The Gaussian Poincaré inequality

Let $X = (X_1, \ldots, X_n)$ be centered Gaussian vector with values in $\mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. It holds true that

$$\mathrm{Var}\big(\max_{1 \leq k \leq n} X_k\big) \,\leq\, \max_{1 \leq k \leq n} \mathrm{Var}(X_k). \tag{1}$$

This easy to remember inequality is both very general and useful (and, at this level of generality, optimal). While it may be addressed via the Gaussian isoperimetric inequality [1], it may be deduced in a straightforward manner from a functional inequality known as the Gaussian Poincaré inequality, expressing that for every smooth function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Var}_{\gamma_n}(f) \,\leq\, \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n. \tag{2}$$

Here γ_n stands for the standard Gaussian probability measure on the Borel sets of $\mathbb{R}^n$ 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, and, for a function $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathrm{L}^2(\gamma_n)$,

$$\mathrm{Var}_{\gamma_n}(f) \,=\, \int_{\mathbb{R}^n} f^2 d\gamma_n - \bigg(\int_{\mathbb{R}^n} f \, d\gamma_n\bigg)^2$$

is the variance of f with respect to γ_n. The gradient ∇f of f is the vector of partial derivatives $(\partial_k f)_{1 \leq k \leq n}$, and $|\nabla f|^2 = \sum_{k=1}^n (\partial_k f)^2$.

To deduce the variance inequality (1) from the Poincaré inequality (2), apply the latter to a smooth approximation of the function $f : x \mapsto \max_{1 \leq k \leq n} (Mx)_k$ where $\Sigma = M^{\top} M$ is the covariance matrix of the law of X. Consider for example, for any $\beta > 0$,

$$f_\beta(x) \,=\, \frac{1}{\beta} \log\bigg(\sum_{k=1}^n e^{\beta (Mx)_k}\bigg), \quad x \in \mathbb{R}^n.$$

Drafted by M. L. **v1** February 2024

Then, for every $k = 1, \dots, n$,

$$\partial_k f_\beta(x) \,=\, \frac{\sum_{\ell=1}^n M_{k\ell}\, e^{\beta(Mx)_\ell}}{\sum_{\ell=1}^n e^{\beta(Mx)_\ell}} \,=\, \sum_{\ell=1}^n p_\ell(x) M_{k\ell}$$

where $p_\ell(x) = e^{\beta(Mx)_\ell} / \sum_{j=1}^n e^{\beta(Mx)_j}$, $\ell = 1, \dots, n$. By Jensen's inequality with respect to the probability weights $p_\ell(x)$, $\ell = 1, \dots, n$, at every $x \in \mathbb{R}^n$,

$$\sum_{k=1}^n \big(\partial_k f_\beta(x)\big)^2 \,\leq\, \sum_{\ell=1}^n p_\ell(x) \sum_{k=1}^n M_{k\ell}^2 \,=\, \sum_{\ell=1}^n p_\ell(x)\, \Sigma_{\ell\ell} \,\leq\, \max_{1\leq \ell\leq n} \Sigma_{\ell\ell}.$$

Therefore, as an application of the Poincaré inequality (2) to f_β, it follows that

$$\mathrm{Var}_{\gamma_n}(f_\beta) \,\leq\, \max_{1\leq k\leq n} \mathrm{Var}(X_k).$$

It remains to observe that $\lim_{\beta\to\infty} f_\beta(x) = \max_{1\leq k\leq n}(Mx)_k$, $x \in \mathbb{R}^n$, and that $x \mapsto Mx$ under γ_n is distributed as X.

This note is devoted to a brief discussion of the Poincaré inequality (2), including two independent proofs (among numerous alternate proofs available in the literature). Stronger inequalities are discussed in the companion note [2]. While the exposition is limited to the Poincaré inequality for finite-dimensional Gaussian distributions, it may be extended to arbitrary (infinite-dimensional) Gaussian distributions (cf. [6]).

