Integrability of norms of Gaussian random vectors and processes

Let X be a centered Gaussian real random variable with variance σ^2 defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. It is clear from the expression of the density $\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}x^2}$, $x \in \mathbb{R}$, of its law that $\mathbb{E}(e^{\alpha X^2}) < \infty$ if and only if $\alpha < \frac{1}{2\sigma^2}$.

If $X = (X_1, \ldots, X_n)$ is a centered Gaussian random vector, and $S = \max_{1 \leq k \leq n} X_k$ (or $S = \max_{1 \leq k \leq n} |X_k|$),

$$\mathbb{E}\big(e^{\alpha S^2}\big) \,\leq\, \mathbb{E}\big(e^{\alpha \max_{1 \leq k \leq n} X_k^2}\big) \,\leq\, \sum_{k=1}^{n} \mathbb{E}\big(e^{\alpha X_k^2}\big) \,<\, \infty$$

for some $\alpha > 0$. As a consequence, for any norm $\| \cdot \|$ on $\mathbb{R}^n$, there exists $\alpha > 0$ such that $\mathbb{E}(e^{\alpha \|X\|^2}) < \infty$, although the optimal value of α is perhaps less immediate.

Consider now, on $(\Omega, \mathcal{A}, \mathbb{P})$, a centered Gaussian sequence X_k, $k \geq 1$, that is $(X_{k_1}, \ldots, X_{k_n})$ is a (centered) Gaussian vector in $\mathbb{R}^n$ for any $k_1, \ldots, k_n \geq 1$, such that $S = \sup_{k \geq 1} X_k < \infty$ almost surely. What can be said about the integrability properties of S ? As a main result discussed in this note,

$$\mathbb{E}\big(e^{\alpha S^2}\big) < \infty \qquad \text{if and only if} \qquad \alpha < \frac{1}{2\sigma^2}$$

where $\sigma^2 = \sup_{k \geq 1} \mathbb{E}(X_k^2) < \infty$.

Names attached to the history of this statement are J.-P. Kahane, A. Skorokhod, X. Fernique, H. Landau, L. Shepp, M. Marcus. The result may be addressed in the convenient

Drafted by M. L. **v1** December 2023

setting of Gaussian random vectors with values in a (real separable) Banach, or, equivalently, for Gaussian processes. It turns out that the only necessary key to achieve the conclusion is the rotational invariance of Gaussian measures, expressed by the fact that if Y and Z are independent centered Gaussian vectors, for any $\theta \in \mathbb{R}$, $Y(\theta) = Y\sin(\theta) + Z\cos(\theta)$ and $Y'(\theta) = Y\cos(\theta) - Z\sin(\theta)$ are independent with the same law as Y.

Throughout the note, all Gaussian vectors and processes will be centered, without further notice. Most of the material may be found in the classical references [10, ?, 9, 6, 4, 12].

Table of contents

1 A first strong integrability theorem

Let thus E be real separable Banach space equipped with its Borel σ-algebra $\mathcal{B}$, and with norm $\|\cdot\|$. A random variable, or vector, X on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is Gaussian if its law, on the Borel sets of E, is Gaussian, that is, for every element ξ of the dual space E^* of E, $\langle \xi, X\rangle$ is a real Gaussian variable (cf. [1]).

Since E is separable, the norm $\|\cdot\|$ on E may be described as a supremum over a countable set $(\xi_k)_{k\geq 1}$ of elements of the unit ball of the dual space E^*, that is, for every x in E,

$$\|x\| = \sup_{k\geq 1} \langle \xi_k, x\rangle.$$

The framework thus conveniently covers the case of random vectors and supremum of Gaussian processes (sequences).

The following basic integrability property is going back to X. Fernique [5] and H. Landau and L. Shepp [7], after preliminary observations by J.-P. Kahane and A. Skorokhod (cf. [10]).

Theorem 1. *Let X be a Gaussian vector with values in $(E, \mathcal{B}, \|\cdot\|)$. There exists $\alpha > 0$ such that*

$$\mathbb{E}\big(e^{\alpha\|X\|^2}\big) \,<\, \infty.$$

The proof of Theorem 1 in [5] is solely based on the aforementioned rotational invariance, while in [7] deeper isoperimetric-type arguments are developed.

