ON GAUSSIAN AND BERNOULLI COVARIANCE REPRESENTATIONS

S. G. Bobkov *   F. Götze   C. Houdré

September 29, 1999

Abstract

We show that a covariance representation in Gauss space implies large deviation inequalities for smooth functions of Gauss random vectors. This covariance representation characterizes Gaussian measures. of the Gaussian. This representation has an analogue for Bernoulli measures and implies sharp large deviation inequalities for Bernoulli sums as well.

Key words: covariance identities, large deviations, Gaussian measures, Bernoulli sums

1 Introduction

Let γ_n be the canonical Gaussian measure on $\mathbf{R}^n$ with density $d\gamma_n(x)/dx = (2\pi)^{-n/2}\exp(-|x|^2/2)$, and let X and Y be two independent random vectors in $\mathbf{R}^n$ with distribution γ_n. Then, for any smooth functions f and g on $\mathbf{R}^n$,

$$\mathrm{cov}(f(X), g(X)) = \int_0^1 \mathbf{E}\left\langle \nabla f\,(X), \nabla g\,(\alpha\, X + \sqrt{1-\alpha^2}\, Y)\right\rangle d\alpha, \qquad (1.1)$$

where $\mathrm{cov}(f,g) = \mathbf{E}fg - \mathbf{E}f\mathbf{E}g$, where ∇ stands for the usual gradient, and where $\langle\cdot\,,\cdot\rangle$ is the euclidian inner product in $\mathbf{R}^n$. In an infinite dimensional setting (for functionals of the Wiener process) the above identity appeared in [H-PA], and was then explicitly connected to the Ornstein-Uhlenbeck semigroup by M. Ledoux [L3], a related approach is the work of Herbst and Pitt [HP]. Indeed, if P_t, $t \geq 0$, is the Ornstein–Uhlenbeck operator associated to γ_n and acting in the space $L^1(\gamma_n)$ via

$$(P_t g)(x) = \int_{\mathbf{R}^n} g(e^{-t}x + \sqrt{1-e^{-2t}}\, y)\, d\gamma_n(y), \quad x \in \mathbf{R}^n,$$

*Research supported in part by the Alexander von Humboldt Foundation

then (1.1) becomes

$$\operatorname{cov}(f,g) = \int_0^{+\infty} \mathbf{E}\,\langle \nabla f, \nabla P_t g\rangle\, dt. \tag{1.2}$$

The covariance representation (1.1) which somehow hides the presence of the Ornstein-Uhlenbeck operator was further studied in [H-P-S]. There, it is treated as a particular case of a more general representation for the expectation of functions of infinitely divisible random vectors (see [H-P-S] and the references therein). It turns out that (1.1) is a powerful tool in the study of a number of problems on Gaussian measures, such as correlation inequalities and comparison results [H]. In relation to (1.2) properties of the Ornstein–Uhlenbeck and other Markovian semigroups have been deeply investigated, in the isoperimetric context by D. Bakry and M. Ledoux (cf. [B], [B-L], [L1–L6]). In particular ([L5]), the semigroup technique is a tool to recover the Gaussian isoperimetric inequality due to V. N. Sudakov, B. S. Tsirel'son [S-T] and C. Borell [Bor]. This isoperimetric inequality implies the following inequality for the deviation of a Lipschitz function f on $\mathbf{R}^n$ from its medians $m(f)$ with respect to γ_n:

$$\gamma_n\{|f - m(f)| \geq h\} \leq 2(1 - \Phi(h)), \quad h > 0. \tag{1.3}$$

Here $\Phi(h) = \gamma_1((-\infty, h])$ is the distribution function of γ_1. In this note, we present (among other results) another related application based on (1.1). Namely, we prove that, for every Lipschitz function f on $\mathbf{R}^n$, i.e., such that $\|f\|_{\mathrm{Lip}} \leq 1$,

$$\gamma_n\{|f - \mathbf{E}f| \geq h\} \leq \mathbf{E}|f - \mathbf{E}f|\, \frac{e^{-h^2/2}}{h}, \quad h > 0. \tag{1.4}$$

The proof of this estimate is given in section 2 where we also discuss how it improves some of the known deviation inequalities. However, we would like to stress here that it is unlikely that it is possible to derive (1.4) from (1.3) or from the Gaussian isoperimetric inequality in its full strength. The latter allows one to reduce (1.4) to dimension one, but even in this case we are faced with analytic difficulties.

Another easy application of (1.1) worth mentioning is the following exponential moment inequality: For any smooth function f defined on $\mathbf{R}^n$ and such that $\mathbf{E}f = 0$,

$$\mathbf{E}e^f \leq \mathbf{E}e^{|\nabla f|^2}. \tag{1.5}$$

With a worse constant $(\pi^2/8)$ (1.5) is due to Pisier [P], while as stated above it was proved in [BG]. Actually, the result obtained in [BG] shows that under a log–Sobolev inequality, a result similar to (1.5) always holds.

If μ_α denotes the Gaussian measure on $\mathbf{R}^n \times \mathbf{R}^n$ which is the distribution of the random vector $\left(X, \alpha X + \sqrt{1-\alpha^2}\, Y\right)$, and if π_n denotes the probability measure $\int_0^1 \mu_\alpha\, d\alpha$, then (1.1) becomes

$$\operatorname{cov}(f(X), g(X)) = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \langle \nabla f\,(x), \nabla g\,(y)\rangle\; d\pi_n(x,y). \tag{1.6}$$

It is the form (1.6) that we apply to obtain (1.4), only using the fact that the marginals of π_n are γ_n. One may also wonder whether or not, it is possible to find a representation similar to (1.6) for random vectors X in $\mathbf{R}^n$ with other, non-Gaussian distributions. As it turns out, for $n \geq 2$ this is not the case: the existence of a measure π_n in (1.6) implies that X is Gaussian. We prove this characterization at the end of section 2.

