2 Concentration of Gaussian measure

It is of course impossible to even think *the word Gaussian without immediately mentioning the most important property of Gaussian processes, that is concentration of measure.*

M. Talagrand[1]

2a Why be Lipschitz?

2a1 Definition. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a C-*Lipschitz* function (symbolically, $f \in \operatorname{Lip}(C, \mathbb{R}^n \to \mathbb{R})$ or just $f \in \operatorname{Lip}(C)$), if

$$|f(x) - f(y)| \le C|x - y| \quad \text{for all } x, y \in \mathbb{R}^n ;$$

here $|x - y|$ is the Euclidean distance, and $C \in [0, \infty)$.

2a2 Exercise. (a) The function $(x_1, \dots, x_n) \mapsto \|x\|_\infty = \max(|x_1|, \dots, |x_n|)$ is 1-Lipschitz;

(b) the function $(x_1, \dots, x_n) \mapsto \sqrt{x_1^2 + \dots + x_n^2}$ is 1-Lipschitz;

(c) every norm (or seminorm) $\|\cdot\|$ on $\mathbb{R}^n$ is C-Lipschitz where $C = \max\{\|x\| : |x| \le 1\}$;

(d) the Lipschitz constants in (a), (b), (c) are exact.

Prove it.

Hint: every norm is the supremum of some linear functions.

[1] See page 189 of "Mean field model for spin glasses: a first course", Lecture Notes in Math. **1816** (2003), 181–285.

2a3 Exercise. The function $f(x_1,\dots,x_n) = \ln(\mathrm{e}^{x_1}+\dots+\mathrm{e}^{x_n})$ is 1-Lipschitz.
Prove it.
Hint: f is increasing, and $f(x_1+a,\dots,x_n+a) = f(x_1,\dots,x_n)+a$.

2a4 Exercise. For $x_1,\dots,x_{2n-1} \in \mathbb{R}$ define $\mathrm{Me}(x_1,\dots,x_{2n-1})$ to be the (evidently existing and unique) number $x_{(n)}$ such that

$$\#\{k : x_k \le x_{(n)}\} \ge n \quad \text{and} \quad \#\{k : x_k \ge x_{(n)}\} \ge n\,;$$

prove that $\mathrm{Me} \in \mathrm{Lip}(1)$.
Hint: similar to 2a3.

In place of $\mathbb{R}^n$ we may use $\mathbb{R}^E$ for a finite set E. Especially, let E be the set of edges of a connected graph $G = (V,E)$. Given $x \in \mathbb{R}^E$, every $a \in \mathbb{R}$ leads to a subgraph $G_a = (V,E_a)$, where $E_a = \{e \in E : x(e) \ge a\}$. Given also two (different) vertices $v_1, v_2 \in V$, we may consider the highest a such that v_1, v_2 are connected in G_a.[1] Define $f : \mathbb{R}^E \to \mathbb{R}$ by $f(x) = a$ (a being as above).

2a5 Exercise. Prove that f is 1-Lipschitz.
Hint: similar to 2a3.

More complicated percolation-type properties may be treated in the same way.

2a6 Exercise. Show that 2a4, 2a5 are special cases of the general claim below, and prove the claim.

Let E be a finite set, and $\mathcal{E}$ some nonempty set of nonempty subsets of E. Define $f : \mathbb{R}^E \to \mathbb{R}$ by $f(x) = a$, where a is the greatest number such that the set $E_a = \{e \in E : x(e) \ge a\}$ contains some (at least one) element of $\mathcal{E}$. Then f is 1-Lipschitz.

We turn to Lipschitz functions of matrices. The n^2-dimensional space of all $n \times n$-matrices (with real elements) may be written as $\mathbb{R}^E$, $E = \{1,\dots,n\} \times \{1,\dots,n\}$, its Euclidean norm being called the Hilbert-Schmidt (matrix) norm,

$$\|A\|_{\mathrm{HS}} = \Big(\sum_{k,l} A_{k,l}^2 \Big)^{1/2} .$$

However, we deal with the $\frac12 n(n+1)$-dimensional (sub)space of all *symmetric* (in the sense that $A_{k,l} = A_{l,k}$) $n \times n$-matrices A. It inherits the norm,

$$\|A\|_{\mathrm{HS}} = \Big(\sum_{k,l} A_{k,l}^2 \Big)^{1/2} = \Big(\sum_k A_{k,k}^2 + 2 \sum_{k<l} A_{k,l}^2 \Big)^{1/2} ,$$

[1] The two towns are connected until the flood is strong enough...

and we may consider Lipschitz functions on this (Euclidean) space.

Each symmetric matrix has its spectrum (eigenvalues, with multiplicities)

$$\operatorname{spec}(A) = (\lambda_1, \dots, \lambda_n)\,,$$

$-\infty < \lambda_1 \le \dots \le \lambda_n < \infty$. The spectrum is uniquely determined by the equality

$$\operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}\bigr) = \mathrm{e}^{\mathrm{i}t\lambda_1} + \dots + \mathrm{e}^{\mathrm{i}t\lambda_n} \quad \text{for all } t \in \mathbb{R}\,;$$

recall that $\operatorname{trace}(A) = A_{1,1} + \dots + A_{n,n}$ and $\mathrm{e}^{\mathrm{i}tA} = \sum_{k=0}^{\infty} \frac{1}{k!}(itA)^k$.

