5. Some consequences of the Gaussian isoperimetric inequality

Equivalent formulations: We present two equivalent ways of writing the Gaussian isoperimetric inequality. The first one, without explicit reference to half-spaces is

$$\Phi^{-1}(\gamma_m(A^{\varepsilon})) \geq \Phi^{-1}(\gamma_m(A)) + \varepsilon \quad \text{for all Borel sets } A \text{ and any } \varepsilon > 0. \tag{2}$$

Exercise 10. Deduce (2) and Theorem 1 from each other.

Here is yet another formulation[5].

Proposition 11. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lip(κ) function. Then there exists a Lip(κ) function $g : \mathbb{R} \to \mathbb{R}$ such that $\gamma_n \circ f^{-1} = \gamma_1 \circ g^{-1}$. In other words, the distribution of the random variable f on the probability space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \gamma_n)$ is the same as the distribution of the random variable g on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \gamma_1)$.*

Exercise 12. Deduce Proposition 11 and Theorem 1 from each other.

Proposition 11 shows the dimension-free nature of isoperimetric inequality. In other words, the isoperimetric inequality will hold for standard Gaussian measures in infinite dimensions, once we make sense of such an object! This would not have been the case if Proposition 11 only asserted that g is Lip($\kappa \log n$), for example.

Log-concave densities: What made the proof of isoperimetric inequality in one dimension click? Looking back, we see that the key point was that $\varphi(u+\varepsilon)/\varphi(u)$ is decreasing in u, for any fixed $\varepsilon > 0$. Any other density f satisfying this will also satisfy the isoperimetric inequality (perhaps we need symmetry?). This condition is equivalent to $\log f(u+\varepsilon) - \log f(u)$ being decreasing in u. Assuming smoothness for simplicity, this happens if and only if $(\log f)'(u)$ is decreasing in u, which in turn is equivalent to $(\log f)''(u)$ being negative. In other words, equivalent to $\log f$ being a concave function.

Any density (in any dimension) for which $\log f$ is concave, is called *log-concave*. Examples in one dimension are symmetric exponential density $\frac{1}{2}e^{-|x|}$, uniform density on an interval, and of course the Gaussian. Examples in higher dimensions are uniform measures on compact convex sets and the densities $\exp\{-|\mathbf{x}|^p\}$ for $p \geq 1$. One can get many more from these few, since log-concave densities are closed under convolutions and under linear transformations (eg., marginals). Log-concave densities are a very important class of densities that share many properties of Gaussian measures, in particular, concentration properties.

Gaussian Brunn-Minkowski inequality: In Euclidean space, we deduced the isoperimetric inequality from the Brunn-Minkowski inequality. Is there an analogue for the Gaussian measure? Ehrhard initiated this study and proved the inequality below for convex sets, again using his symmetrization procedure. The convexity assumption was relaxed by Latala and completely removed by Borell.

Result 13 (Ehrhard, Latala, Borell)**.** If $A, B \subseteq \mathbb{R}^n$ (Borel sets), and $\alpha \in [0,1]$, then $\Phi^{-1}(\gamma_n(\alpha A + (1-\alpha)B)) \geq \alpha\Phi^{-1}(\gamma_n(A)) + (1-\alpha)\Phi^{-1}(\gamma_n(B))$.

We shall not use this and hence not give a proof[6].

Concentration inequalities: The isoperimetric inequality implies concentration inequalities for various functions of Gaussian random variables. This is its primary importance in probability. It is possible to deduce some of these concentration bounds, albeit with poorer constants, but the isoperimetric inequality yields the sharpest general bounds.

[5] Taken from Boris Tsirelson's lecture notes available on his home page.

[6] Potential presentation topic! See Borell's paper *The Ehrhard inequality*. Another potential topic is a very different proof of the Gaussian isoperimteric inequality by Bobkov, see *An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space.*

Theorem 14. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lip(κ) function. Let M_f be a median of f, defined by $\gamma_n\{f \geq M_f\} \geq \frac{1}{2}$ and $\gamma_n\{f \leq M_f\} \geq \frac{1}{2}$. Then, for every $t > 0$, we have*

$$\gamma_n\left\{f - M_f \geq t\right\} \leq \overline{\Phi}\left(\frac{t}{\kappa}\right) \leq e^{-\frac{t^2}{2\kappa^2}}, \tag{3}$$

$$\gamma_n\left\{|f - M_f| \geq t\right\} \leq 2\overline{\Phi}\left(\frac{t}{\kappa}\right) \leq 2e^{-\frac{t^2}{2\kappa^2}}. \tag{4}$$