There exists however one remarkable exception: if we let Δ denote a discrete gradient for functions on the discrete cube $\{-1,1\}^n$, it is possible to find a discrete analogue of (1.6) for the canonical Bernoulli measure μ_n on $\{-1,1\}^n$ (which assigns the mass 2^{-n} to each point). Our main result asserts that, with respect to μ_n, and for all f and g on $\{-1,1\}^n$,

$$\operatorname{cov}(f,g) = \int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \langle \Delta f(x), \Delta g(y) \rangle \, d\nu_n(x,y), \tag{1.7}$$

where the mixing measure ν_n on $\{-1,1\}^n \times \{-1,1\}^n$ has density

$$\frac{d\nu_n(x,y)}{d\mu_n(x) d\mu_n(y)} = \int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i) \, d\alpha.$$

The above representation has also a semigroup version. Indeed, introducing the operator

$$(Q_t g)(x) = \int_{\{-1,1\}^n} \prod_{i=1}^n (1 + e^{-t} x_i y_i) \, g(y) \, d\mu_n(y), \quad x \in \{-1,1\}^n, \ t \geq 0,$$

where g is an arbitrary function on $\{-1,1\}^n$, then (1.7) becomes

$$\operatorname{cov}(f,g) = \int_0^\infty \mathbf{E} \langle \Delta f, \Delta Q_t g \rangle \, dt. \tag{1.8}$$

The semigroup Q_t enjoy properties similar to the Ornstein–Uhlenbeck semigroup P_t, and we refer to Bakry [B] for the precise relationship between P_t and Q_t.

In contrast to the Gaussian case, the range of applications of the representation (1.7) seems however to be much more restricted, due to the fact that the discrete gradient is not local. An easy direct application of (1.7)–(1.8) to large deviations of functions f on $\{0,1\}^n$ when we can freely work with the discrete gradient and discrete perimeter involves only classical functions $f(x) = \dfrac{x_1 + \ldots + x_n}{\sqrt{n}}$. In this case, we will show that for a universal constant K,

$$\mu_n \Big\{ f \geq h \Big\} \leq K \, \frac{e^{-h^2/2}}{h}, \quad h > 0. \tag{1.9}$$

As in (1.4), the main point in (1.9) is the factor $1/h$ on the right hand side. Without this factor, the inequality is trivial and holds for arbitrary linear functions $f(x) = a_1 x_1 + \ldots + a_n x_n$ with $a_1^2 + \ldots + a_n^2 = 1$ (the so-called subgaussian inequality), and even more generally, for all f on $\{-1,1\}^n$ with $|\Delta f| \leq 1$, $\mathbf{E} f = 0$ (as a consequence of the Gross' logarithmic Sobolev inequality on the discrete cube). However, we do not known if the "true" Gaussian tails (with the factor $1/h$) can be reached for such a general family of functions. In the class of linear functions, the problem of getting (1.9) known in literarure as Eaton's problem was affirmatively solved by I. Pinelis [Pin1] by using comparison of suitable moments of f with moments of f as a random variable on Gauss space. We finish the notes by giving a simple inductive proof of this result.

2 Deviation inequalities in Gauss space

Since on $\mathbf{R}^n \times \mathbf{R}^n$ the probability measure π_n in (1.5) has marginals γ_n, as an immediate consequence we obtain

Proposition 2.1. For every locally Lipschitz functions f and g on $\mathbf{R}^n$ such that $\|f\|_{\mathrm{Lip}} \leq 1$,

$$\operatorname{cov}(f,g) \leq \mathbf{E}|\nabla g|, \tag{2.1}$$

where the covariance and the expectation taken with respect to γ_n.

To better understand how this inequality is sharp, let us observe that for the linear functions $f(x) = \langle a, x\rangle$ with $|a| = 1$ and for the indicator functions g of halfspaces H of the form $H = \{x \in \mathbf{R}^n : \langle a, x\rangle \geq c\}$ there is equality in (2.1).

Let us apply (2.1) to $g = e^{tf}$, $t \geq 0$, assuming for a while that $\mathbf{E} f = 0$:

$$\mathbf{E} f\, e^{tf} = \operatorname{cov}(f, e^{tf}) \leq t\, \mathbf{E}|\nabla f| e^{tf} \leq t\, \mathbf{E} e^{tf}.$$

For the function $u(t) = \log \mathbf{E} e^{tf}$, we have $\mathbf{E} f\, e^{tf} = u'(t) e^{u(t)}$, hence, $u'(t) \leq t$. Since $u(0) = 0$, we conclude that, $u(t) \leq t^2/2$, that is, $\mathbf{E} e^{tf} \leq e^{t^2/2}$. Thus, for all f on $\mathbf{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$,

$$\mathbf{E} e^{tf} \leq e^{t\,\mathbf{E} f + \frac{t^2}{2}}. \tag{2.2}$$

By Chebyshev's inequality, it follows that for all $h \geq 0$,

$$\gamma_n\{|f - \mathbf{E} f| \geq h\} \leq 2\, e^{-h^2/2}. \tag{2.3}$$

Using stochastic integrals, the inequality (2.2) was proved by B. Maurey ([P, pp.180–181]). This type of ideas goes back to I. A. Ibragimov, V. N. Sudakov and B. S. Tsirel'son [C-I-S]. They expressed the functional f in the form $f = \mathbf{E} f + W(\tau)$ where W is the standard Wiener process and τ is a stopping time with $0 \leq \tau \leq 1$ so that $f - \mathbf{E} f \leq \sup_{0 \leq t \leq 1} W(t)$. Since the last supremum is distributed as $|W(1)|$, they deduced from this the following:

$$\gamma_n\{f - \mathbf{E} f \geq h\} \leq 2(1 - \Phi(h)).$$

Applying the above to $-f$, we get an analogue of (1.3) which is of course stronger than (2.3) for large h:

$$\gamma_n\{|f - \mathbf{E}f| \geq h\} \leq 4(1 - \Phi(h)). \tag{2.4}$$

Different proofs of (2.2) based on semigroup techniques and on Gross' logarithmic inequality were given by M. Ledoux ([L1], [L2]). A stronger infimum-convolution inequality going back to [T] and [M] can be found in [BG]. We now precise the argument used in derivation of (2.2) from (2.1) and show in particular how the constant 4 in (2.4) can be improved.

