Gaussian concentration inequalities

Let γ_n be the standard Gaussian probability measure γ_n on $\mathbb{R}^n$, with density $\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}|x|^2}$, $x \in \mathbb{R}^n$, with respect to the Lebesgue measure. As a consequence of the Gaussian isoperimetric inequality [1], if $F : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function on $\mathbb{R}^n$, with Lipschitz coefficient $\|F\|_{\mathrm{Lip}}$, and if $m(F)$ is a median of F under γ_n, for every $r \geq 0$,

$$\gamma_n\big(|F - m(F)| \geq r\big) \,\leq\, e^{-r^2/2\|F\|^2_{\mathrm{Lip}}}. \tag{1}$$

This result is a prototype of a *concentration inequality*: under some smoothness assumption on a function $F : \mathbb{R}^n \to \mathbb{R}$, for example $\|F\|_{\mathrm{Lip}} \leq 1$, already for values of r of the order of 5 or 10, F is within r from a fixed value (median) with very high probability. In vigorous words, a (mildly) smooth function is almost constant on almost (in the sense of the measure γ_n) all the space. As developed below, (1) also holds with the mean $\int_{\mathbb{R}^n} F d\gamma_n$ instead of a median $m(F)$.

A significant feature of this inequality is that it holds for large families of functions, and is dimension-free, independent of the dimension of the underlying state space $\mathbb{R}^n$. (Dimension is actually hidden in the median or mean value, as witnessed for example by the function $F(x) = |x|$, $x \in \mathbb{R}^n$, for which $m(F)$ and $\int_{\mathbb{R}^n} F d\gamma_n$ are of the order of $\sqrt{n}$ while $\|F\|_{\mathrm{Lip}} = 1$.) As such, this and related concentration inequalities extend to arbitrary Gaussian measures, on finite or infinite-dimensional spaces. They give rise in particular to the sharp integrability properties of norms of Gaussian vectors and processes, emphasized in the note [2] and recalled below.

The Gaussian concentration inequalities are part of the more general concentration of measure phenomenon, a widely shared property which emerged within asymptotic geometric

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

analysis and applies to numerous probabilistic models depending on a large number of (independent) coordinates (cf. [6]).

While the concentration statement (1) is deduced from the somewhat delicate Gaussian isoperimetric inequality, simple functional analytic tools may be developed to achieve similar statements. It is the purpose of this post to briefly emphasize and present some of these arguments. A prior, weaker, form of the Laplace transform inequality is the Maurey-Pisier inequality presented in the companion post [2].

Table of contents

1 Concentration and isoperimetry

One basic concentration inequality arises as an natural consequence of the Gaussian isoperimetric inequality. Recall namely (cf. [1]) that if A is a Borel set in $\mathbb{R}^n$ such that $\gamma_n(A) \geq \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}x^2} dx$ for some $a \in \mathbb{R}$, it holds true that

$$\gamma_n(A_r) \,\geq\, \Phi(a+r) \tag{2}$$

for any $r \geq 0$, where A_r is the (closed) r-th neighborhood of A in the Euclidean metric.

Consider then a set A such that $\gamma_n(A) \geq \frac{1}{2} = \Phi(0)$. Then, for any $r \geq 0$,

$$\gamma_n(A_r) \,\geq\, \Phi(r) \,\geq\, 1 - \frac{1}{2}\, e^{-\frac{1}{2}r^2}. \tag{3}$$

That is, starting from a set A with $\gamma_n(A) \geq \frac{1}{2}$, for r already of the order of 5 or 10, the enlargement A_r has a measure close to 1, one illustration therefore of the terminology "concentration of measure". The value $\frac{1}{2}$ is nothing special at this stage, the claim would be similar for any $\gamma_n(A) \geq \varepsilon > 0$ (choosing r large enough depending on ε), cf. [6]. It will be useful nevertheless when dealing with medians of Lipschitz functions.

The result (3) extends to arbitrary, finite or infinite-dimensional, Gaussian measures μ on the basis of the isoperimetric inequality

$$\mu(A + r\,\mathcal{K}) \,\geq\, \Phi(a+r), \quad r \geq 0,$$

whenever A is a Borel set such that $\mu(A) \geq \Phi(a)$, $a \in \mathbb{R}$, and $\mathcal{K}$ is the unit ball of the reproducing kernel Hilbert space $\mathcal{H}$ associated to μ [3]. For example, if μ is finite-dimensional with covariance matrix $M^{\top}M$, $\mathcal{K}$ is the image by M of the Euclidean unit ball.