The argument developed in the next section, due to B. Maurey and G. Pisier [14], and which extends the idea of X. Fernique in [5], provides a simple direct proof of this theorem. This argument actually covers a more general framework, and emphasizes at the same time the concentration inequality, for a Gaussian vector X,

$$\mathbb{P}\big(\big|\|X\| - \mathbb{E}(\|X\|)\big| \geq r\big) \,\leq\, 2\,e^{-\frac{2r^2}{\pi^2\sigma^2}}, \quad r \geq 0, \tag{1}$$

where $\sigma^2 = \sup_{\xi \in E^*, \|\xi\| \leq 1} \mathbb{E}(\langle \xi, X\rangle^2)$, which actually goes much beyond the integrability theorem itself. In particular, this concentration property of the norm of X around its mean arises at the exponential rate σ^2, and it is of fundamental importance to realize that $\mathbb{E}(\|X\|)$ is in general much bigger than σ, as may be checked on the example of the Euclidean norm of a standard normal vector X in $\mathbb{R}^n$ (for which $\mathbb{E}(\|X\|)$ is of the order of $\sqrt{n}$ while $\sigma = 1$).

The inequality (1) is actually part of the family of Gaussian concentration inequalities presented in the companion post [2].

2 The Fernique-Maurey-Pisier argument

Let E and F be finite-dimensional Banach spaces (think of $\mathbb{R}^n$ and $\mathbb{R}^m$, with arbitrary norms), and let $f : E \to F$ be locally Lipschitz. The function f has, at almost every point $y \in E$, a derivative $f'(y)$ which is a linear map from E into F. For z in E, denote by $f'(y) \cdot z$ the value of $f'(y)$ on z, so that

$$f'(y) \cdot z \,=\, \lim_{t \to 0} \frac{1}{t}\,\big[f(y + tz) - f(y)\big].$$

Proposition 2 (The Maurey-Pisier inequality)**.** *Let Y be a Gaussian random vector on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in E. Let $f : E \to F$ be locally Lipschitz, and $\Psi : F \to \mathbb{R}$ be a convex function such that $\Psi \circ f(Y)$ is integrable. Then*

$$\mathbb{E}\big(\Psi\big(f(Y) - \mathbb{E}\big(f(Y)\big)\big)\big) \,\leq\, \mathbb{E}\Big(\Psi\Big(\frac{\pi}{2}\, f'(Y) \cdot Z\Big)\Big) \tag{2}$$

where Z denotes an independent copy of Y.

Proof. Write

$$f(Y) - f(Z) = \int_0^{\pi/2} \frac{d}{d\theta} f\big(Y(\theta)\big) d\theta = \int_0^{\pi/2} f'\big(Y(\theta)\big) \cdot Y'(\theta) d\theta$$

where $Y(\theta) = Y \sin(\theta) + Z \cos(\theta)$, $Y'(\theta) = Y \cos(\theta) - Z \sin(\theta)$. By convexity of Ψ,

$$\Psi\big(f(Y) - f(Z)\big) \le \frac{2}{\pi} \int_0^{\pi/2} \Psi\Big(\frac{\pi}{2} f'\big(Y(\theta)\big) \cdot Y'(\theta)\Big) d\theta.$$

Take now expectation of both sides of this inequality. By the rotational invariance of Gaussian measures, for every fixed $\theta \in \mathbb{R}$, the couple $(Y(\theta), Y'(\theta))$ has the same distribution as the couple (Y, Z), so that

$$\mathbb{E}\big(\Psi\big(f(Y) - f(Z)\big)\big) \le \mathbb{E}\Big(\Psi\Big(\frac{\pi}{2} f'(Y) \cdot Z\Big)\Big).$$

It remains to observe that by Jensen's inequality (and independence of Y and Z),

$$\mathbb{E}\big(\Psi\big(f(Y) - f(Z)\big)\big) \ge \mathbb{E}\big(\Psi\big(f(Y) - \mathbb{E}\big(f(Y)\big)\big)\big).$$

□

An important feature of the Maurey-Pisier inequality (2) is that it is fully dimension-free, allowing thus for extensions to infinite-dimensional spaces. One such extension is considered next, but the Maurey-Pisier inequality goes beyond this illustration since vector-valued functions f may be considered. With respect to [5], it may be noticed that the Fernique proof uses rotational invariance of Gaussian random vectors for one angle, and on distribution functions, whereas the Maurey-Pisier inequality relies on a continuum of angles along measurable functions of the vectors.