Theorem 2.2 *For every Lipschitz function f on $\mathbf{R}^n$ with $\|f\|_{\rm Lip} \leq 1$, the function*

$$T_f(h) = e^{h^2/2}\, \mathbf{E}(f - \mathbf{E}f)\, \mathbf{1}_{\{f - \mathbf{E}f \geq h\}}$$

is non-increasing in $h \geq 0$. In particular, for all $h > 0$,

$$\gamma_n\{f - \mathbf{E}f \geq h\} \leq \mathbf{E}(f - \mathbf{E}f)^+\, \frac{e^{-h^2/2}}{h}, \tag{2.5}$$

$$\gamma_n\{|f - \mathbf{E}f| \geq h\} \leq \mathbf{E}|f - \mathbf{E}f|\, \frac{e^{-h^2/2}}{h}. \tag{2.6}$$

Note that T_f is a constant function when f is linear. To connect (2.6) with (2.4), let us recall Pisier's inequality $\mathbf{E}|f - \mathbf{E}f| \leq \frac{2}{\sqrt{2\pi}}\, \mathbf{E}|\nabla f|$ (cf.[P]) from which it follows that, $\mathbf{E}|f - \mathbf{E}f| \leq \frac{2}{\sqrt{2\pi}}$ when $\|f\|_{\rm Lip} \leq 1$. Hence, the inequality (2.6) implies that

$$\gamma_n\{|f - \mathbf{E}f| \geq h\} \leq \frac{2}{\sqrt{2\pi}}\, \frac{e^{-h^2/2}}{h}.$$

Recalling the asymptotic

$$1 - \Phi(h) = \frac{e^{-h^2/2}}{h\sqrt{2\pi}}\, (1 + {\rm o}(1)), \quad \text{as} \quad h \to +\infty,$$

we also obtain

$$\gamma_n\{|f - \mathbf{E}f| \geq h\} \leq 2\, (1 - \Phi(h))\, (1 + {\rm o}(1)), \tag{2.7}$$

as $h \to +\infty$. Thus, for h large, (2.7) is a little better than (2.4). As in (1.3), here the constant 2 is optimal, as the example of linear functions shows. It can be made even smaller in some special situations when we have additional information about the value of $\mathbf{E}|f - \mathbf{E}f|$ (e.g. for functions $f(x) = \max_{i \leq n} x_i$).

Similar inequalities may be written for one-sided deviations, but we cannot then apply the Pisier inequality. Instead, in order to get a sharp constant, one should apply a related result of Pinelis [Pin2] showing that, for every convex function Ψ

on $\mathbf{R}$, the value of $\mathbf{E}\Psi(f-\mathbf{E}f)$ is maximal for linear functions (of euclidean norm 1). In particular, taking $\Psi(x)=x^+$, we obtain that $\mathbf{E}(f-\mathbf{E}f)^+ \leq \frac{1}{\sqrt{2\pi}}$. Thus, by (2.6), for all $h>0$,

$$\gamma_n\{f-\mathbf{E}f \geq h\} \leq \frac{1}{\sqrt{2\pi}}\,\frac{e^{-h^2/2}}{h} = (1-\Phi(h))\,(1+\mathrm{o}(1)).$$

Proof of Theorem 2.2 One may assume that $\mathbf{E}f=0$ and that moreover as a random variable on $(\mathbf{R}^n,\gamma_n)$ f has a continuous positive density p on the whole real line. Otherwise one can apply the statement of the theorem to functions $f_\delta(x_1,\cdots,x_n,x_{n+1}) = (1-\delta)f(x_1,\cdots,x_n)+\delta x_{n+1}$, $0<\delta<1$, and next let δ tend to 0. As in getting (2.2), applying (2.1) to $g=U(f)$, where U is a non–decreasing (piecewise) differentiable function on $\mathbf{R}$, gives

$$\mathbf{E}f\,U(f) \leq \mathbf{E}U'(f). \tag{2.8}$$

Let F be the distribution function of f. Given $h>0$ and $\varepsilon>0$, applying (2.8) to the function $U(x)=\min((x-h)^+,\,\varepsilon)$ leads to

$$\int_h^{h+\varepsilon} x(x-h)\,dF(x) + \varepsilon\int_{h+\varepsilon}^{\infty} x\,dF(x) \leq F(h+\varepsilon)-F(h)$$

Dividing by ε and letting ε tend to 0, we get, for all $h>0$,

$$\int_h^\infty x\,dF(x) \leq p(h).$$

Thus, the function $V(h) = \int_h^\infty x\,dF(x) = \int_h^\infty xp(x)\,dx$ satisfies the differential inequality $V(h) \leq -V'(h)/h$, i.e.,

$$(\log V(h))' \leq (-h^2/2)'.$$

This is equivalent to saying that the function $\log V(h)+h^2/2$ does not increase, and therefore the function $T_f(h)=V(h)\exp(h^2/2)$ is also non–increasing. Combining (2.5) with the same inequality applied to $-f$, we get (2.6). Theorem 2.2 is proved. $\square$

We now present the proof of the exponential inequality (1.5).

Theorem 2.3. *For any smooth function f defined on $\mathbf{R}^n$, and such that $\mathbf{E}f=0$,*

$$\mathbf{E}e^f \leq \mathbf{E}e^{|\nabla f|^2}.$$

Proof. Define the constant β by

$$\mathbf{E}_{\pi_n}e^{\langle\nabla f(X),\nabla f(Y)\rangle} = e^\beta,$$

where π_n is as in (1.6). Defining β is equivalent to saying that for all $g = g(x, y) \geq 0$,

$$\mathbf{E}_{\pi_n}\left(\langle \nabla f(X), \nabla f(Y)\rangle - \beta\right) g \leq \mathrm{Ent}_{\pi_n}\, g = \sup_{\{h:\mathbf{E}_{\pi_n} e^h \leq 1\}} \mathbf{E}_{\pi_n} gh$$

Taking $g = e^{f(y)}$ gives:

$$\begin{aligned}
\mathbf{E} f e^f - \beta \mathbf{E} e^f &= \mathrm{cov}(f, e^f) - \beta \mathbf{E} e^f \\
&= \mathbf{E}_{\pi_n} \langle \nabla f(X), \nabla f(Y)\rangle e^{f(Y)} - \beta \mathbf{E} e^f \\
&\leq \mathrm{Ent}\ e^f \\
&= \mathbf{E} f e^f - \mathbf{E} e^f \log \mathbf{E} e^f.
\end{aligned}$$

Hence, $\beta \geq \log \mathbf{E} e^f$, i.e., $e^\beta \geq \mathbf{E} e^f$, and thus,

$$\mathbf{E} e^f \leq \mathbf{E}_{\pi_n} e^{\langle \nabla f(X), \nabla f(Y)\rangle} \leq \mathbf{E}_{\pi_n} e^{\frac{|\nabla f(X)|^2 + |\nabla f(Y)|^2}{2}}.$$

The Cauchy–Schwarz inequality finishes the proof. □

To finish this section, we turn to a characterization of Gaussian measures via (1.5).