2 Concentration of Lipschitz functions

The preceding concentration property (3) on sets may be expressed equivalently on Lipschitz functions. If $F : \mathbb{R}^n \to \mathbb{R}$, its Lipschitz coefficient is defined by

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

(where $|\cdot|$ is the Euclidean length in $\mathbb{R}^n$).

By homogeneity, it is often convenient to deal with Lipschitz functions $F : \mathbb{R}^n \to \mathbb{R}$ such that $\|F\|_{\mathrm{Lip}} \leq 1$, called 1-Lipschitz in the sequel. The various inequalities derived below for 1-Lipschitz functions are immediately extended to arbitrary Lipschitz functions (as in (1)).

Let thus $F : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz, and let $m \in \mathbb{R}$ be such that $\gamma_n(F \leq m) \geq \frac{1}{2} = \Phi(0)$. It is immediate by the Lipschitz property that, for any $r \geq 0$,

$$\big(\{F \leq m\}\big)_r \subset \{F \leq m + r\}.$$

Hence, from (3),

$$\gamma_n(F \leq m + r) \,\geq\, \Phi(r) \,\geq\, 1 - \frac{1}{2}\, e^{-\frac{1}{2}r^2}.$$

Thus (and by continuity of the lower-bound),

$$\gamma_n(F \geq m + r) \,\leq\, 1 - \Phi(r) \,\leq\, \frac{1}{2}\, e^{-\frac{1}{2}r^2}, \quad r \geq 0. \tag{4}$$

The choice of $F = d(\cdot, A)$, where d is the Euclidean distance, shows that (4) is actually equivalent to (3).

Now, if $m(F)$ denotes a median of F, i.e. such that $\gamma_n(F \geq m(F)) \geq \frac{1}{2}$ and $\gamma_n(F \leq m(F)) \geq \frac{1}{2}$, the preceding inequality (4) applied to both F and $-F$ yields, by the union bound,

$$\gamma_n\big(|F - m(F)| \geq r\big) \,\leq\, e^{-\frac{1}{2}r^2}, \quad r \geq 0. \tag{5}$$

The Lipschitz function F therefore "concentrates" around the value $m(F)$ on a set of measure close to 1 for the large values of r.

For an arbitrary, finite or infinite-dimensional, Gaussian measures μ on a Banach space E (cf. [3]), the Lipschitz property of a function $F : E \to \mathbb{R}$ has to be understood with respect to the (reproducing kernel) Hilbertian structure $\mathcal{H}$ induced by the measure, that is

$$\big|F(x+h) - F(x)\big| \,\leq\, |h|_{\mathcal{H}}$$

for any $x \in E$, $h \in \mathcal{H}$.

3 Laplace transform inequality

While the preceding concentration inequalities are deduced from the isoperimetric property of Gaussian measures, a rather simple direct approach may be provided. This is the content of the following estimate on the Laplace transform of a given Lipschitz function.

Theorem 1 (Laplace transform inequality)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz. For any $\lambda \in \mathbb{R}$,*

$$\int_{\mathbb{R}^n} e^{\lambda F} d\gamma_n \,\leq\, e^{\lambda \int_{\mathbb{R}^n} F d\gamma_n + \frac{1}{2}\lambda^2}. \tag{6}$$

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

$$\gamma_n\big(F \geq \textstyle\int_{\mathbb{R}^n} F d\gamma_n + r\big) \,\leq\, e^{-\lambda r + \frac{1}{2}\lambda^2}.$$

After optimization in λ $(= r)$, for every $r \geq 0$,

$$\gamma_n\big(F \geq \textstyle\int_{\mathbb{R}^n} F d\gamma_n + r\big) \,\leq\, e^{-\frac{1}{2}r^2}. \tag{7}$$

Since (7) similarly holds for $-F$, by the union bound

$$\gamma_n\big(\big|F - \textstyle\int_{\mathbb{R}^n} F d\gamma_n\big| \geq r\big) \,\leq\, 2\, e^{-\frac{1}{2}r^2}, \quad r \geq 0. \tag{8}$$

Before turning to a proof of Theorem 1 in the next section, it has to be observed that the concentration inequalities on Lipschitz functions just deduced from it are rather close to the ones obtained in the previous section, with the mean instead of a median. Actually, integrating (5) with respect to $r \geq 0$ shows that $\big|\int_{\mathbb{R}^n} F\, d\gamma_n - m(F)\big| \leq \sqrt{\frac{\pi}{2}}$ for any median $m(F)$.

Conversely, choosing $r_0 > 0$ such that $2\,e^{-\frac{1}{2}r_0^2} < \frac{1}{2}$ in (8) shows that $\big|\int_{\mathbb{R}^n} F\,d\gamma_n - m(F)\big| \leq r_0$. Hence, up to numerical constants, both in the exponential and in front of it, the two inequalities (5) and (8) are essentially equivalent.