Applied to $E = \mathbb{R}^n$, $F = \mathbb{R}$, Y with law $\mathcal{N}(0, \mathrm{Id})$ and to the family of convex functions $\Psi(u) = e^{\lambda u}$, $u \in \mathbb{R}$, $\lambda \in \mathbb{R}$, partial integration of the inequality (2) with respect to Z yields

$$\mathbb{E}\big(e^{\lambda[f(Y) - \mathbb{E}(f(Y))]}\big) = \mathbb{E}\big(e^{\frac{\pi^2}{8}\lambda^2 |f'(Y)|^2}\big).$$

As a consequence, the following corollary may be stated. For a function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

Corollary 3. *Let Y with law $\mathcal{N}(0, \mathrm{Id})$ on $\mathbb{R}^n$. For any Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ and any $\lambda \in \mathbb{R}$,*

$$\mathbb{E}\big(e^{\lambda[f(Y) - \mathbb{E}(f(Y))]}\big) \le e^{\frac{\pi^2}{8}\lambda^2 \|f\|_{\mathrm{Lip}}^2}.$$

The companion post [2] presents an improved version of this corollary (replacing $\frac{\pi^2}{8}$ by $\frac{1}{2}$) by different, although related, tools, and is used in Section 5 to reach the sharp integrability exponents.

3 Proof of the integrability Theorem 1

While Theorem 1 directly follows from the suitable infinite-dimensional version of Proposition 2, due to its dimension-free character, a simple finite-dimensional approximation provides an easy approach.

Given a Gaussian random vector X with values in E, the proof will, as announced, establish the stronger concentration inequality (1), including that $\mathbb{E}(\|X\|) < \infty$ and $\sigma < \infty$. Theorem 1 is then immediate by integration of this exponential tail inequality, with $\alpha < \frac{2}{\pi^2 \sigma^2}$.

The constant $\frac{\pi}{2}$ is optimal in the general form (2) of the Maurey-Pisier inequality, for vector-valued functions f. But, as developed in the next section, the optimal exponent in the integrability theorem may be achieved from the more precise Gaussian concentration inequalities.

The task is therefore to establish the concentration inequality (1), which may easily be accessed by a finite-dimensional argument on the basis of Corollary 3.

First observe that

$$\sigma^2 \,=\, \sup_{\xi \in E^*, \|\xi\| \leq 1} E\big(\langle \xi, X \rangle^2\big) \,<\, \infty.$$

Indeed, let $m > 0$ be such that $\mathbb{P}(\|X\| \leq m) \geq \frac{3}{4}$. Then, for every element ξ in E^* with $\|\xi\| \leq 1$, $\mathbb{P}(\langle \xi, X \rangle \leq m) \geq \frac{3}{4}$. Now $\langle \xi, X \rangle$ is real Gaussian with variance $\mathbb{E}(\langle \xi, X \rangle^2)$. Since $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1/2} e^{-\frac{1}{2}x^2} dx < \frac{3}{4}$, it follows that $[\mathbb{E}(\langle \xi, X \rangle^2)]^{1/2} \leq 2m$.