The function

$$A \mapsto \operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}\bigr)$$

maps the space of symmetric matrices into $\mathbb{C}$ (rather than $\mathbb{R}$); still, its Lipschitz constant is defined evidently (since $\mathbb{C}$ is also a metric space). The function appears to be $(\sqrt{n}\,|t|)$-Lipschitz; that is,

$$|\operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}\bigr) - \operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tB}\bigr)| \le \sqrt{n}\,|t| \|A - B\|_{\mathrm{HS}}\,. \tag{2a7}$$

The proof uses the gradient of the function,

$$\begin{aligned}\operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}t(A+B)}\bigr) &= \operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}(\mathbf{1} + \mathrm{i}tB)\bigr) + o(\|B\|) = \\ &= \operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}\bigr) + \mathrm{i}t\operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}B\bigr) + o(\|B\|)\end{aligned}$$

for $\|B\| \to 0$ (but do not think that $\mathrm{e}^{\mathrm{i}t(A+B)} = \mathrm{e}^{\mathrm{i}tA}(\mathbf{1} + \mathrm{i}tB) + o(\|B\|)$, it need not hold for noncommuting A, B), and the estimation $|\operatorname{trace}\bigl(\mathrm{e}^{\mathrm{i}tA}B\bigr)| = |\langle B, \mathrm{e}^{-\mathrm{i}tA}\rangle_{\mathrm{HS}}| \le \|B\|_{\mathrm{HS}}\|\mathrm{e}^{-\mathrm{i}tA}\|_{\mathrm{HS}} \le \|B\|_{\mathrm{HS}} \cdot \sqrt{n}\|\mathrm{e}^{-\mathrm{i}tA}\| = \sqrt{n}\|B\|_{\mathrm{HS}}$.

2b No fear of high dimension

The higher the dimension, the more complicated the mathematics, says the usual wisdom. However, it fails in the Gaussian paradise; here, dimension does not matter, in the following sense.

2b1 Theorem. For every $n = 1, 2, \dots$ and every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ there exists a 1-Lipschitz function $g : \mathbb{R} \to \mathbb{R}$ such that f and g are identically distributed (on $(\mathbb{R}^n, \gamma^n)$ and $(\mathbb{R}, \gamma^1)$ respectively).

The theorem will be proven later (in 2h, 2i).

Less formally, we may say that 'the distribution of f on $(\mathbb{R}^n, \gamma^n)$ is more concentrated than $\mathrm{N}(0, 1)$', by which I mean existence of g as above. Alternatively we may say 'f is more concentrated than γ^1' etc.

2b2 Exercise. Deduce from Theorem 2b1 the following claim.

For every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ there exists $a \in \mathbb{R}$ such that

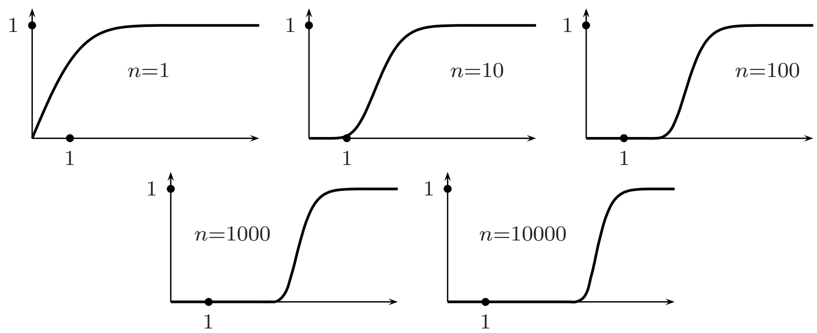
$$\gamma^n\{x \in \mathbb{R}^n : a - 2.6 \le f(x) \le a + 2.6\} \ge 0.99\,.$$

Hint: $\Phi(2.6) = 0.9953\dots$ (do not bother to prove this fact).

2c Elementary examples

Our first example is the norm $\|x\|_\infty = \max(|x_1|, \dots, |x_n|)$ (recall 2a2(a)). Its c.d.f. (cumulative distribution function) is just

$$\gamma^n\{x : \|x\|_\infty \le a\} = \big(\gamma^1[-a,a]\big)^n = \big(2\Phi(a) - 1\big)^n \,.$$



2c1 Exercise. Prove that $1 - \Phi(x) \sim \frac{1}{x}\Phi'(x)$ as $x \to \infty$.

Hint: $\int_0^\infty \Phi'(a+x)\,\mathrm{d}x = \Phi'(a)\int_0^\infty \mathrm{e}^{-ax-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{a}\Phi'(a)\int_0^\infty \mathrm{e}^{-x-\frac{x^2}{2a^2}}\,\mathrm{d}x$.

2c2 Exercise. Let a_n satisfy $1 - \Phi(a_n) \sim \frac{1}{2n}$. Prove that $a_n \to \infty$ and

$$1 - \Phi\Big(a_n + \frac{x}{a_n}\Big) \sim \frac{1}{2n}\mathrm{e}^{-x} \quad \text{as } n \to \infty$$

for every $x \in \mathbb{R}$.

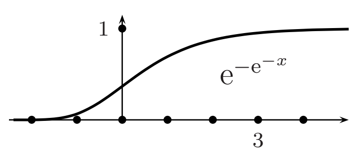
Hint: $\Phi'\big(a_n + \frac{x}{a_n}\big) = \Phi'(a_n)\exp\big(-x + \frac{x^2}{2a_n^2}\big)$.

2c3 Exercise. Prove that

$$\mathbb{P}\left(a_n - \frac{3}{a_n} \le \|x\|_\infty \le a_n + \frac{3}{a_n}\right) \ge 0.9$$

for all n large enough. Here a_n are as in 2c2.

Hint: $\big(2\Phi(a_n + \frac{x}{a_n}) - 1\big)^n \to \mathrm{e}^{-\mathrm{e}^{-x}}$ as $n \to \infty$.