In the applications to integrability of norms of Gaussian random vectors (in the last section), it is sometimes more convenient to use (5) since it does not require to check a priori that the expectation is finite.

It is also worthwhile mentioning that Theorem 1, or (7), is good enough to get close to the isoperimetric statement. For example, applied to $F(x) = \min(d(x, A), r)$, $x \in \mathbb{R}^n$, $r > 0$, for some Borel set A with $\gamma_n(A) \geq \alpha$, $0 < \alpha < 1$, for which $\int_{\mathbb{R}^n} F\,d\gamma_n \leq (1-\gamma_n(A))r \leq (1-\alpha)r$, (7) yields

$$\gamma_n\big(A_{\alpha r}\big) \;\geq\; 1 - e^{-\frac{1}{2}r^2}.$$

With some more effort, it is possible to show that

$$\gamma_n\big(A_r\big) \;\geq\; 1 - e^{-\frac{1}{2}r^2 + \delta(\gamma_n(A))r} \tag{9}$$

for every $r \geq 0$, where $\delta(\gamma_n(A)) \to 0$ as $\gamma_n(A) \to 1$ (cf. [6]).

4 Functional analytic proof

A weaker form of Theorem 1 is presented in [2] (the Maurey-Pisier inequality). In the post [5], it is shown how Theorem 1 may be deduced from the logarithmic Sobolev inequality via the Herbst argument. A simple proof of the logarithmic Sobolev inequality itself is provided by heat (Mehler kernel) flow arguments. It is actually instructive to give a direct heat flow proof of Theorem 1.

Let

$$h_t(x) \;=\; \frac{1}{(4\pi t)^{\frac{n}{2}}}\,e^{-\frac{1}{4t}|x|^2}, \quad t > 0,\; x \in \mathbb{R}^n,$$

be the standard heat kernel, fundamental solution of the heat equation $\partial_t h_t = \Delta h_t$. The convolution semigroup $H_t f(x) = f * h_t(x)$, $t > 0$, solves $\partial_t H_t f = \Delta H_t f = H_t \Delta f$ with initial data f. At $t = \frac{1}{2}$, h_t is just the standard Gaussian density so that $H_{\frac{1}{2}} f(0) = \int_{\mathbb{R}^n} f d\gamma_n$ (while $H_0 f = f$).

Let $F : \mathbb{R}^n \to \mathbb{R}$ be bounded, 1-Lipschitz, and smooth so that $|\nabla F| \leq 1$ everywhere. For $\lambda \in \mathbb{R}$, and $t > 0$ fixed, consider the function $J(s) = H_s(e^{\lambda H_{t-s} F})$, $s \in [0, t]$ (at any fixed point x in $\mathbb{R}^n$, omitted in the notation). By the heat equation and the chain rule,

$$\begin{aligned} J'(s) &= \Delta H_s\big(e^{\lambda H_{t-s}F}\big) - H_s\big(\lambda\, e^{\lambda H_{t-s}F}\Delta H_{t-s}f\big) \\ &= H_s\Big(\Delta\, e^{\lambda H_{t-s}F} - \lambda\, e^{\lambda H_{t-s}F}\Delta H_{t-s}f\Big) \\ &= \lambda^2 H_s\Big(e^{\lambda H_{t-s}F}\,|\nabla H_{t-s}F|^2\Big). \end{aligned}$$

Now $|\nabla H_{t-s}F| = |H_{t-s}(\nabla F)| \leq H_{t-s}(|\nabla F|) \leq 1$, so that

$$J'(s) \,\leq\, \lambda^2 J(s), \quad s \in [0,t].$$

Integration of this differential inequality yields

$$\log\Big(\frac{J(t)}{J(0)}\Big) \,=\, \log J(t) - \log J(0) \,\leq\, \lambda^2 t,$$

that is

$$H_t(e^{\lambda F}) \,\leq\, e^{\lambda H_t F + \lambda^2 t}.$$

At $t = \frac{1}{2}$ (and at the point $x = 0$ for example), this is the announced claim. If F is an arbitrary 1-Lipschitz function, apply the preceding to $\min(N, \max(H_\varepsilon F, -N))$, $N \geq 1$, $\varepsilon > 0$, and let $N \to \infty$, $\varepsilon \to 0$.