Recall then that, for every x in E, $\|x\| = \sup_{k \geq 1} \langle \xi_k, x \rangle$ where $(\xi_k)_{k \geq 1}$ is a sequence in the unit ball of the dual space E^*. For every integer $n \geq 1$, consider the (finite-dimensional) Gaussian vector $X_n = (\langle \xi_1, X \rangle, \ldots, \langle \xi_n, X \rangle)$ in $\mathbb{R}^n$. The vector X_n has the same distribution as AY where Y has law $\mathcal{N}(0, \mathrm{Id})$ and $\Sigma = A^{\top} A$ is the covariance matrix of X_n. For any measurable map $f : \mathbb{R}^n \to \mathbb{R}$, Corollary 3 yields that

$$\mathbb{E}\big(e^{\lambda [f(X) - \mathbb{E}(f(X))]}\big) \,=\, \mathbb{E}\big(e^{\lambda [f \circ A(Y) - \mathbb{E}(f \circ A(Y))]}\big) \,\leq\, e^{\frac{\pi^2}{8} \lambda^2 \|f \circ A\|_{\mathrm{Lip}}^2}$$

for every $\lambda \in \mathbb{R}$, provided that $\|f \circ A\|_{\mathrm{Lip}} < \infty$. Apply this inequality to the function

$$f(x) \,=\, f(x_1, \ldots, x_n) \,=\, \max_{1 \leq k \leq n} x_k \,=\, \|x\|_n.$$

For every $x, y \in \mathbb{R}^n$, $k = 1, \ldots, n$,

$$\big|(Ax)_k - (Ay)_k\big|^2 \,=\, \bigg| \sum_{\ell=1}^{n} A_{k\ell}(x_\ell - y_\ell) \bigg|^2 \,\leq\, \sum_{\ell=1}^{n} A_{k\ell}^2 \, |x - y|^2 \,=\, \Sigma_{kk} \, |x - y|^2$$

so that

$$\|f \circ A\|_{\mathrm{Lip}}^2 \,\leq\, \max_{1 \leq k \leq n} \mathbb{E}\big(\langle \xi_k, X \rangle^2\big) \,\leq\, \sup_{\xi \in E^*, \|\xi\| \leq 1} E\big(\langle \xi, X \rangle^2\big) \,=\, \sigma^2.$$

As a consequence of the preceding, for every $\lambda \in \mathbb{R}$ and every $n \geq 1$,

$$\mathbb{E}\big(e^{\lambda[\|X\|_n - \mathbb{E}(\|X\|_n)]}\big) \,\leq\, e^{\frac{\pi^2}{8}\lambda^2\sigma^2}.$$

By Markov's inequality, for every $r \geq 0$ and $\lambda \geq 0$,

$$\mathbb{P}\big(\|X\|_n - \mathbb{E}(\|X\|_n) \geq r\big) \,\leq\, e^{-\lambda r + \frac{\pi^2}{8}\lambda^2\sigma^2}.$$

After optimization in λ, it follows that

$$\mathbb{P}\big(\|X\|_n - \mathbb{E}(\|X\|_n) \geq r\big) \,\leq\, e^{-\frac{2r^2}{\pi^2\sigma^2}} \tag{3}$$

for every $r \geq 0$. Together with the same argument for $-f$, and the union bound,

$$\mathbb{P}\big(\big|\|X\|_n - \mathbb{E}(\|X\|_n)\big| \geq r\big) \,\leq\, 2\,e^{-\frac{2r^2}{\pi^2\sigma^2}}. \tag{4}$$

The final step is to let n tend to infinity, so to replace $\|X\|_n$ by $\|X\|$. A little care is necessary here to handle the expectations $\mathbb{E}(\|X\|_n)$. Let $m > 0$ be such that $\mathbb{P}(\|X\| \leq m) \geq \frac{1}{2}$, and let $r_0 > 0$ be such that $2\,e^{-2r_0^2/\pi^2\sigma^2} < \frac{1}{2}$. Hence, by (4), $\mathbb{E}(\|X\|_n) \leq m + r_0$ independently of n, and therefore $\mathbb{E}(\|X\|) < \infty$. By monotone convergence as $n \to \infty$, (4) (and similarly (3)) then turns into the announced (1).

Although this is not strictly necessary, to be on the (positive) safe side, it might be easier to work, exactly in the same way, with $\|x\| = \sup_{k\geq 1}|\langle \xi_k, x\rangle|$ and the functions on $\mathbb{R}^n$, $f(x) = \max_{1\leq k\leq n}|x_k|$. This comment is of interest for supremum of (centered) Gaussian processes.