Within the family of so-called "Poincaré inequalities", the one considered here is more of the form of the Wirtinger inequality (on the sphere – see in particular the historical note of J. Mawhin [9]). The Gaussian Poincaré inequality was probably known already in the years 1930 in the physics literature, as part of some folklore along expansions in Fourier-Hermite (harmonic oscillator) polynomials. It is later stated explicitly in a pde context in [10] and in statistics [8]. It then developed in a number of areas of studies and applications. Some account may be found in [6, 5]...

Table of contents

1 The Gaussian Poincaré inequality

The Poincaré inequality for the standard Gaussian measure γ_n emphasized in the introduction is summarized in the following statement.

Theorem 1 (The Gaussian Poincaré inequality)**.** *For any locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ in $\mathrm{L}^2(\gamma_n)$,*

$$\mathrm{Var}_{\gamma_n}(f) \le \int_{\mathbb{R}^n} |\nabla f|^2 d\gamma_n. \tag{3}$$

It is not difficult to verify that linear functions are extremals of the inequality.

If X is an arbitrary Gaussian random vector with values in $\mathbb{R}^n$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with covariance matrix Σ, the Poincaré inequality takes the form

$$\mathrm{Var}\big(f(X)\big) \le \mathbb{E}\big(\langle \Sigma \, \nabla f, \nabla f \rangle\big) \tag{4}$$

for every smooth function $f : \mathbb{R}^n \to \mathbb{R}$. This follows, as in the introduction, from the fact that X has the same distribution as $\mathbb{E}(X) + MG$ where $\Sigma = M^{\top} M$ and G is standard normal.

2 Tensorization

An important and useful property of the Poincaré inequality (3) is the tensorization property. That is, due to the product property of γ_n, it is enough to establish (3) in dimension one. For a sketch of the argument for $n = 2$,

$$\begin{aligned}
\int_{\mathbb{R}^2} f(x_1, x_2)^2 d\gamma_2(x_1, x_2) &= \int_{\mathbb{R}} \bigg(\int_{\mathbb{R}} f(x_1, x_2)^2 d\gamma_1(x_2) \bigg) d\gamma_1(x_1) \\
&\le \int_{\mathbb{R}} \bigg(\int_{\mathbb{R}} f(x_1, x_2) d\gamma_1(x_2) \bigg)^2 d\gamma_1(x_1) \\
&\quad + \int_{\mathbb{R}} \bigg(\int_{\mathbb{R}} \big[\partial_2 f(x_1, x_2)\big]^2 d\gamma_1(x_2) \bigg) d\gamma_1(x_1)
\end{aligned}$$

where (3) has been used along the x_2 coordinate. Apply next (3) to $h(x_1) = \int_{\mathbb{R}} f(x_1, x_2) d\gamma_1(x_2)$, $x_1 \in \mathbb{R}$, for which

$$h'(x_1)^2 = \bigg(\int_{\mathbb{R}} \partial_1 f(x_1, x_2) d\gamma_1(x_2) \bigg)^2 \le \int_{\mathbb{R}} \big[\partial_1 f(x_1, x_2)\big]^2 d\gamma_1(x_2)$$

to conclude that

$$\begin{aligned}\int_{\mathbb{R}^2} f(x_1,x_2)^2 d\gamma_2(x_1,x_2) \;\leq\; & \left(\int_{\mathbb{R}}\int_{\mathbb{R}} f(x_1,x_2)d\gamma_1(x_2)d\gamma_1(x_1)\right)^2 \\ & + \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \big[\partial_2 f(x_1,x_2)\big]^2 d\gamma_1(x_2)\right)d\gamma_1(x_1) \\ & + \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \big[\partial_1 f(x_1,x_2)\big]^2 d\gamma_1(x_2)\right)d\gamma_1(x_1)\end{aligned}$$

which amounts to the Poincaré inequality (3) in dimension 2.