Theorem 2.4 *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbf{R}^n$, $n \geq 2$, and assume there exists a finite measure π on $\mathbf{R}^n \times \mathbf{R}^n$ such that, for all bounded smooth functions f and g on $\mathbf{R}^n$,*

$$\mathrm{cov}(f(X), g(X)) = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \langle \nabla f\,(x), \nabla g\,(y)\rangle\; d\pi(x, y). \tag{2.9}$$

Then, for some $a \in \mathbf{R}^n$ and $\sigma \geq 0$, X has characteristic function

$$\psi(v) \equiv \mathbf{E}\, e^{i\langle v, X\rangle} = e^{i\langle a, v\rangle + \sigma^2 \frac{|v|^2}{2}}, \qquad v \in \mathbf{R}^n.$$

Proof. For the functions $f(x) = e^{i\langle v_1, x\rangle}$, $g(x) = e^{i\langle v_2, x\rangle}$, the identity (2.9) has the form

$$\psi(v_1 + v_2) - \psi(v_1)\psi(v_2) = -\langle v_1, v_2\rangle\, \Psi(v_1, v_2),$$

where Ψ is Fourier transform of the measure π. In particular,

$$\langle v_1, v_2\rangle = 0 \Longrightarrow \psi(v_1 + v_2) = \psi(v_1)\psi(v_2).$$

This extends easily to n vectors $v_1, \ldots, v_n \in \mathbf{R}^n$:

$$\langle v_i, v_j\rangle = 0 \ \text{ for all } i \neq j \implies \psi(v_1 + \ldots + v_n) = \psi(v_1) \ldots \psi(v_n).$$

Hence, if $(e_i)_{1 \leq n}$ is orthonormal basis of $\mathbf{R}^n$, for all $t_1, \ldots, t_n \in \mathbf{R}$,

$$\psi(t_1 e_1 + \ldots + t_n e_n) = \psi(t_1 e_1) \ldots \psi(t_n e_n).$$

But this means that the random variables $\langle X, e_1 \rangle, \dots, \langle X, e_n \rangle$ are independent. In particular, $X_1, \dots, X_n$ are independent. Now by the assumption $n \geq 2$, we can take two linear forms of these independent random variables, say, $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$, and since Y_1 and Y_2 are independent, we conclude by the Darmois–Skitovitch Theorem (cf. [K-L-R]) that X_1 and X_2 are Gaussian. In the same way, all the X_i's are Gaussian. Now, the rest of the proof is clear. □

3 Covariance Representations on the Discrete Cube

For functions f on $\{-1,1\}^n$, we define a discrete gradient $\Delta f(x)$, $x \in \{-1,1\}^n$, as the vector $(\Delta_1 f(x), \dots, \Delta_n f(x))$ with coordinates

$$\Delta_k f(x) = \frac{f(x_1, \dots, x_{k-1}, 1, x_{k+1}, \dots, x_n) - f(x_1, \dots, x_{k-1}, -1, x_{k+1}, \dots, x_n)}{2}.$$

That is $\Delta_k f(x)$ represents a sign invariant normalized increment of f along the k-th coordinate. In particular, it does not depend on x_k. There is another important difference operator $\nabla f(x) = (\nabla_1 f(x), \dots, \nabla_n f(x))$, defined by

$$\nabla_k f(x) = x_k \Delta_k f(x) = \frac{f(x) - f(s_k(x))}{2}$$

where $[s_k(x)]_i = \begin{cases} x_i, & i \neq k \\ -x_k, & i = k \end{cases}$, is the neighbor of x along the kth coordinate.

Recall that the measure ν_n on $\{-1,1\}^n \times \{-1,1\}^n$ is defined by its density

$$\frac{d\nu_n(x,y)}{d\mu_n(x) d\mu_n(y)} = \int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i)\, d\alpha,$$

with respect to the uniform probability measure μ_n on $\{-1,1\}^n$. Clearly, ν_n is a probability measure with marginals μ_n. In the rest of the paper, expectations, variances and covariances of functions on $\{-1,1\}^n$ will always be understood with respect to μ_n. We turn now to the basic covariance representation:

Theorem 3.1 *For all functions f and g on $\{-1,1\}^n$,*

$$\mathrm{cov}(f,g) = \int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \langle \Delta f(x), \Delta g(y) \rangle \, d\nu_n(x,y). \tag{3.1}$$

As in the Gaussian case, (3.1) can be rewritten in terms of the semigroup Q_t where $Q_t(u_1, \dots, u_n) = (Q_t u_1, \dots, Q_t u_n)$ and becomes

$$\begin{aligned} \mathrm{cov}(f,g) &= \int_0^\infty \mathbf{E}\langle \nabla f, Q_t(\nabla g) \rangle \, dt \\ &= \int_0^\infty \mathbf{E}\langle \nabla f, \nabla Q_t(g) \rangle \, dt \\ &= \int_0^\infty \mathbf{E}\langle \Delta f, \Delta Q_t(g) \rangle \, dt. \end{aligned} \tag{3.2}$$

For large values of n, the covariance functional can also be developed in terms of the difference operators of higher orders (see also [H-PA], [L3], [H-P-S], [H]). We illustrate this general procedure with the example of the second order difference operators $\Delta_{ij} f(x) = \Delta_i(\Delta_j f(x))$, where we assume that $n \geq 2$, and $1 \leq i, j \leq n$. Note that $\Delta_{ij} f = \Delta_{ji} f$, that this function does not depend on (x_i, x_j) and that $\Delta_{ii} f = 0$. When $n = 2$,

$$\Delta_{1,2} f(x) = \frac{f(1,1) - f(1,-1) - f(-1,1) + f(-1,1)}{4},$$

and a similar expression should be written in the general case. Introduce the matrix $\Delta^2 f(x) = (\Delta_{ij} f(x))_{i,j=1}^n$ and note that the trace $\mathrm{tr}(\Delta^2 f(x) \Delta^2 g(y)) = \sum_{i,j} \Delta_{ij} f(x) \Delta_{ij} g(y)$. With these notations, we have:

Theorem 3.2 *For all functions f and g on $\{-1,1\}^n$, $n \geq 2$,*

$$\mathrm{cov}(f,g) = \langle \mathbf{E}\Delta f, \mathbf{E}\Delta g \rangle + \int_{\{-1,1\}^n} \int_{\{-1,1\}^n} \mathrm{tr}(\Delta^2 f(x) \Delta^2 g(y)) \, d\lambda_n(x,y), \tag{3.3}$$

where $\dfrac{d\lambda_n(x,y)}{d\mu_n(x) d\mu_n(y)} = \displaystyle\int_0^1 \prod_{i=1}^n (1 + \alpha x_i y_i)(1-\alpha) \, d\alpha.$

Both sides of the identities (3.1)–(3.3) are bilinear forms in f, g, so to prove them it suffices to verify their validity for some generating system of functions, e.g. for the Walsh's functions (we are then in a typical situation where it is more difficult to find an identity than to prove it).