The preceding argument is, in spirit, not very far from the Maurey-Pisier inequality presented in [2]. It may indeed be developed completely similarly with the Mehler kernel and Ornstein-Uhlenbeck semigroup (cf. [4, 5]), which is represented as

$$P_t f(x) \,=\, \int_{\mathbb{R}^n} f\big(\sin(\theta)x + \cos(\theta)y\big)d\gamma_n(y)$$

with $e^{-t} = \sin(\theta)$.

5 Concentration of Gaussian vectors

The concentration inequalities for the canonical Gaussian measure γ_n presented in the previous sections extend to arbitrary Gaussian measures. It is actually convenient to present them for random vectors.

Let thus X be a centered Gaussian vector on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^n$, with covariance matrix $\Sigma = M^{\top} M$. If Y is a random vector with law γ_n, then MY has the law of X. Therefore, whenever $F : \mathbb{R}^n \to \mathbb{R}$, (7) for example yields

$$\mathbb{P}\big(F(X) \geq \mathbb{E}(F(X)) + r\big) \,\leq\, e^{-r^2/2\sigma_F^2}, \quad r \geq 0,$$

where $\sigma_F = \|F \circ M\|_{\mathrm{Lip}}$ (provided it is finite). It may be observed that, in general,

$$\|F \circ M\|_{\mathrm{Lip}} \leq \sup_{|c|\leq 1} \sqrt{\langle \Sigma c, c\rangle}\, \|F\|_{\mathrm{Lip}},$$

but depending on the nature of F, sharper bounds might be available.

For instance, if $F(x) = \max_{1\leq k\leq n} x_k$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\|F \circ M\|_{\mathrm{Lip}} \leq \max_{1\leq k\leq n} \Sigma_{kk} = \max_{1\leq k\leq n} \mathbb{E}(X_k^2) = \sigma^2.$$

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

$$\big|(Mx)_k - (My)_k\big|^2 = \bigg|\sum_{\ell=1}^n M_{k\ell}(x_\ell - y_\ell)\bigg|^2 \leq \sum_{\ell=1}^n M_{k\ell}^2 \, |x-y|^2 = \Sigma_{kk}\, |x-y|^2.$$

As a consequence, for any centered Gaussian vector $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$ and any $r \geq 0$,

$$\mathbb{P}\Big(\max_{1\leq k\leq n} X_k \geq \mathbb{E}\big(\max_{1\leq k\leq n} X_k\big) + t\Big) \leq e^{-r^2/2\sigma^2}. \tag{10}$$

It holds as well, for any $r \geq 0$,

$$\mathbb{P}\Big(\Big|\max_{1\leq k\leq n} X_k - \mathbb{E}\big(\max_{1\leq k\leq n} X_k\big)\Big| \geq r\Big) \leq 2\, e^{-r^2/2\sigma^2}, \tag{11}$$

and similarly with $\max_{1\leq k\leq n} |X_k|$ instead of $\max_{1\leq k\leq n} X_k$. These properties are used in the study of the integrability of norms of Gaussian vectors and processes (next section and [2]). Again, it is important to realize the relative sizes of $\mathbb{E}\big(\max_{1\leq k\leq n} X_k\big)$ and $\sigma^2 = \max_{1\leq k\leq n} \mathbb{E}(X_k^2)$ in these concentration inequalities. For example, for a sample of independent standard normal variables $X_1, \ldots, X_n$, the first quantity is of order of $\sqrt{\log n}$ (cf. [4]) while $\sigma^2 = 1$.

6 Integrability of norms of Gaussian vectors

This section briefly resumes the conclusions of the note [2] on the basis of the concentration inequalities emphasized in the previous sections.

Theorem 2 (Concentration and integrability of norms of Gaussian vectors)**.** *Let X be a centered Gaussian random vector on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a real separable Banach space $(E, \|\cdot\|)$. Then $\mathbb{E}(\|X\|) < \infty$, and*

$$\mathbb{P}\big(\big|\|X\| - m\big| \geq r\big) \leq 2\, e^{-\frac{r^2}{2\sigma^2}}, \quad r \geq 0.$$

where m is either a median of $\|X\|$ or $\mathbb{E}(\|X\|)$, and

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

As a consequence, $\mathbb{E}(e^{\alpha\|X\|^2}) < \infty$ if and only if $\alpha < \frac{1}{2\sigma^2}$.

References

[1] The Gaussian isoperimetric inequality. *The Gaussian Blog.*

[2] Integrability of norms of Gaussian random vectors and processes. *The Gaussian Blog.*

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

[4] Some basic properties and characterizations of Gaussian measures and variables. *The Gaussian Blog.*

[5] Logarithmic Sobolev and transportation inequalities. *The Gaussian Blog.*

[6] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs 89. American Mathematical Society (2001).