1}(x))$, we have

$$\partial_i J(x; M) = K(\Phi^{-1}(x_{\hat{i}}) - M_{i\hat{i}}\Phi^{-1}(x_i); \overline{M_i}) \tag{3.3}$$

(where by $\Phi^{-1}(x_{\hat{i}})$, we mean the vector obtained by applying Φ^{-1} to $x_{\hat{i}}$ element-wise). Now let $I(x) = \phi(\Phi^{-1}(x))$ and define, for $j \neq i$,

$$J_{ij}(x; M) = I(x_i)\partial_j K(\Phi^{-1}(x_{\hat{i}}) - M_{i\hat{i}}\Phi^{-1}(x_i); \overline{M_i});$$

by the chain rule applied to (3.3), we have

$$\begin{aligned}\partial_i\partial_j J(x; M) &= \frac{1}{I(x_j)}\partial_j K(\Phi^{-1}(x_{\hat{i}}) - M_{i\hat{i}}\Phi^{-1}(x_i); \overline{M_i}) \\ &= \frac{1}{I(x_i)I(x_j)} J_{ij}(x; M).\end{aligned} \tag{3.4}$$

It is worth mentioning that this last equation shows that in fact $J_{ij} = J_{ji}$. This is not obvious from the definition of J_{ij}, although it may also be checked by the tedious process of calculating the derivative in that definition.

To compute the repeated second derivatives of J, we use (3.3) and the chain rule to write

$$\begin{aligned}\partial_i^2 J(x; M) &= -\sum_{j\neq i} \frac{m_{ij}}{I(x_i)}\partial_j K(\Phi^{-1}(x_{\hat{i}}) - M_{i\hat{i}}\Phi^{-1}(x_i); \overline{M_i}) \\ &= -\frac{1}{I^2(x_i)} \sum_{j\neq i} m_{ij} J_{ij}(x; M).\end{aligned} \tag{3.5}$$

Now let $\mathcal{I}(x)$ be the $k \times k$ diagonal matrix with $1/I(x_i)$ as the ith diagonal entry. Note that by (3.4), the ij entry of $M \odot H_J$ is given by $\frac{m_{ij}}{I(x_i)I(x_j)} J_{ij}$, while the ii entry of $M \odot H_J$ is just given by (3.5) (since $m_{ii} = 1$). Hence we may write

$$M \odot H_J(x; M) = \mathcal{I}(x)A(x)\mathcal{I}(x), \tag{3.6}$$

where $a_{ij} = m_{ij}J_{ij}$ and $a_{ii} = -\sum_{j\neq i} a_{ij}$.

Lemma 3.2. *If A is a symmetric matrix such that $a_{ij} \geq 0$ for $i \neq j$ and $a_{ii} = -\sum_{j\neq i} a_{ij}$ then $A \leq 0$.*

Proof. In fact, the proof follows immediately from some well-known facts in linear algebra, such as the fact that $-A$ is diagonally dominant. However, we may also give a simple proof by noting that the quadratic form of A is nothing but

$$v^T A v = -\sum_{i<j} a_{ij}(v_i - v_j)^2 \leq 0. \qquad \square$$

To apply Lemma 3.2 with the matrix A given in (3.6), note that J_{ij} is non-negative because by definition it is a positive number times a derivative of $K(\cdot, \overline{M_i})$, which is non-decreasing in each coordinate. Then Lemma 3.2 implies Proposition 3.1, which completes the proof of (1.1).

3.1 The equality cases

In order to prove Theorem 1.5, we first remark on some points in the previous proof. First of all, if M is strictly positive definite then $J_{ij}(x; M)$ is strictly positive everywhere.

Indeed, if M is positive definite then $\overline{M_i}$ is also, and the definition of K then ensures that $K(y;\overline{M_i})$ is strictly increasing in each coordinate of y; hence $J_{ij} > 0$. In particular, if M is positive definite and irreducible then the matrix $A(x)$ defined in (3.6) is irreducible for all x.

Next, if A is irreducible in Lemma 3.2 then the kernel of A is the span of the all-ones vector. This follows from the formula $v^T A v = -\sum_{i<j} a_{ij}(v_i - v_j)^2$ and the fact that if A is irreducible and $i \neq j$ then there is some sequence $i = i_0, \ldots, i_\ell = j$ such that $a_{i_{k-1},i_k} > 0$ for all $1 \leq k \leq \ell$. Combining this with the previous paragraph, we see that the kernel of $A(x)$ in (3.6) is the span of the all-ones vector.

Now we may prove Theorem 1.5 using Corollary 2.2: if $\mathbb{E}J(F) = J(\mathbb{E}F)$ then for every $t > 0$,

$$\begin{aligned} 0 &= \left((M \odot H_\Psi(P_t F^Q)) \otimes I_n\right) \begin{pmatrix} (\nabla P_t f_1)^{Q_1} \\ \vdots \\ (\nabla P_t f_k)^{Q_k} \end{pmatrix} \\ &= \left((\mathcal{I}(P_t F^Q) A(P_t F^Q) \mathcal{I}(P_t F^Q)) \otimes I_n\right) \begin{pmatrix} (\nabla P_t f_1)^{Q_1} \\ \vdots \\ (\nabla P_t f_k)^{Q_k} \end{pmatrix} \\ &= (\mathcal{I}(P_t F^Q) \otimes I_n)(A(P_t F^Q) \otimes I_n)(\mathcal{I}(P_t F^Q)) \otimes I_n) \begin{pmatrix} (\nabla P_t f_1)^{Q_1} \\ \vdots \\ (\nabla P_t f_k)^{Q_k} \end{pmatrix}, \end{aligned} \tag{3.7}$$

where the second equality follows from (3.6). Since $\mathcal{I}$ is always non-singular, we may drop the first instance of it from (3.7). Defining $w_i = \Phi^{-1} \circ P_t f_i$, we have $\nabla w_i = \nabla P_t f_i / I(P_t f_i)$. By multiplying out the last two terms of (3.7), we have

$$0 = (A(P_t F^Q) \otimes I_n) \begin{pmatrix} \frac{(\nabla P_t f_1)^{Q_1}}{I(P_t f_1)^{Q_1}} \\ \vdots \\ \frac{(\nabla P_t f_k)^{Q_k}}{I(P_t f_k)^{Q_k}} \end{pmatrix} = (A(P_t F^Q) \otimes I_n) \begin{pmatrix} (\nabla w_1)^{Q_1} \\ \vdots \\ (\nabla w_k)^{Q_k} \end{pmatrix}.$$