The Walsh functions $\varepsilon_\pi : \{-1,1\}^n \longrightarrow \{-1,1\}$ are defined for an arbitrary set $\pi \subset \{1, \ldots, n\}$ by

$$\varepsilon_\pi(x) = \begin{cases} \prod_{i \in \pi} x_i, & \pi \neq \emptyset, \\ 1, & \pi = \emptyset. \end{cases}$$

For simplicity of notation, we set $\varepsilon_k = \varepsilon_{\{k\}}$, $k = 1, \ldots, n$, and thus $\varepsilon_\pi = \prod_{i \in \pi} \varepsilon_i$ for $\pi \neq \emptyset$. Also, we denote by $\mathbf{1}_\pi$ the indicator function of π, while $\mathbf{1}_{\pi_1 = \pi_2}$ is understood as 1 if $\pi_1 = \pi_2$, and 0 otherwise.

Let us recall some elementary properties of the Walsh functions:

1) $\varepsilon_{\pi_1} \varepsilon_{\pi_2} = \varepsilon_{\pi_1 \Delta \pi_2}$, where Δ is the symmetric difference of the corresponding sets;

2) $\mathbf{E}\varepsilon_\pi = 0$, if $\pi \neq \emptyset$; $\quad \mathbf{E}\varepsilon_\emptyset = 1$;

3) $\mathbf{E}\varepsilon_{\pi_1} \varepsilon_{\pi_2} = \mathbf{1}_{(\pi_1 = \pi_2)}$;

4) $\nabla_k \varepsilon_\pi = \varepsilon_\pi \mathbf{1}_\pi(k)$, for all $k = 1, \ldots, n$;

5) $\Delta_k \varepsilon_\pi = \varepsilon_{\pi \setminus \{k\}} \mathbf{1}_\pi(k)$, for all $k = 1, \ldots, n$;

6) $(\Delta^2 \varepsilon_\pi)_{ij} \equiv \Delta_{ij}\varepsilon_\pi = \varepsilon_{\pi\setminus\{i,j\}}\mathbf{1}_\pi(i)\mathbf{1}_\pi(j)$, for all $i \neq j$, $1 \leq i,j \leq n$ $(n \geq 2)$. In particular, $\Delta_{ij}\varepsilon_\pi = 0$, if $|\pi| = \mathrm{card}\ \pi \leq 1$.

We are now ready to prove Theorems 3.1 and 3.2.

Proof of (3.1)**.** Let $f = \varepsilon_{\pi_1}$, $g = \varepsilon_{\pi_2}$. We may also assume that $\pi_1, \pi_2 \neq \emptyset$. Then $\mathrm{cov}(f,g) = \mathbf{1}_{\pi_1 \neq \pi_2}$. On the other hand,

$$\begin{cases} \Delta_k f(x) & = \Delta_k \varepsilon_{\pi_1}(x) = \varepsilon_{\pi_1\setminus\{k\}}(x)\mathbf{1}_{\pi_1}(k), \\ \Delta_k g(y) & = \Delta_k \varepsilon_{\pi_2}(y) = \varepsilon_{\pi_2\setminus\{k\}}(y)\mathbf{1}_{\pi_2}(k), \end{cases}$$

so,

$$\begin{aligned} \langle \Delta f(x), \Delta g(y)\rangle &= \sum_{k=1}^n \Big(\varepsilon_{\pi_1\setminus\{k\}}(x)\mathbf{1}_{\pi_1(k)}\Big)\Big(\varepsilon_{\pi_2\setminus\{k\}}(g)\mathbf{1}_{\pi_2}(k)\Big) \\ &= \sum_{k\in\pi_1\cap\pi_2} \varepsilon_{\pi_1\setminus\{k\}}(x)\varepsilon_{\pi_2\setminus\{k\}}(y). \end{aligned}$$

In addition, $\prod_{i=1}^n (1+\alpha x_i y_i) = \sum_\pi \alpha^{|\pi|}\varepsilon_\pi(x)\varepsilon_\pi(y)$, hence,

$$\begin{aligned} &\prod_{i=1}^n (1+\alpha x_i y_i)\langle \Delta f(x), \Delta g(y)\rangle \\ &= \sum_\pi \sum_{k\in\pi_1\cap\pi_2} \alpha^{|\pi|}\varepsilon_\pi(x)\varepsilon_{\pi_1\setminus\{k\}}(x)\varepsilon_\pi(y)\varepsilon_{\pi_2\setminus\{k\}}(y). \end{aligned}$$

Thus,

$$\begin{aligned} &\int_{\{-1,1\}^n}\int_{\{-1,1\}^n} \prod_{i=1}^n (1+\alpha x_i y_i)\langle \Delta f(x), \Delta g(y)\rangle \, d\mu_n(x) d\mu_n(y) \\ &= \sum_\pi \sum_{k\in\pi_1\cap\pi_2} \alpha^{|\pi|}\mathbf{E}\varepsilon_\pi\varepsilon_{\pi_1\setminus\{k\}}\mathbf{E}\varepsilon_\pi\varepsilon_{\pi_2\setminus\{k\}} \\ &= \sum_\pi \sum_{k\in\pi_1\cap\pi_2} \alpha^{|\pi|}\mathbf{1}_{(\pi=\pi_1\setminus\{k\})}\mathbf{1}_{(\pi=\pi_2\setminus\{k\})} = 0, \qquad \text{if } \pi_1 \neq \pi_2. \end{aligned}$$