In fact, $a_n \sim \sqrt{2\ln n}$.

We see that for large n the distribution has a standard shape, up to a linear transformation with two parameters; the location parameter $a_n \to \infty$, and the scaling parameter $1/a_n \to 0$. Theorem 2b1 restricts the scaling parameter only; namely, it must be bounded. In this example it tends to 0, which is another story.

Our second elementary example is Euclidean norm, $|x| = \sqrt{x_1^2 + \dots + x_n^2}$ (recall 2a2(b)). Consider first the squared norm, $|x|^2 = x_1^2 + \dots + x_n^2$; it is the sum of n i.i.d. (independent, identically distributed) random variables.

2c4 Exercise. For all $n = 1, 2, \dots$

(a) $\int_{\mathbb{R}^n} |x|^2 \, \gamma^n(\mathrm{d}x) = n$;

(b) $\int_{\mathbb{R}^n} (|x|^2 - n)^2 \, \gamma^n(\mathrm{d}x) = 2n$;

(c) $\gamma^n\{x \in \mathbb{R}^n : n - 10\sqrt{2n} \le |x|^2 \le n + 10\sqrt{2n}\} \ge 0.99$.

Prove it.

Hints: $\int_{-\infty}^{+\infty} u^2 \gamma^1(\mathrm{d}u) = 1$; $\int_{-\infty}^{+\infty} (u^2 - 1)^2 \gamma^1(\mathrm{d}u) = 2$; use Chebyshev's inequality.

It follows that

$$\gamma^n\{x \in \mathbb{R}^n : \sqrt{n} - 10 \le |x| \le \sqrt{n} + 10\} \ge 0.99 \tag{2c5}$$

for all n large enough. Note that the interval $[\sqrt{n} - 10, \sqrt{n} + 10]$ moves to ∞, but its length remains bounded, in agreement with Theorem 2b1. By analogy with the previous example (of $\|x\|_\infty$) you could think that $\gamma^n\{x : \sqrt{n} - 10 \le |x| \le \sqrt{n} + 10\}$ converges to 1, but it does not. In fact, CLT (the central limit theorem) ensures that the distribution of $\frac{|x|^2 - n}{\sqrt{2n}}$ converges to γ^1 (as $n \to \infty$). It follows that the distribution of $\sqrt{2}(|x| - \sqrt{n})$ converges to γ^1.

2d Non-elementary examples

Let γ be a Gaussian measure on $\mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} x_k^2 \, \gamma(\mathrm{d}x_1 \dots \mathrm{d}x_n) \le 1 \quad \text{for } k = 1, \dots, n\,. \tag{2d1}$$

Note that covariances $\int x_k x_l \gamma(\mathrm{d}x_1 \dots \mathrm{d}x_n)$ need not vanish.

Feel free to use Theorem 2b1 (within 2d, not 2e).

2d2 Exercise. In the special case $\gamma = \gamma^n$ the distribution of the Euclidean norm $|x| = \sqrt{x_1^2 + \dots + x_n^2}$ on $(\mathbb{R}^n, \gamma)$ is more concentrated than $\mathrm{N}(0, 1)$, but in general it is not.

Prove it.

Hint: try a one-dimensional γ.

2d3 Exercise. The distribution of the norm $\|x\|_\infty = \max(|x_1|, \dots, |x_n|)$ on $(\mathbb{R}^n, \gamma)$ is more concentrated than $\mathrm{N}(0,1)$.

Prove it.

Hint: by 1f3, γ is the image of γ^m under a linear embedding $\mathbb{R}^m \to \mathbb{R}^n$; each coordinate on $\mathbb{R}^n$ corresponds to a 1-Lipschitz function on $\mathbb{R}^m$; the same holds for $\|\cdot\|_\infty$.

Similarly to 2a1, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called C-Lipschitz on $(\mathbb{R}^n, \|\cdot\|_\infty)$, if

$$|f(x) - f(y)| \le C\|x - y\|_\infty \quad \text{for all } x, y \in \mathbb{R}^n\,. \tag{2d4}$$

2d5 Exercise. Prove that every 1-Lipschitz function on $\mathbb{R}^n$ (that is, on the Euclidean space $(\mathbb{R}^n, |\cdot|)$) is a 1-Lipschitz function on $(\mathbb{R}^n, \|\cdot\|_\infty)$, but the converse does not hold.

Hint: $\|x\|_\infty \le |x|$.

2d6 Exercise. If $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz on $(\mathbb{R}^n, \|\cdot\|_\infty)$ then the distribution of f on $(\mathbb{R}^n, \gamma)$ is more concentrated than $\mathrm{N}(0,1)$.

Prove it.

Hint: similar to 2d3.

2d7 Exercise. Prove that the functions of 2a3–2a6 are 1-Lipschitz on $(\mathbb{R}^n, \|\cdot\|_\infty)$.

We see that the distribution of $f(x_1, \dots, x_n) = \ln\big(\mathrm{e}^{x_1} + \dots + \mathrm{e}^{x_n}\big)$ (recall 2a3) on $(\mathbb{R}^n, \gamma)$ is more concentrated than $\mathrm{N}(0,1)$ for every γ (satisfying (2d1)). Probably we could check it elementarily for $\gamma = \gamma^n$, but not in general!

The same can be said about $\mathrm{Me}(x_1, \dots, x_{2n-1})$ (recall 2a4). Its distribution can be calculated explicitly for $\gamma = \gamma^n$, but not in general!