Now, recall that under the assumptions of Theorem 1.5, the kernel of A is the all-ones vector. It follows then that if

$$(A \otimes I_n) \begin{pmatrix} (\nabla w_1)^{Q_1} \\ \vdots \\ (\nabla w_k)^{Q_k} \end{pmatrix} = 0$$

then the $(\nabla w_i)^{Q_i}$ are all equal. Since this holds pointwise, and since Q (recall that $Q^T = (Q_1^T, \ldots, Q_k^T)$) has a trivial kernel, we see that ∇w_i must all be almost surely equal to the same constant, a say. Since each w_i is a smooth function, we have $w_i(x) = a \cdot x + b_i$ for some b_i. Recalling the definition of w_i, we have $(P_t f_i)(x) = \Phi(a \cdot x + b_i)$. Carlen and Kerce [5] observed (and this observation was subsequently used in [13] and [13]) that under this condition, and if $f_i = 1_{A_i}$, then A_i is a half-space (up to a set of measure zero) and a is normal to it. Since we have the same a for every A_i, it follows that $A_1, \ldots, A_k$ is a family of parallel half-spaces, which completes the proof of Theorem 1.4.

In order to be more self-contained, we sketch a proof (from [13]) of why $(P_t 1_A)(x) = \Phi(a \cdot x + b_i)$ implies that A is a half-space. First, one checks that if A is a half-space and ν its outward unit normal, then

$$(P_t 1_A)(x) = \Phi(k_t \nu \cdot x + b)$$

for some $b \in \mathbb{R}$, where $k_t = (e^{2t} - 1)^{-1/2}$. Moreover, as A ranges over all half-spaces with normal ν then b ranges over $\mathbb{R}$. After checking that $P_t : L^2(\gamma_n) \to L^2(\gamma_n)$ is one-to-one (for example, because it acts diagonally on the Hermite basis), this implies that if $(P_t 1_A)(x) = \Phi(a \cdot x + b)$ with $|a| = k_t$, then A is (up to a null set) a half-space with normal a. It remains to see what happens when $|a| \neq k_t$. First of all, Bakry and Ledoux showed that $|\nabla(\Phi^{-1} \circ P_t f)| \leq k_t$ for any $f : \mathbb{R}^n \to [0,1]$; hence $|a| \leq k_t$. But if $|a| < k_t$ then there is some $s > 0$ with $|a| = k_{t+s}$. It follows from the previous argument, then, that there is a half-space B with $P_{t+s} 1_B = \Phi(a \cdot x + b) = P_t 1_A$. We then have $P_s 1_B = 1_A$, which is a contradiction since $P_s 1_B$ is always a smooth function.

References

[1] D. Bakry and M. Ledoux. Lévy–Gromov's isoperimetric inequality for an infinite dimensional diffusion generator. *Invent. math.*, 123(2):259–281, 1996. MR-1374200

[2] S. G. Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. Probab.*, 25(1):206–214, 1997. MR-1428506

[3] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985. MR-0795785

[4] A. Burchard and M. Schmuckenschläger. Comparison theorems for exit times. *Geom. Funct. Anal.*, 11(4):651–692, 2001. MR-1866798

[5] E. A. Carlen and C. Kerce. On the cases of equality in Bobkov's inequality and Gaussian rearrangement. *Calc. Var. Partial Dif.*, 13(1):1–18, 2001. MR-1854254

[6] M. L. Eaton. *Multivariate statistics: a vector space approach*. Wiley, 1983. MR-0716321

[7] A. Ehrhard. Inégalités isopérimétriques et intégrales de Dirichlet gaussiennes. *Ann. Sci. École Norm. Sup. (4)*, 17(2):317–332, 1984. MR-0760680

[8] M. Isaksson and E. Mossel. Maximally stable Gaussian partitions with discrete applications. *Isr. J. Math.*, 189(1):347–396, 2012. MR-2931402

[9] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? In *Proc. 45th Annu. IEEE Symp. Foundations of Computer Science*, pages 146–154. IEEE, 2004.

[10] G. Kindler and R. O'Donnell. Gaussian noise sensitivity and Fourier tails. In *IEEE Conf. Computational Complexity*, pages 137–147, 2012.

[11] M. Ledoux. The geometry of Markov diffusion generators. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2):305–366, 2000. Probability theory. MR-1813804

[12] E. Mossel. Gaussian bounds for noise correlation of functions. *Geom. Funct. Anal.*, 19:1713–1756, 2010. MR-2594620

[13] E. Mossel and J. Neeman. Robust dimension free isoperimetry in Gaussian space. 2012.

[14] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. Math.*, 171(1):295–341, 2010. MR-2630040

[15] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proc. 40th Annu. Symp. Theory of Computing*, pages 245–254, 2008. MR-2582901

[16] N.T. Varopoulos. Hardy-Littlewood theory for semigroups. *J. Funct. Anal*, 63(2):240–260, 1985. MR-0803094