This tensorization argument may actually be formalized via the inequality (in arbitrary dimension)

$$\mathrm{Var}_{\gamma_n}(f) \;\leq\; \sum_{k=1}^{n}\int_{\mathbb{R}^{n-1}} \mathrm{Var}_{d\gamma_1(x_k)}(f_k)\, d\gamma_{n-1}(x_1,\dots,x_{k-1},x_{k+1},\dots,x_n)$$

(valid for any product measure) where $f_k(x_k) = f(x_1,\dots,x_n)$ with $x_1,\dots,x_{k-1},x_{k+1},\dots,x_n$ fixed.

3 Hermite expansion proof

The Poincaré inequality of Theorem 1 may be given a simple proof by a series expansion in Hermite polynomials. Although this is not strictly necessary, the argument is easily developed in dimension one (the general case following by tensorization).

Start therefore with the expansion of a function $f \in \mathrm{L}^2(\gamma_1)$ in Hermite polynomials

$$f - \int_{\mathbb{R}} f\, d\gamma_1 \;=\; \sum_{k=1}^{\infty} f_k h_k$$

where f_k are real coefficients, and h_k, $k \geq 0$, is the sequence of Hermite polynomials (normalized in $\mathrm{L}^2(\gamma_1)$), see [3]. If necessary, start with a finite sum. Since $h_k' = \sqrt{k}\, h_{k-1}$, $k \geq 1$,

$$f' \;=\; \sum_{k=1}^{\infty} f_k \sqrt{k}\, h_{k-1}.$$

Since the Hermite polynomials form an orthonormal basis of $\mathrm{L}^2(\gamma_1)$,

$$\int_{\mathbb{R}} \big[f - \textstyle\int_{\mathbb{R}} f d\gamma_1\big]^2 d\gamma_1 \;=\; \displaystyle\sum_{k=1}^{\infty} f_k^2$$

while

$$\int_{\mathbb{R}} f'^2 d\gamma_1 \;=\; \sum_{k=1}^{\infty} k\, f_k^2$$

from which the inequality $\int_{\mathbb{R}} [f - \int_{\mathbb{R}} f d\gamma_1]^2 d\gamma_1 \leq \int_{\mathbb{R}} f'^2 d\gamma_1$ is immediate. A density argument then completes the proof.

4 Heat flow proof

An alternate approach to the Poincaré inequality goes via the heat semigroup $(H_t)_{t\geq 0}$. Recall

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

the standard heat kernel, fundamental solution of the heat equation $\partial_t h_t = \Delta p_t$. 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 $H_{\frac{1}{2}} f(0) = \int_{\mathbb{R}^n} f d\gamma$ (while $H_0 f = f$).

Fix $t > 0$ (later taken to be $\frac{1}{2}$), and for a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, write, at any point (omitted in the notation),

$$H_t(f^2) - (H_t f)^2 \;=\; \int_0^t \frac{d}{ds} H_s\big((H_{t-s} f)^2\big) ds.$$

By the chain rule and the heat equation,

$$\begin{aligned}
\frac{d}{ds} H_s\big((H_{t-s} f)^2\big) &= \Delta H_s\big((H_{t-s} f)^2\big) - 2H_s\big(H_{t-s} f\, \Delta H_{t-s} f\big) \\
&= H_s\big(\Delta\big((H_{t-s} f)^2\big)\big) - 2H_s\big(H_{t-s} f\, \Delta H_{t-s} f\big).
\end{aligned}$$

Now, for any smooth function u, $\Delta(u^2) - 2u\Delta u = 2|\nabla u|^2$ so that

$$\frac{d}{ds} H_s\big((H_{t-s} f)^2\big) \;=\; 2H_s\big(|\nabla H_{t-s} f|^2\big).$$

Next, as is immediately verified on the convolution definition of H_t,

$$|\nabla H_{t-s} f|^2 \;=\; \big|H_{t-s}(\nabla f)\big|^2 \;\leq\; H_{t-s}\big(|\nabla f|^2\big)$$

so that

$$\begin{aligned}
H_t(f^2) - \big(H_t f\big)^2 &= 2\int_0^t H_s\big(|\nabla H_{t-s} f|^2\big) ds \\
&\leq 2\int_0^t H_s\big(H_{t-s}(|\nabla f|^2)\big) ds \\
&= 2t H_t\big(|\nabla f|^2\big)
\end{aligned}$$

where the last step follows from the semigroup property. As announced at the beginning of the argument, at $t = \frac{1}{2}$ (and at the point $x = 0$ for example), the latter inequality exactly amounts to the Poincaré inequality.