The distribution of the function of 2a5 ('the critical flood level') depends on the graph and the covariances. Even for $\gamma = \gamma^n$ we are unable to calculate the distribution, if the graph is not too simple. Still, it is more concentrated than $\mathrm{N}(0,1)$, irrespective of the graph and the covariances!

The same can be said about 2a6.

However, (2a7) does not generalize to $\|\cdot\|_\infty$ (as far as I know). For a random matrix with orthogaussian elements (but symmetric), the distribution of $\frac{1}{\sqrt{n}\,|t|} \operatorname{trace}\big(\mathrm{e}^{\mathrm{i}tA}\big)$ is more concentrated than $\mathrm{N}(0,1)$.

Theorem 2b1 is both general and strong. A much weaker statement would be enough for some applications, such as (say)

$$\gamma\{x : a - 1000 \le f(x) \le a + 1000\} \ge 0.5 \quad \text{for some } a \tag{2d8}$$

(recall 2b2); it is vital that a single estimation holds for Gaussian measures of all dimensions (thus, also for infinite dimension, as we will see). We have no way to (2d8) for f of 2d3 other than concentration of Gaussian measures. The more so for 2d7, and the spectra of random matrices. We cannot economize on the generality of Theorem 2b1. However, we can economize on its strength, and will do so in the next subsection.

2e Crude results, easy proofs

For large n, a typical point of $(\mathbb{R}^n, \gamma^n)$ is at a distance of $\approx \sqrt{n}$ from the origin, see (2c5). Two typical points x, y are at a distance of $\approx \sqrt{2n}$ from each other, see below.

2e1 Exercise. (a) If f, g are orthogaussian then $\frac{1}{\sqrt{2}}(f+g), \frac{1}{\sqrt{2}}(f-g)$ are orthogaussian. In particular, $\frac{1}{\sqrt{2}}(f-g) \sim \mathrm{N}(0,1)$.

(b) If $f_1, \dots, f_n, g_1, \dots, g_n$ are orthogaussian, then $\frac{1}{\sqrt{2}}(f_1+g_1), \dots, \frac{1}{\sqrt{2}}(f_n+g_n), \frac{1}{\sqrt{2}}(f_1-g_1), \dots, \frac{1}{\sqrt{2}}(f_n-g_n)$ are orthogaussian. In particular, $\bigl(\frac{1}{\sqrt{2}}(f_1-g_1), \dots, \frac{1}{\sqrt{2}}(f_n-g_n)\bigr) \sim \gamma^n$.

(c) $(\gamma^n \times \gamma^n)\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : \sqrt{n}-10 \le \frac{1}{\sqrt{2}}|x-y| \le \sqrt{n}+10\} \ge 0.99$.

Prove it.

Hint: use (2c5).

For a 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ we get $|f(x)-f(y)| \le |x-y| \approx \sqrt{2n}$ (for typical x, y), while Theorem 2b1 (via 2b2) gives $|f(x)-f(y)| \le 6$ for 98% of pairs (x,y). Quite strange! Could we refute the theorem by constructing a counterexample, a 1-Lipschitz f such that $|f(x)-f(y)| \approx |x-y|$ for typical x, y? To this end the gradient ∇f should be (roughly) collinear with $x-y$ along the straight segment

$$\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\} \subset \mathbb{R}^n \,. \tag{2e2}$$

However, we face an obstacle: these segments intersect. A lot of segments pass through a single point in various directions.

An obstacle to a counterexample could be a clue to a proof; let us try. We have

$$\begin{gathered}
|f(x)-f(y)| = \biggl| \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\alpha} f(\alpha x + (1-\alpha)y) \, \mathrm{d}\alpha \biggr| \le \\
\le \int_0^1 \biggl| \frac{\mathrm{d}}{\mathrm{d}\alpha} f(\alpha x + (1-\alpha)y) \biggr| \, \mathrm{d}\alpha = \int_0^1 \bigl| \langle \nabla f(\alpha x + (1-\alpha)y), x-y \rangle \bigr| \, \mathrm{d}\alpha \,;
\end{gathered}$$

$$\iint |f(x)-f(y)|\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le \tag{2e3}$$
$$\le \int_0^1 \mathrm{d}\alpha \iint \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)\big|\langle \nabla f(\alpha x+(1-\alpha)y), x-y\rangle\big|\,.$$

Note that $|\nabla f| \le 1$ everywhere, since f is 1-Lipschitz. Consider first $\alpha = 0.5$; hopefully, other α are similar. We want to estimate from above the integral

$$\iint \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)\Big|\Big\langle \nabla f\Big(\frac{x+y}{2}\Big), x-y\Big\rangle\Big|\,.$$

Nice! The joint distribution of $u = \frac{1}{\sqrt{2}}(x+y)$ and $v = \frac{1}{\sqrt{2}}(x-y)$ is still $\gamma^n \times \gamma^n$ (recall 2e1), and the integral becomes

$$\iint \gamma^n(\mathrm{d}u)\gamma^n(\mathrm{d}v)\big|\langle \nabla f(u/\sqrt{2}), \sqrt{2}v\rangle\big|\,.$$

2e4 Exercise. Prove that $\int |\langle x,y\rangle|\,\gamma^n(\mathrm{d}y) = \sqrt{2/\pi}\,|x|$ for all $x \in \mathbb{R}^n$.

Hint: $\int_{\mathbb{R}} |t|\,\gamma^1(\mathrm{d}t) = \sqrt{2/\pi}$, since $t\Phi'(t) = -\Phi''(t)$.

Note that $|\langle x,y\rangle| \le |x||y| \approx \sqrt{n}\,|x|$ for typical y; the average over y remains bounded (as $n \to \infty$), since a typical y is nearly orthogonal to x, — just the desired effect!