Proof. If $A = \{f \leq M_f\}$ then $A^t \subseteq \{f \leq M_f + \kappa t\}$. But $\Phi^{-1}(\gamma_n(A)) \geq 0$ and hence by (2) we get $\Phi^{-1}(\gamma_n(A^t)) \geq t$. Hence $\gamma_n\{f \geq M_f + \kappa t\} \leq \overline{\Phi}(t)$ which shows the first claim. The second follows by adding the same estimate for $\gamma_n\{f \leq M_f - t\}$. ■

Remark 15. Since $\overline{\Phi}(t)$ is strictly smaller than $\frac{1}{2}$ for every $t > 0$, it follows that the median is unique! Incidentally, we have been writing statements in terms of measures, but one can equivalently state them in terms of random variables. If $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$ random variables on some probability space, and $V = f(X_1, \ldots, X_n)$ for a Lip(κ) function f, then

$$\mathbf{P}\{|V - \text{Med}[V]| \geq t\} \leq 2e^{-t^2/2\kappa^2}.$$

The random variable is concentrated around its median. Incidentally, inequalities of this type, with perhaps not the optimal constants on the right, can be obtained by easier methods (see the end of this section). Often that suffices in applications but we decided to go through the isoperimetric inequality for its natural appeal, in addition to sharpness of constants.

Example 16. Some examples of Lipschitz functions of interest are $\max_i x_i$, $\|\mathbf{x}\|_p$ (or any norm, for that matter), $d(\mathbf{x}, A)$ for a fixed closed set A. A smooth function is Lipschitz if and only if its gradient is bounded.

What about functions of correlated Gaussians? Here is a simple exercise.

Exercise 17. Suppose $X \sim N_n(\mu, \Sigma)$ and let $f : \mathbb{R}^n \to \mathbb{R}$ is a Lip(κ) function. Let $V = f(X)$. Then $\mathbf{P}\{|V - \text{Med}[V]| \geq t\} \leq 2e^{-t^2/2\lambda_1\kappa^2}$ with λ_1 being the maximal eigenvalue of Σ.

Concentration inequalities of the type given by Theorem 14 are desirable to have for many other probability measures too. Deduce the following from Theorem 14.

Exercise 18. Let V_n be the uniform probability measure on $[0,1]^n$. If $f : [0,1]^n \to \mathbb{R}$ is Lip(κ), show that

$$\gamma_n\left\{f - M_f \geq t\right\} \leq e^{-ct^2/\kappa^2},$$
$$\gamma_n\left\{|f - M_f| \geq t\right\} \leq 2e^{-ct^2/\kappa^2}.$$

Here c is a numerical constant (find it!).

For general product measures, for example uniform measure on the discrete cube $\{0,1\}^n$, getting a similar concentration inequality is hard. This is the famous *Talagrand's inequality*, proved by Talagrand and now a cornerstone in probability.

Concentration about the mean: Usually mean is easier to compute than median and concentration inequalities are often expressed around the mean. Here is a simple way to get a (sub-optimal) concentration inequality around the mean for the same setting as above. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lip(κ) function and let M_f be its median under γ_n and let $E_f = \int f(\mathbf{x}) d\gamma_n(\mathbf{x})$ be its expectation.

Using the bound in Theorem 14 we get

$$\mathbf{E}[(f - M_f)_+] = \int_0^\infty \gamma_n\{f > M_f + t\} dt \leq \int_0^\infty \overline{\Phi}(t/\kappa) dt = \frac{\kappa}{\sqrt{2\pi}}.$$

The same bound holds for $\mathbf{E}[(f-M_f)_-]$ and we get $\mathbf{E}[|f-M_f|] \le \sqrt{\frac{2}{\pi}}\kappa < \kappa$. In particular, $|E_f - M_f| < \kappa$. Therefore, for $t \ge 2$, we get

$$\gamma_n\{f-E_f > t\kappa\} \le \gamma_n\left\{f - M_f > \frac{t}{2}\kappa\right\} \le \overline{\Phi}(t/2),$$

by another application of Theorem 14. For $t \le 2$, we use the trivial bound $\gamma_n\{f-E_f > t\kappa\} \le 1$. Putting all this together and using the same for deviations below E_f we arrive at the following result.