If $\pi_2 = \pi_1$ then above, one takes $\pi = \pi_1\setminus\{k\}$ and after after interchanging the sums $(\sum_\pi\sum_k = \sum_k\sum_\pi)$, this is equal to $\sum\limits_{k\in\pi_1} \alpha^{|\pi_1|-1} = |\pi_1|\alpha^{|\pi_1|-1}$. Finally, $\int_0^1 |\pi_1|\alpha^{|\pi_1|-1}d\alpha = 1$, and the result follows. $\square$

Proof of (3.3)**.** Again, let $f = \varepsilon_{\pi_1}$, $g = \varepsilon_{\pi_2}$, $\pi_1, \pi_2 \neq \emptyset$. If $|\pi_1| = 1$, or $|\pi_2| = 1$ then $\Delta_{ij}f = 0$, resp., $\Delta_{ij}g = 0$ for all i,j and the second term disappears. But in the first term, since

$$\begin{cases} \Delta_k\varepsilon_{\pi_1}(x) & = \varepsilon_{\pi_1\setminus\{k\}}(x)\mathbf{1}_{\pi_1}(k) \\ \Delta_k\varepsilon_{\pi_2}(x) & = \varepsilon_{\pi_2\setminus\{k\}}(x)\mathbf{1}_{\pi_2}(k) \end{cases}$$

we have

$$\begin{cases} \mathbf{E}\Delta_k\varepsilon_{\pi_1} & = \mathbf{1}_{(\pi_1\setminus\{k\}=\emptyset)}\mathbf{1}_{\pi_1}(k) = \mathbf{1}_{(\pi_i=\{k\})}, \\ \mathbf{E}\Delta_k\varepsilon_{\pi_2}(x) & = \mathbf{1}_{(\pi_2=\{k\})}. \end{cases}$$

So, in general

$$\begin{aligned}\langle \mathbf{E}\Delta\varepsilon_{\pi_1}, \mathbf{E}\Delta\varepsilon_{\pi_2}\rangle &= \mathrm{card}\{k \le n : \pi_1 = \pi_2 = \{k\}\} \\ &= \begin{cases} 1, & \text{if } \pi_1 = \pi_2 \text{ and } |\pi_1| = 1, \\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

Thus, (3.3) is proved when $|\pi_1| \le 1$ or $|\pi_2| \le 1$. Consider now the case $|\pi_1| \ge 2$, $|\pi_2| \ge 2$, in which case the first term in (3.3) disappears. Recall that $\Delta_{ii} = 0$, so assume that $i \ne j$. By the above property 6),

$$\begin{cases} \Delta_{ij}\varepsilon_{\pi_1} &= \varepsilon_{\pi_1\setminus\{i,j\}}\mathbf{1}_{\pi_1}(i)\mathbf{1}_{\pi_2}(j), \\ \Delta_{ij}\varepsilon_{\pi_2} &= \varepsilon_{\pi_2\setminus\{i,j\}}\mathbf{1}_{\pi_2}(i)\mathbf{1}_{\pi_2}(j),\end{cases}$$

hence,

$$\begin{aligned}\mathrm{tr}(\Delta^2\varepsilon_{\pi_1}(x)\Delta^2\varepsilon_{\pi_2}(g)) &= \sum_{i\ne j}\varepsilon_{\pi_1\setminus\{i,j\}}(x)\varepsilon_{\pi_2\setminus\{i,j\}}(y)\mathbf{1}_{\pi_1}(i)\mathbf{1}_{\pi_2}(i)\mathbf{1}_{\pi_1}(j)\mathbf{1}_{\pi_2}(j) \\ &= \sum_{\substack{i,j\in\pi_1\cap\pi_2 \\ i\ne j}} \varepsilon_{\pi_1\setminus\{i,j\}}(x)\varepsilon_{\pi_2\setminus\{i,j\}}(y).\end{aligned}$$

As in the proof of (3.1),

$$\begin{aligned}&\prod_{i=1}^n(1+\alpha x_i y_i)\mathrm{tr}(\Delta^2\varepsilon_{\pi_1}(x)\Delta^2\varepsilon_{\pi_2}(y)) \\ &= \sum_{\pi}\sum_{\substack{i,j\in\pi_1\cap\pi_2 \\ i\ne j}} \alpha^{|\pi|}\varepsilon_\pi(x)\varepsilon_{\pi_1\setminus\{i,j\}}(x)\varepsilon_\pi(y)\varepsilon_{\pi_2\setminus\{i,j\}}(y),\end{aligned}$$

therefore,

$$\begin{aligned}&\int_{\{-1,1\}^n}\int_{\{-1,1\}^n}\prod_{i=1}^n(1+\alpha x_i y_i)\mathrm{tr}(\Delta^2\varepsilon_{\pi_1}(x)\Delta^2\varepsilon_{\pi_2}(y))d\mu_n(x)d\mu_n(y) \\ &= \sum_{\pi}\sum_{\substack{i,j\in\pi_1\cap\pi_2 \\ i\ne j}} \alpha^{|\pi|}\mathbf{E}\varepsilon_\pi\varepsilon_{\pi_1\setminus\{i,j\}}\mathbf{E}\varepsilon_\pi\varepsilon_{\pi_2\setminus\{i,j\}} \\ &= \sum_{\pi}\sum_{\substack{i,j\in\pi_1\cap\pi_2 \\ i\ne j}} \alpha^{|\pi|}\mathbf{1}_{(\pi=\pi_1\setminus\{i,j\})}\mathbf{1}_{(\pi=\pi_2\setminus\{i,j\})} = 0,\end{aligned}$$

if $\pi_1 \ne \pi_2$. If $\pi_2 = \pi_1$, this last sum is equal to

$$\begin{aligned}\sum_{\substack{i,j\in\pi_1 \\ i\ne j}}\sum_{\pi}\alpha^{|\pi|}\mathbf{1}_{(\pi=\pi_1\setminus\{i,j\})} &= \alpha^{|\pi|-2}\sum_{\substack{i,j\in\pi_1 \\ i\ne j}} 1 \qquad (\text{since } |\pi| = |\pi_1| - 2) \\ &= |\pi_1|(|\pi_1|-1)\alpha^{|\pi|-2}.\end{aligned}$$

Finally, for $N = |\pi| \geq 2$

$$\int_0^1 N(N-1)\alpha^{N-2}(1-\alpha)d\alpha = 1.$$

Theorems 3.1 and 3.2 are proved. $\square$

Remark 3.3. 1) $\lambda_n(\{-1,1\}^n \times \{-1,1\}^n) = \int_0^1 (1-\alpha)\, d\alpha = 1/2$.