We get

$$\sqrt{2}\int \gamma^n(\mathrm{d}u)\int \gamma^n(\mathrm{d}v)\big|\langle \nabla f(u/\sqrt{2}), v\rangle\big| = \sqrt{2}\int \gamma^n(\mathrm{d}u)\sqrt{\tfrac{2}{\pi}}|\nabla f(u/\sqrt{2})| \le \tfrac{2}{\sqrt{\pi}}\,,$$

which is bounded in n. The point $\alpha = 0.5$ of (2e2) is done. In order to generalize our argument for other $\alpha \in [0,1]$ we need to estimate $\iint \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)|\langle \nabla f(\alpha x+(1-\alpha)y), x-y\rangle|$. However, we cannot! Take for instance $\alpha = 1$. The average of $\langle \nabla f(x), x-y\rangle$ over y is $\langle \nabla f(x), x\rangle$, which can be large. Say, for $f(x) = |x|$ we have $\nabla f(x) = x/|x|$ and $\langle \nabla f(x), x\rangle = |x| \approx \sqrt{n}$. What is wrong?

Here is what is wrong. The typical $|f(x)-f(y)|$ cannot be large (according to Theorem 2b1), however, the typical $\big|f(x) - f\big(\frac{x+y}{2}\big)\big|$ can be large (try $f(x) = |x|$), since $\big|\frac{x+y}{2}\big| \approx \sqrt{n/2}$ is far from $\sqrt{n}$. We should not connect two points, situated (roughly) on the $\sqrt{n}$-sphere, by a straight line. Rather, we should connect them by an arc on the sphere! It means

$$\{x\cos\alpha + y\sin\alpha : \alpha \in [0,\pi/2]\} \subset \mathbb{R}^n \tag{2e5}$$

instead of (2e2). We get

$$\iint |f(x)-f(y)|\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le$$
$$\le \int_0^{\pi/2} \mathrm{d}\alpha \iint \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y)\big|\langle \nabla f(x\cos\alpha + y\sin\alpha), -x\sin\alpha + y\cos\alpha\rangle\big| =$$
$$= \int_0^{\pi/2} \mathrm{d}\alpha \iint \gamma^n(\mathrm{d}u)\gamma^n(\mathrm{d}v)\big|\langle \nabla f(u), v\rangle\big| =$$
$$= \frac{\pi}{2}\cdot\sqrt{\frac{2}{\pi}}\int |\nabla f(u)|\,\gamma^n(\mathrm{d}u) \le \sqrt{\frac{\pi}{2}}\,,$$

which is bounded in n. Nice!

2e6 Exercise. Fill in the details in the calculation above.

Hint: 2e1(b) means $\alpha = \pi/4$; generalize it to all α.

2e7 Exercise. Prove the inequality $\iint |f(x)-f(y)|\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le \sqrt{\pi/2}$ for all (not just smooth) 1-Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$.

Hint: a 1-Lipschitz function need not be smooth, but can be approximated uniformly by smooth 1-Lipschitz functions. For example (see also 2f10),

$$f_\varepsilon(x) = \int f(x+\varepsilon y)\,\gamma^n(\mathrm{d}y)\,;$$
$$|f_\varepsilon(x) - f(x)| \le C_n\varepsilon \quad \text{for all } x, \quad \text{where } C_n = \int |y|\,\gamma^n(\mathrm{d}y) < \infty\,.$$

2e8 Exercise. For every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ there exists $a \in \mathbb{R}$ such that

(a) $\int_{\mathbb{R}^n} |f(x) - a|\,\gamma^n(\mathrm{d}x) \le \sqrt{\pi/2}$;

(b) $\gamma^n\{x \in \mathbb{R}^n : a - 130 \le f(x) \le a + 130\} \ge 0.99$.

Prove it.

Hint: 2e7, Fubini theorem, and Markov/Chebyshev inequality.

Compare (b) with (2d8). Fortunately, the number 130 is bounded in n; unfortunately, it is much worse than 2.6 (recall 2b2). Striving to a better estimation, let us try the second moment. Here is a counterpart of 2e4.

2e9 Exercise. Prove that $\int |\langle x, y\rangle|^2\,\gamma^n(\mathrm{d}y) = |x|^2$ for all $x \in \mathbb{R}^n$.

2e10 Exercise. Prove the inequality $\iint |f(x)-f(y)|^2\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le (\pi/2)^2$ for all 1-Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$.

Hint: $|\int_0^{\pi/2}(\dots)\,\mathrm{d}\alpha|^2 \le \frac{\pi}{2}\int_0^{\pi/2} |\dots|^2\,\mathrm{d}\alpha$.

The Fubini theorem gives $\int_{\mathbb{R}^n} |f(x)-a|^2\,\gamma^n(\mathrm{d}x) \le (\pi/2)^2$ for some a; but we can do twice better.

2e11 Exercise. Prove that $\iint |f(x)-f(y)|^2\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) = 2\int |f(x)-a|^2\,\gamma^n(\mathrm{d}x)$, where $a = \int f(x)\,\gamma^n(\mathrm{d}x)$.

Hint: $\iint (f(x)-a)(f(y)-a)\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) = 0$.

2e12 Exercise. For every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$,

(a) $\int_{\mathbb{R}^n} |f(x)-a|^2\,\gamma^n(\mathrm{d}x) \le \pi^2/8$,

(b) $\gamma^n\{x\in\mathbb{R}^n : a-12 \le f(x) \le a+12\} \ge 0.99$

for some $a\in\mathbb{R}$, namely, $a = \int f(x)\,\gamma^n(\mathrm{d}x)$.

Prove it.