Theorem 19. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be a Lip(κ) function. Let $E_f = \int f d\gamma_n$. Then, for every $t>0$, we have (with $C = 1/\overline{\Phi}(1)$)*

$$\gamma_n\left\{f - E_f \ge t\right\} \le C\overline{\Phi}\left(\frac{t}{2\kappa}\right) \le Ce^{-\frac{t^2}{8\kappa^2}}, \tag{5}$$

$$\gamma_n\left\{|f - E_f| \ge t\right\} \le C\overline{\Phi}\left(\frac{t}{2\kappa}\right) \le Ce^{-\frac{t^2}{8\kappa^2}}. \tag{6}$$

Weaker forms of concentration by easier methods: As we remarked earlier, weaker forms of concentration inequalities can be obtained by easier methods some of which we mention here[7].

Let $f:\mathbb{R}^n \to \mathbb{R}$ be a Lip(1) function and let $X \sim \gamma_n$. We look for a number A_f such that $f(X)$ is well-concentrated about A_f. The crudest bound is as follows. Let Y be an independent copy of X on the same probability space, and use $\mathbf{E}[|f(X)-f(Y)|] \le \mathbf{E}[\|X-Y\|] \asymp \sqrt{n}$. Observing that $\min_a \mathbf{E}[|f(X)-a|] \le \mathbf{E}[|f(X)-f(Y)|]$, we get a number A_f such that $\mathbf{E}[|f(X)-A_f|] \lesssim \sqrt{n}$. By Markov's inequality this gives weak bounds like $\mathbf{P}\{|f(X)-A_f| \ge t\} \lesssim \frac{\sqrt{n}}{t}$. This compares poorly with the bound in (14).

To improve this, we introduce a technique that will be used many times later. Interpolate between X and Y by setting $Z(\theta) = (\cos\theta)X + (\sin\theta)Y$ for $0 \le \theta \le \frac{\pi}{2}$ so that $Z(0) = X$ and $Z(\pi/2) = Y$. The key property of this interpolation is that for any θ, the random vectors $Z_\theta = (\cos\theta)X + (\sin\theta)Y$ and $\dot{Z}_\theta = -(\sin\theta)X + (\cos\theta)Y$ are independent and have γ_n distribution.

Now assume that f is smooth, then the Lipschitz condition is equivalent to $|\nabla f(\mathbf{x})| \le \kappa$ for all $\mathbf{x} \in \mathbb{R}^n$. It is easy to approximate Lipschitz functions uniformly by smooth Lipschitz functions and thus extend the bounds obtained below to all Lipschitz functions, a step we shall not elaborate on. Then, write $f(X)-f(Y)$ as the integral of $\frac{d}{d\theta}f(Z_\theta) = \langle \nabla f(Z_\theta), \dot{Z}_\theta\rangle$ to get

$$\begin{aligned}\mathbf{E}[|f(X)-f(Y)|] &\le \int_0^{\pi/2} \mathbf{E}[\langle \nabla f(Z_\theta), \dot{Z}_\theta\rangle] d\theta \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\pi/2} \mathbf{E}[\|\nabla f(Z_\theta)\|] d\theta \\ &\le \sqrt{\frac{\pi}{2}}.\end{aligned}$$

From this we get some number A_f such that $\mathbf{P}\{|f(X)-A_f| \ge t\} \lesssim \frac{1}{t}$. This does not decay fast in t, but is free of n, already a remarkable improvement over the crude bound.

By bounding $\mathbf{E}[G(|f(X)-f(Y)|)]$ for some convex increasing function G we can get better bounds along the same lines.

Exercise 20. For $b>0$ and $x \in \mathbb{R}^n$ define $G_b(x) = (|x|-b)_+$. Use the convexity of G to obtain the bound $\mathbf{E}[|G_b(X) - G_b(Y)| \ge t] \le \mathbf{E}[G(\frac{\pi}{2}X)]$.

What concentration of $f(X)$ does this yield?

[7] For a spectacular presentation leading up from simpler inequalities up to the Borell-TIS inequality, see the lecture notes of Boris Tsirelson `http://www.tau.ac.il/~tsirel/Courses/Gaussian/lect2.pdf`. Here I have taken a couple of points from those notes.