Indeed, the integrability Theorem 1 is not specific to norms on Banach spaces, actually measurable semi-norms may be used similarly (really only basic convexity is used in the proof argument). In particular, all the preceding integrability properties apply to the supremum of a centered Gaussian process $(X_t)_{t\in T}$, assuming proper measurability assumptions to deal with suprema. For example, if T is countable, and $\sup_{t\in T}|X_t| < \infty$ almost surely, the conclusions apply then with $\|X\|$ replaced by this supremum $\sup_{t\in T}|X_t|$. As a statement, there exists $\alpha > 0$ such that

$$\mathbb{E}\big(e^{\alpha\,[\sup_{t\in T}|X_t|]^2}\big) \,<\, \infty. \tag{5}$$

The point being that the main argument relies on dimension-free finite dimensional inequalities, and the maximum function is used exactly in the same way. If necessary, further details are provided in [10, **?**, 9] for instance. Note that if $\sup_{t\in T} X_t < \infty$ almost surely only, by symmetry (due to the centering), $\sup_{t\in T}|X_t| < \infty$ almost surely (and thus also $\mathbb{E}(e^{\alpha\,[\sup_{t\in T} X_t]^2}) < \infty$).

4 Moment equivalence

The proof of Theorem 1, and more precisely (1), has an other interesting consequence as the equivalence of all moments of Gaussian random vectors. That is, for every $0 < p, q < \infty$, there exists a constant $C_{p,q} > 0$ only depending on p and q such that, for any Gaussian vector X with values in some Banach space E with norm $\|\cdot\|$,

$$\Big(\mathbb{E}\big(\|X\|^p\big)\Big)^{1/p} \le C_{p,q}\Big(\mathbb{E}\big(\|X\|^q\big)\Big)^{1/q}. \tag{6}$$

The claim easily follows from the integration in $t \ge 0$ of the concentration inequality (1), together with the fact that $\mathbb{E}(\|X\|) \le m + r_0$ and $\sigma \le C_p[\mathbb{E}(\|X\|^p)]^{1/p}$ for any $p > 0$. The result applies similarly to $\|X\| = \sup_{t\in T}|X_t|$ for a Gaussian process $X = (X_t)_{t\in T}$. It is shown in [8], as a consequence of the solution of the so-called S-conjecture (cf. [3]), that the constants $C_{p,q}$ are the same than in the real case.

5 Sharp integrability

The more precise Gaussian concentration inequalities (presented e.g. in [2]) show that Corollary 3 can be improved into

$$\mathbb{E}\big(e^{\lambda[f(Y)-\mathbb{E}(f(Y))]}\big) \le e^{\frac{\lambda^2}{2}\|f\|^2_{\mathrm{Lip}}}, \quad \lambda \in \mathbb{R}.$$

It is then clear that all the tail inequalities of the previous section are also improved from the factor $\frac{2}{\pi^2}$ to $\frac{1}{2}$ in the exponential. In particular the basic concentration inequality (1) actually reads

$$\mathbb{P}\big(\big|\|X\| - \mathbb{E}(\|X\|)\big| \ge r\big) \le 2\,e^{-\frac{r^2}{2\sigma^2}}, \quad r \ge 0. \tag{7}$$

The values $\alpha > 0$ in Theorem 1 thus satisfies $\alpha < \frac{1}{2\sigma^2}$, which actually describes the sharp exponent (first obtained in [13] by different means). Indeed, optimality follows from the one-dimensional case as, for every $\xi \in E^*$ with $\|\xi\| \le 1$,

$$\infty > \mathbb{E}\big(e^{\alpha\|X\|^2}\big) \ge \mathbb{E}\big(e^{\alpha\langle\xi,X\rangle^2}\big) = \frac{1}{\sqrt{\frac{1}{\mathbb{E}(\langle\xi,X\rangle^2)} - 2\alpha}},$$

so that indeed $\alpha < \frac{1}{2\sigma^2}$. The same holds true for (5) with $\alpha < \frac{1}{2\sup_{t\in T}\mathbb{E}(X_t^2)}$. Dealing with the partial supremum $S = \sup_{t\in T} X_t$ might require a few more details. For every $t \in T$ and $r \ge 0$,

$$1 - \Phi\bigg(\frac{r}{\sqrt{\mathbb{E}(X_t^2)}}\bigg) = \mathbb{P}(X_t \ge r) \le \mathbb{P}(S \ge r).$$