Compare 2e12(b) with 2e8(b) and 2b2. Further in the same direction, we may choose a convex function $M : \mathbb{R}\to[0,\infty)$, $M(0)=0$ (not just $M(t)=t$ or $M(t)=t^2$), note that $M\big(\frac{2}{\pi}\int_0^{\pi/2}(\dots)\,\mathrm{d}\alpha\big) \le \frac{2}{\pi}\int_0^{\pi/2} M(\dots)\,\mathrm{d}\alpha$, and get

$$\iint M\big(f(x)-f(y)\big)\,\gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le$$
$$\le \frac{2}{\pi}\int_0^{\pi/2}\mathrm{d}\alpha \iint \gamma^n(\mathrm{d}u)\gamma^n(\mathrm{d}v) M\Big(\frac{\pi}{2}\langle\nabla f(u),v\rangle\Big) \le \int M\Big(\frac{\pi}{2}t\Big)\,\gamma^1(\mathrm{d}t)\,;$$

$$\int M\big(f(x)-a\big)\,\gamma^n(\mathrm{d}x) \le \int M\Big(\frac{\pi}{2}t\Big)\,\gamma^1(\mathrm{d}t) \quad \text{for some } a\,;$$
$$\gamma^n\{x\in\mathbb{R}^n : a-b \le f(x) \le a+b\} \ge 1 - \frac{\int M\big(\frac{\pi}{2}t\big)\,\gamma^1(\mathrm{d}t)}{M(b)}\,.$$

Now we may minimize the ratio in M. It is enough to check functions of the form $M(t) = (|t|-c)^+$. In this case, $\int M\big(\frac{\pi}{2}t\big)\,\gamma^1(\mathrm{d}t) = \pi\Phi'\big(\frac{2}{\pi}c\big) - 2c\Phi\big(-\frac{2}{\pi}c\big)$. Denote it by $g(c)$, then $g'(c) = -2\Phi\big(-\frac{2}{\pi}c\big)$; we see that g is convex, and the optimal c satisfies $g(c) = -(b-c)g'(c)$ (unless b is too small), and $\frac{g(c)}{b-c} = -g'(c) = 2\Phi\big(-\frac{2}{\pi}c\big)$. We get $b = c - \frac{g(c)}{g'(c)} = \frac{\pi}{2}\frac{\Phi'}{\Phi}\big(-\frac{2}{\pi}c\big)$ and so,

$$\gamma^n\Big\{x\in\mathbb{R}^n : a - \frac{\pi}{2}\frac{\Phi'}{\Phi}\big(-\frac{2}{\pi}c\big) \le f(x) \le a + \frac{\pi}{2}\frac{\Phi'}{\Phi}\big(-\frac{2}{\pi}c\big)\Big\} \ge 1 - 2\Phi\Big(-\frac{2}{\pi}c\Big)\,.$$

For instance, $c = 4.05$ gives

$$\gamma^n\{x\in\mathbb{R}^n : a-4.6 \le f(x) \le a+4.6\} \ge 0.99\,.$$

Compare it with 2e12(b) and 2b2. For large c we have $\frac{\Phi'}{\Phi}\big(-\frac{2}{\pi}c\big) \sim \frac{2}{\pi}c$ (by 2c1), thus,

$$\gamma^n\big\{x\in\mathbb{R}^n : a-(c+o(c)) \le f(x) \le a+(c+o(c))\big\} \ge 1 - 2\Phi\Big(-\frac{2}{\pi}c\Big)\,,$$

as if $\frac{f-a}{\pi/2}$ would be distributed (approximately) $\mathrm{N}(0,1)$. A good result! Tails of f are (at most) normal. However, the constant $\pi/2$ is a loss.

2f Subtler arguments, stronger results

The origin of the loss is evident: the arc (2e5) is longer (by $\pi/2$) than the radius ($\sqrt{n}$) of the sphere. It is not at all evident, how to avoid the loss. The experience of (2e2) tells us that we cannot leave the sphere; and surely we cannot go straight on the sphere!

Well, we can go (almost) straight on the sphere, provided that we do not go far. Instead of $\alpha \in [0, \pi/2]$ in (2e5), let us try $\alpha \in [0, \varepsilon]$ for a small ε.

2f1 Exercise. Prove the inequality $\iint |f(x) - f(x\cos\varepsilon + y\sin\varepsilon)|^2 \, \gamma^n(\mathrm{d}x)\gamma^n(\mathrm{d}y) \le \varepsilon^2$ for all 1-Lipschitz functions $f : \mathbb{R}^n \to \mathbb{R}$.

Hint: similar to 2e10.

So what? Yes, $\|f_0 - f_\varepsilon\| \le \varepsilon$ in $L_2(\gamma^n \times \gamma^n)$, but we need $\|f_0 - f_{\pi/2}\|$ and we get $\pi/2$ again.

However, small ε opens a new way.

2f2 Exercise. (a) For every smooth Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, prove that

$$\int |f(x) - f(x\cos\varepsilon + y\sin\varepsilon)|^2 \, \gamma^n(\mathrm{d}y) = \varepsilon^2 |\nabla f(x)|^2 + o(\varepsilon^2) \quad \text{as } \varepsilon \to 0$$

for all $x \in \mathbb{R}^n$.

(b) Does it hold for all (not just Lipschitz) smooth functions?

(c) Can we replace $f(x)$ with $f(x\cos\varepsilon)$?

Hints: (a) $f(x+\Delta x) = f(x) + \langle \nabla f(x), \Delta x\rangle + o(|\Delta x|)$; use the dominated convergence theorem; (b) f need not be bounded.