2) $2\lambda_n$ has marginals μ_n.

4 Large deviations for Bernoulli sums

Since on $\{-1,1\}^n \times \{-1,1\}^n$ the probability measure ν_n in (3.1) has marginals μ_n, as an immediate consequence of Theorem 3.1 and in complete similarity to the Gaussian case we obtain

Theorem 4.1 *For every functions f, g on $\{-1,1\}^n$ such that $|\Delta f(x)| \leq 1$, for all $x \in \{-1,1\}^n$,*

$$\mathrm{cov}(f,g) \leq \mathbf{E}|\Delta g|. \tag{4.1}$$

where the covariance and the expectation are taken with respect to μ_n.

This is the discrete analogue of the Gaussian inequality (2.1), and at the same time, it is more general: if we apply (4.1) in dimension nk to functions of the form $f\left(\dfrac{x_1 + \ldots + x_k}{\sqrt{k}}\right)$, $g\left(\dfrac{x_1 + \ldots + x_k}{\sqrt{k}}\right)$, where $x_1, \ldots, x_k \in \{-1,1\}^n$ and where f and g are smooth functions on $\mathbf{R}^n$ with $\|f\|_{\mathrm{Lip}} \leq 1$, in the limit we obtain (2.1) (by the central limit theorem in $\mathbf{R}^n$).

Now, let us see how to apply (4.1) to the problem of large deviations. Consider the Bernoulli sums $S_n = \varepsilon_1 + \ldots + \varepsilon_n$. Applying (4.1) to $f = \dfrac{S_n}{\sqrt{n}}$, $g = \mathbf{1}_{\{\frac{S_n}{\sqrt{n}} \geq h\}}$, we get

$$h\ \mu_n\Big\{\frac{S_n}{\sqrt{n}} \geq h\Big\} \leq \mathbf{E}\ \frac{S_n}{\sqrt{n}}\,\mathbf{1}_{\{\frac{S_n}{\sqrt{n}} \geq h\}} \leq \mu_n^+\Big\{\frac{S_n}{\sqrt{n}} \geq h\Big\}.$$

Here, we use the notation

$$\begin{aligned} \mu_n^+(A) &= \mathbf{E}\ |\Delta \mathbf{1}_A| \ = \ \mathbf{E}\ |\nabla \mathbf{1}_A| \\ &= \frac{1}{2}\,\mathbf{E}\,\mathrm{card}^{\frac{1}{2}}\{i \leq n : (x \in A, s_i(x) \notin A) \text{ or } (x \notin A, s_i(x) \in A)\} \end{aligned}$$

which defines the discrete perimeter of a set $A \subset \{-1,1\}^n$ ($\mathbf{1}_A$ is the indicator function of A). Thus, in order to get the Gaussian tails, i.e., the inequality of the form

$$\mu_n\Big\{\frac{S_n}{\sqrt{n}} \geq h\Big\} \leq K\,\frac{e^{-h^2/2}}{h}, \quad h \geq 1, \tag{4.2}$$

where K is a universal constant, we need just an estimate

$$\mu_n^+\left\{\frac{S_n}{\sqrt{n}} \geq h\right\} \leq K\, e^{-h^2/2}. \tag{4.3}$$

Introduce $T_n = \frac{S_n + n}{n} = \delta_1 + \ldots + \delta_n$, where $\delta_i = (x_i + 1)/2$ so that T_n is the sum of $\{0,1\}$-valued independent Bernoulli random variables (with probability of success $1/2$). Then,

$$\left\{\frac{S_n}{\sqrt{n}} > h\right\} = \{T_n \geq k+1\}, \quad \text{where} \quad k = \left[\frac{h\sqrt{n}+n}{2}\right]$$

($[\,\cdot\,]$ stands for the integer part of a real number). Thus, $\mu_n^+\left\{\frac{S_n}{\sqrt{n}} > h\right\} = \mu_n^+\{T_n \geq k+1\}$. Clearly, $k \geq n/2$ since $h \geq 1$. By the above definition of the discrete perimeter, for $k \leq n-1$,

$$\mu_n^+\{T_n \geq k+1\} = \frac{1}{2}\left(\sqrt{k+1}\ \mu_n\{T_n = k+1\} + \sqrt{n-k}\ \mu_n\{T_n = k\}\right),$$

and $\mu_n^+\{T_n \geq k\} = 0$, for $k \geq n$. Hence, only the values $k \in [n/2, n-1]$ and thus $n \geq 2$, need to be considered. Next, since $\mu_n\{T_n = k+1\} = \dfrac{n-k}{k+1}\,\mu_n\{T_n = k\}$, we may write

$$\begin{aligned}
\mu_n^+\{T_n \geq k+1\} &= \frac{1}{2}\left(\frac{n-k}{\sqrt{k+1}} + \sqrt{n-k}\right)\mu_n\{T_n = k\} \\
&= \frac{\sqrt{n-k}}{2}\left(\sqrt{\frac{n-k}{k+1}} + 1\right)\mu_n\{T_n = k\} \\
&\leq \sqrt{n-k}\ \mu_n\{T_n = k\},
\end{aligned}$$

where at the last step we have also made of use of $\dfrac{n-k}{k+1} \leq 1$. Now, to prove (4.3), we need an estimate of the form

$$\sqrt{n-k}\ \mu_n\{T_n = k\} \leq K\, e^{-h^2/2}.$$

And indeed, it can be obtained by a standard argument using Stirling's formula.