Hence, with $\sigma^2 = \sup_{t \in T} \mathbb{E}(X_t^2)$, for every $r \geq 0$,

$$1 - \Phi\Big(\frac{r}{\sigma}\Big) \leq \mathbb{P}(S \geq r),$$

a tail behavior which prevents the fact that $\mathbb{E}(e^{\alpha S^2}) < \infty$ for some $\alpha \geq \frac{1}{2\sigma^2}$.

These conclusions may be summarized in the following general statement, which describes the optimal integrability properties of norms of (centered) Gaussian vectors and supremum of Gaussian processes.

Theorem 4 (The Gaussian integrability theorem)**.** *Let X be a Gaussian vector on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a real separable Banach space $(E, \mathcal{B}, \| \cdot \|)$. Then*

$$\mathbb{E}\big(e^{\alpha \|X\|^2}\big) < \infty \qquad \textit{if and only if} \qquad \alpha < \frac{1}{2 \sup_{\|\xi\| \leq 1} \mathbb{E}(\langle \xi, X \rangle^2)} \, .$$

Similarly, if $X = (X_t)_{t \in T}$ is a (separable) Gaussian process on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $\sup_{t \in T} X_t$, or equivalently $\sup_{t \in T} |X_t|$, is finite almost surely, then

$$\mathbb{E}\big(e^{\alpha S^2}\big) < \infty \qquad \textit{if and only if} \qquad \alpha < \frac{1}{2 \sup_{t \in T} \mathbb{E}(X_t^2)}$$

where $S = \sup_{t \in T} X_t$ or $\sup_{t \in T} |X_t|$.

If a random vector X is not centered, the conclusion of Theorem 4 (applied to $X - \mathbb{E}(X)$) still holds true with $\sup_{\|\xi\| \leq 1} \mathrm{Var}(\xi, X\rangle)$ instead of $\sup_{\|\xi\| \leq 1} \mathbb{E}(\langle \xi, X \rangle^2)$. Things are a bit more delicate for processes, unless it is considered $S = \sup_{t \in T} |X_t|$.

References

[1] Admissible shift, reproducing kernel Hilbert space, and abstract Wiener space. *The Gaussian Blog.*

[2] Gaussian concentration inequalities. *The Gaussian Blog.*

[3] Some geometric inequalities for Gaussian measures. *The Gaussian Blog.*

[4] V. Bogachev. *Gaussian measures.* Math. Surveys Monogr. 62. American Mathematical Society (1998).

[5] X. Fernique. Intégrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris* 270, 1698–1699 (1970).

[6] S. Janson. *Gaussian Hilbert spaces*. Cambridge Tracts in Math. 129. Cambridge University Press (1997).

[7] H. Landau, L. Shepp. On the supremum of a Gaussian process. *Sankhyà: The Indian Journal of Statistics* 32, 369–378 (1970).

[8] R. Latała, K. Oleszkiewicz. Gaussian measures of dilatations of convex symmetric sets. *Ann. Probab.* 27, 1922–1938 (1999).

[9] M. Ledoux. Isoperimetry and Gaussian Analysis. *École d'Été de Probabilités de St-Flour 1994. Lecture Notes in Math.* 1648, 165–294. Springer (1996).

[10] M. Ledoux, M. Talagrand. *Probability in Banach spaces (Isoperimetry and processes)*. Ergebnisse der Mathematik und ihrer Grenzgebiete 23. Springer (1991).

[11] M. Lifshits. *Gaussian random functions*. Kluwer (1995).

[12] M. Lifshits. *Lectures on Gaussian processes*. SpringerBriefs Math. (2012).

[13] M. Marcus, L. Shepp. Sample behavior of Gaussian processes. *Proc. of the Sixth Berkeley Symposium on Math. Statist. and Probab.* 2, 423–441 (1972).

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