2f3 Exercise. For every smooth Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, prove that

$$\int \left| f(x+\varepsilon y) - \int f(x+\varepsilon y_1) \, \gamma^n(\mathrm{d}y_1) \right|^2 \gamma^n(\mathrm{d}y) = \varepsilon^2 |\nabla f(x)|^2 + o(\varepsilon^2) \quad \text{as } \varepsilon \to 0$$

for all $x \in \mathbb{R}^n$.

Hint: $\int f(x + \varepsilon y_1) \, \gamma^n(\mathrm{d}y_1) = f(x) + o(\varepsilon)$.

The language of convolutions will be convenient to us,

$$(f * \gamma^n)(x) = \int f(x+y) \, \gamma^n(\mathrm{d}y) \, ,$$

$$(f * \gamma^n_t)(x) = \int f(x + \sqrt{t}\, y) \, \gamma^n(\mathrm{d}y) \, ;$$

these are well-defined for all Lipschitz functions f, and moreover, for all continuous (or just measurable) functions f such that $|f(x)| = O(\mathrm{e}^{\varepsilon|x|^2})$ as $|x| \to \infty$ for all $\varepsilon > 0$. (Polynomials of Lipschitz functions fit.) Note that γ_t^n (or just γ_t) is the Gaussian measure on $\mathbb{R}^n$ homothetic to γ^n with the coefficient $\sqrt{t}$. Why not t? Here is why.

2f4 Exercise. Prove that $(f * \gamma_s) * \gamma_t = f * \gamma_{s+t}$ for all $s, t \in [0, \infty)$.

Hint: $\int f(x + ur\cos\alpha + vr\sin\alpha)\,\gamma^n(\mathrm{d}u)\gamma^n(\mathrm{d}v) = \int f(x + ry)\,\gamma^n(\mathrm{d}y)$, see the hint to 2e7.

In fact, $\gamma_s * \gamma_t = \gamma_{s+t}$.

If f is 1-Lipschitz then $f * \gamma_t$ is 1-Lipschitz for any t (see also the hint to 2e7).

2f5 Exercise. For every smooth Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\bigl(f^2 * \gamma_t - (f * \gamma_t)^2\bigr)(x) = t|\nabla f(x)|^2 + o(t) \quad \text{as } t \to 0$$

for all $x \in \mathbb{R}^n$.

Prove it.

Hint: it is a reformulation of 2f3.

2f6 Exercise. Deduce from Theorem 2b1 that

$$f^2 * \gamma^n - (f * \gamma^n)^2 \le 1 \quad \text{on } \mathbb{R}^n$$

for every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$.

Hints: apply the theorem to a shifted function, $g(y) = f(x+y)$; note that $\int f^2\,\mathrm{d}\gamma - (\int f\,\mathrm{d}\gamma)^2 = \min_a \int (f-a)^2\,\mathrm{d}\gamma$.

2f7 Exercise. Deduce from 2f6 that

$$f^2 * \gamma_t - (f * \gamma_t)^2 \le t \quad \text{on } \mathbb{R}^n$$

for every $t \in [0, \infty)$ and every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$.

Hint: apply 2f6 to rescaled function, $g(x) = f(x\sqrt{t})/\sqrt{t}$.

Without Theorem 2b1 we have the result of 2f7 for small t, up to $o(t)$ (recall 2f5). How to go further? Note that

$$f^2 * \gamma_1 - (f * \gamma_1)^2 = (f * \gamma_{1-t})^2 * \gamma_t \Big|_{t=0}^{t=1}.$$

This elegant interpolation between $f^2 * \gamma_1$ and $(f * \gamma_1)^2$ gives us a clever idea: maybe,

$$\frac{\mathrm{d}}{\mathrm{d}t}(f * \gamma_{1-t})^2 * \gamma_t \le 1\,; \tag{2f8}$$

this would be sufficient for proving 2f6 without Theorem 2b1!

We have (recall 2f4)

$$
\text{(2f9)} \quad (f * \gamma_{1-s})^2 * \gamma_s \big|_{s=t-\varepsilon}^{s=t} = (f * \gamma_{1-t})^2 * \gamma_t - (f * \gamma_{1-t+\varepsilon})^2 * \gamma_{t-\varepsilon} = \\ = \big(g^2 * \gamma_\varepsilon - (g * \gamma_\varepsilon)^2 \big) * \gamma_{t-\varepsilon} \,,
$$

where $g = f * \gamma_{1-t}$ is a 1-Lipschitz function; 2f5 gives $g^2 * \gamma_\varepsilon - (g * \gamma_\varepsilon)^2 \le \varepsilon + o(\varepsilon)$. Can we conclude that $(f * \gamma_{1-s})^2 * \gamma_s \big|_{s=t-\varepsilon}^{s=t} \le \varepsilon + o(\varepsilon)$? These $o(\varepsilon)$ are functions of x and ε, not just ε. We know that $\limsup_{\varepsilon \to 0} \frac{1}{\varepsilon} \big(g^2 * \gamma_\varepsilon - (g * \gamma_\varepsilon)^2 \big) \le 1$, therefore $\limsup_{\varepsilon \to 0} \frac{1}{\varepsilon} \big(\dots \big) \varphi_{t-\varepsilon} \le \varphi_t$, where φ_s is the density of γ_s; however, we cannot integrate it without an integrable majorant. Some additional effort is needed.

2f10 Exercise. (a) Let f be a 1-Lipschitz function $\mathbb{R}^n \to \mathbb{R}$, and $g = f * \gamma_t$; then g is twice continuously differentiable, and

$$
\frac{\partial^2}{\partial x_k \partial x_l} g(x_1, \dots, x_n) \le \sqrt{\frac{2}{\pi t}}
$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$.