One could try to prove the inequality (4.2) for arbitrary linear functions f, using Theorem 4.1 and the same argument. However, we do not know how to prove an appropriate generalization of the anti-isoperimetric inequality (4.3). Nevertheless, we would like to show here an alternative simple proof of Pinelis' inequality:

Theorem 4.2 ([Pin1]) *There exists a universal constant K such that, for all linear functions $f(x) = a_1x_1 + \cdots + a_nx_n$ with $a_1^2 + \cdots + a_n^2 = 1$, we have*

$$\mu_n\{f \geq h\} \leq K(1 - \Phi(h)), \quad h > 0. \tag{4.4}$$

Proof. For $0 < h \leq \sqrt{3}$, since $\mu_n\{f \geq h\} \leq 1/2$, the inequality (4.4) holds with $K = 1/(2(1-\Phi(\sqrt{3}))$. Thus, we may assume that $h \geq \sqrt{2}$. For these values, we prove (4.4) by induction over n. The case $n = 1$ is clear. Let $n > 1$ and assume (4.4) holds for $n-1$. If $\sqrt{2} \leq h \leq \sqrt{3}$, there is nothing to prove, as explained above. Let $h \geq \sqrt{3}$. Assume that all $a_i \geq 0$ and set $a = a_n$, $b = \sqrt{1-a^2}$, so that $f(x) = bg(x_1, \cdots, x_{n-1}) + ax_n$, where g is a linear functional of euclidean norm 1. Thus,

$$\mu_n\{f \geq h\} = \frac{\mu_{n-1}\{g \geq \frac{h-a}{b}\} + \mu_{n-1}\{g \geq \frac{h+a}{b}\}}{2}.$$

Since for $h \geq \sqrt{3}$, $\frac{h-a}{b} \geq \sqrt{2}$, and therefore we may apply the induction hypothesis, in order to perform the induction step, we need just to verify that, for all $h \geq \sqrt{3}$ and for all $a, b \geq 0$ with $a^2 + b^2 = 1$,

$$\frac{\Phi(\frac{h-a}{b}) + \Phi(\frac{h+a}{b})}{2} \geq \Phi(h).$$

It suffices to see that the function $u(a) = \Phi(\frac{h-a}{b}) + \Phi(\frac{h+a}{b})$ is not decreasing in $0 \leq a < 1$. Since $b' = -\frac{a}{b}$, $(\frac{h-a}{b})' = -\frac{1-ah}{b^3}$, $(\frac{h+a}{b})' = -\frac{1+ah}{b^3}$, the inequality $u'(a) \geq 0$ is equivalent to

$$(1+ah)\varphi\left(\frac{h+a}{b}\right) \geq (1-ah)\varphi\left(\frac{h-a}{b}\right). \tag{4.5}$$

It is evident for $a \geq 1/h$. For $0 \leq a < 1/h$, taking the logarithm, (4.5) becomes

$$v(a) = \log(1+ah) - \log(1-ah) - \frac{2ah}{b^2} \geq 0.$$

After another differentiation, we see that $v'(a) \geq 0$ if and only if

$$\frac{1}{1-a^2h^2} \geq \frac{1+a^2}{b^4},$$

i.e., if and only if $a^2(1+h^2) + h^2 - 3 \geq 0$. Thus, v increases if $h \geq \sqrt{3}$, and since $v(0) = 0$, we get $v(a) \geq 0$, for all $a \in [0,1]$. Theorem 4.2 follows.

References

[B] Bakry, D. L' hypercontractivité et son utilisation en théorie des semi–groupes *École d'Été de Probabilités de St-Flour* (1992), Lecture Notes in Math., **1581**, Springer, (1994).

[B-L] Bakry, D., Ledoux, M. Lévy–Gromov isoperimetric inequality for an infinite-dimensional diffusion generator. *Invent. math.*, 123 (1996), 259–281.

[B-G] Bobkov, S.G., Götze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163 (1999), 1–28.

[Bor] Borell, C. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.* 30 (1975), 207–211.

[I-S-T] Cirel'son, B.S., Ibragimov, I.A., Sudakov, V.N., Norms of Gaussian sample functions. *Proceedings of the third Japan-USSR Symposium on Probability Theory. Lecture Notes in Math.* Springer-Verlag, 550 (1976), 20–41.

[HP] Herbst, I. and Pitt, L. D. (1991) Diffusion equation techniques in stochastic monotonicity and positive correlations. *Prob. Theory Rel. Fields* **87**, 275–312.

[H] Houdré, C. Comparison and deviation from a representation formula. In *Stochastic processes and related topics*, 207–218, Birkhäuser, Boston, Boston, MA.

[H-PA] Houdré, C., Pérez-Abreu, V. Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.* 23 (1995), 400–419.

[H-P-S] Houdré, C., Pérez-Abreu, V., Surgailis, D. Interpolation, correlation identities and inequalities for infinitely divisible processes. Preprint (1996). To appear in: Journal of Fourier Analysis and Applications.

[K-L-R] Kagan, A.M., Linnik, Yu. V., Rao, C.R. Characterizations problems in mathematical statistics. John Wiley and sons, New York, 1973.

[L1] Ledoux, M. A heat semigroup approach to concentration on the sphere and on a compact Riemannian manifold. *Geom. and Funct. Anal.*, 2 (1992), No.2, 444–465.

[L2] Ledoux, M. Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bull. Sci. math.* 118 (1994), 485–510.

[L3] Ledoux, M. L'algèbre de Lie des gradients itérés d'un générateur Markovien-développements de moyennes et entropies. *Ann. Scient. Éc. Norm. Sup.* 28 (1995), 435-460.

[L4] Ledoux, M. Isoperimetry and Gaussian Analysis. *Ecole d'été de Probabilités de Saint-Flour.* (1994). *Lect. Notes in Math.*, 1648 (1996), 165–294.

[L5] Ledoux M. A short proof of the Gaussian isoperimetric inequality. In *High dimensional probability (Oberwolfach, 1996)*, 229–232, Birkhäuser, Basel, 1998.

[L6] Ledoux M. Concentration of measure and logarithmic Sobolev inequalities. Preprint (1997).

[Pin1] Pinelis, I. Extremal probabilistic problems and Hotelling's T^2 test under a symmetry condition. *Ann. Stat.*, 22 (1994), No.1, 357–368.

[Pin2] Pinelis, I. Optimal tail comparison based on comparison of moments. In *High dimensional probability (Oberwolfach, 1996)*, 297–314, Birkhäuser, Basel.

[P] Pisier, G. Probabilistic methods in the geometry of Banach spaces. Probability and Analysis, Varenna (Italy) 1985. *Lect. Notes Math.* 1206 (1986), 167–241.

[S-T] Sudakov, V.N., Tsirel'son, B.S. Extremal properties of half-spaces for spherically invariant measures. *J. Soviet Math.* 9 (1978), 9–18. Translated from: *Zap. Nauch. Sem. L.O.M.I.* 41 (1974), 14–24 (Russian).

S.G.B.: Department of Mathematics, Syktyvkar University, Syktyvkar 167001, Russia

F.G.: Department of Mathematics, Bielefeld University, Bielefeld 33501, Germany

C.H. School of Mathematics, Georgia Institute of Technology, Atlanta. Georgia 30332, USA