A similar, even somewhat easier, proof may be produced via the Ornstein-Uhlenbeck semigroup [4].

5 Characterization

It is shown in [7] that Gaussian measures are the only probability measures μ on the real line satisfying the Poincaré inequality

$$\mathrm{Var}_\mu(f) \,\leq\, \sigma^2 \int_{\mathbb{R}} f'^2 d\mu$$

for every smooth function $f : \mathbb{R} \to \mathbb{R}$ in $\mathrm{L}^2(\mu)$, where σ^2 is the variance of μ.

By homogeneity, it is enough to consider the case when $\sigma = 1$. Denote by ν the measure μ centered at its mean, which also satisfies the Poincaré inequality with constant 1. The task is to show that ν is the standard normal distribution. This is achieved as soon as

$$\int_{\mathbb{R}} h_k \, d\nu \,=\, 0 \tag{5}$$

along the sequence of Hermite polynomials h_k, $k \geq 1$. There is the prior that polynomials are integrable with respect to μ, which is easily verified applying the Poincaré inequality successively to $f(x) = x^2$, x^4 etc.

By the normalization hypothesis, (5) holds true for $k = 2$ since $h_2(x) = \frac{1}{\sqrt{2}}(x^2 - 1)$, $x \in \mathbb{R}$ (and also for $k = 1$ by centering). Assume by induction that it is true up to $k \geq 3$, and prove that it is satisfied for $k+1$. By the three-term recurrence relation satisfied by Hermite polynomials (cf. [3])

$$x \, h_k \,=\, \sqrt{k+1} \, h_{k+1} + \sqrt{k} \, h_{k-1},$$

it is equivalent to show that $\int_{\mathbb{R}} x h_k d\nu = 0$. Now, for every real a, if $f(x) = ax + h_k(x)$, $x \in \mathbb{R}$,

$$\mathrm{Var}_\nu(f) \,=\, a^2 + 2a \int_{\mathbb{R}} x h_k \, d\nu + \mathrm{Var}_\nu(h_k)$$

by the normalization of the variance. On the other hand, since $f'(x) = a + h_k'(x) = a + \sqrt{k} \, h_{k-1}(x)$,

$$\begin{aligned} \int_{\mathbb{R}} f'^2 d\nu \,&=\, a^2 + 2a\sqrt{k} \int_{\mathbb{R}} h_{k-1} d\nu + k \int_{\mathbb{R}} h_{k-1}^2 d\nu \\ &=\, a^2 + k \int_{\mathbb{R}} h_{k-1}^2 d\nu \end{aligned}$$

by the induction hypothesis. The Poincaré inequality for ν thus expresses that

$$2a \int_{\mathbb{R}} x h_k \, d\nu + \mathrm{Var}_\nu(h_k) \;\leq\; k \int_{\mathbb{R}} h_{k-1}^2 d\nu.$$

Since $a \in \mathbb{R}$ is arbitrary, it must be that $\int_{\mathbb{R}} x h_d \nu = 0$, which is the claim.

References

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

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

[3] Hermite polynomials. *The Gaussian Blog.*

[4] Mehler kernel, and the Ornstein-Uhlenbeck operator. *The Gaussian Blog.*

[5] D. Bakry, I. Gentil, M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften 348. Springer (2014).

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

[7] A. Borovkov, S. Utev. An inequality and a characterization of the normal distribution connected with it. *Teor. Veroyatnost. i Primenen.* 28, 209–218 (1983).

[8] H. Chernoff. A note on an inequality involving the normal distribution. *Ann. Probab.* 9, 533-535 (1981).

[9] J. Mawhin. Henri Poincaré and the partial differential equations of mathematical physics. *Hist. Math.* 36, 257–277. American Mathematical Society (2010).

[10] J. Nash. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* 80, 931–954 (1958).