(b) Every 1-Lipschitz function $\mathbb{R}^n \to \mathbb{R}$ can be approximated uniformly by 1-Lipschitz functions with bounded continuous first and second derivatives.

Prove it.

Hint. (a) For $n = 1$ and $t = 1$, $g''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \int f(x + y) \varphi(y) \, \mathrm{d}y = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \int f(z) \varphi(z - x) \, \mathrm{d}z = \int f(z) \varphi''(z - x) \, \mathrm{d}z$ (why?), thus g'' exists. Further, $g''(x) = - \int f'(z) \varphi'(z - x) \, \mathrm{d}z$, thus $|g''(x)| \le \int |\varphi'(z - x)| \, \mathrm{d}z = \sqrt{2/\pi}$ (for smooth f, or anyway?)... (b) See also 2e7.

2f11 Exercise. For every Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ with bounded continuous first and second derivatives prove that

$$
\big| f^2 * \gamma_t - (f * \gamma_t)^2 - t |\nabla f|^2 \big| \le C t^{3/2} \quad \text{on } \mathbb{R}^n
$$

for all $t \in [0, 1]$, with a constant $C \in (0, \infty)$ that depends on f only.

Hint: similar to 2f5; $|f(x + \Delta x) - (f(x) + \langle \nabla f(x), \Delta x \rangle)| \le \text{const} \cdot |\Delta x|^2$.

Returning to (2f9) and applying (to g) 2f11 instead of 2f5 we get $g^2 * \gamma_\varepsilon - (g * \gamma_\varepsilon)^2 \le \varepsilon + C \varepsilon^{3/2}$ where C depends only on $t \in [0, 1)$ (but may tend to ∞ as $t \to 1$). Therefore

$$
(f * \gamma_{1-s})^2 * \gamma_s \big|_{s=t-\varepsilon}^{s=t} \le \varepsilon + C \varepsilon^{3/2} \,.
$$

In order to get (2f8) it remains to prove existence of the derivative.

2f12 Exercise. (a) For every $x \in \mathbb{R}^n$ the function $t \mapsto \big((f * \gamma_{1-t})^2 * \gamma_t\big)(x)$ is continuously differentiable on $(0,1)$ and continuous on $[0,1]$;

(b) (2f8) holds for $t \in (0,1)$.

Prove it.

Now the result of 2f6 is proven (without Theorem 2b1). Substituting the origin, we get the following.

2f13 Proposition. For every 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ there exists $a \in \mathbb{R}$ such that $\int_{\mathbb{R}^n} |f(x) - a|^2 \, \gamma^n(\mathrm{d}x) \le 1$.

Compare it with 2e12(a).

2g Gaussian isoperimetry

For any $A \subset \mathbb{R}^n$ and $\varepsilon > 0$ we define the open ε-neighborhood

$$A_{+\varepsilon} = \{x \in \mathbb{R}^n : \exists a \in A \ |x - a| < \varepsilon\} . \tag{2g1}$$

2g2 Exercise. Prove that

$$A_{+\varepsilon} = \{x \in \mathbb{R}^n : \operatorname{dist}(x, A) < \varepsilon\}$$

and $\operatorname{dist}(x, A)$ defined as $\inf_{a \in A} |x - a|$ is a 1-Lipschitz function of x.

2g3 Exercise. Prove that

$$(A_{+\varepsilon})_{+\delta} \subset A_{+(\varepsilon+\delta)} .$$

Can we replace '$\subset$' with '$=$'?

2g4 Exercise.

$$\Phi^{-1}\big(\gamma^n(A_{+\varepsilon})\big) - \Phi^{-1}\big(\gamma^n(A)\big) \ge \varepsilon$$

for every $\varepsilon > 0$ and every closed set $A \subset \mathbb{R}^n$.

Deduce it from Theorem 2b1.

Hint: consider $f(x) = \operatorname{dist}(x, A)$.

By the way, the classical isoperimetric theorem on $\mathbb{R}^2$ may be formulated similarly:

$$\sqrt{\tfrac{1}{\pi} \operatorname{mes}_2(A_{+\varepsilon})} - \sqrt{\tfrac{1}{\pi} \operatorname{mes}_2(A)} \ge \varepsilon \tag{2g5}$$

for every $\varepsilon > 0$ and every compact set $A \subset \mathbb{R}^2$; here mes_2 stands for the 2-dimensional Lebesgue measure. And, of course,

$$\tfrac{1}{2} \operatorname{mes}(A_{+\varepsilon}) - \tfrac{1}{2} \operatorname{mes}(A) \ge \varepsilon \tag{2g6}$$

for every compact set $A \subset \mathbb{R}$, which is evident (think, why).

2g7 Exercise. (a) For a half-space $A \subset \mathbb{R}^n$ the inequality 2g4 becomes equality.

(b) For a disk $A \subset \mathbb{R}^2$ the inequality (2g5) becomes equality.

Prove it.

In fact, these are the only cases of equality.

A wonder: in Gaussian isoperimetry, the extremal sets are half-spaces (rather than balls)!

2g8 Exercise. (a) Deduce Theorem 2b1 from 2g4.

(b) Deduce from (2g5) that for every 1-Lipschitz function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x) \to +\infty$ as $|x| \to \infty$ there exists an increasing 1-Lipschitz function $g : [0, \infty) \to \mathbb{R}$ such that $g(t) \to +\infty$ as $t \to \infty$, and

$$\operatorname{mes}_2\{x \in \mathbb{R}^2 : f(x) \le a\} = \mu\{t \in [0, \infty) : g(t) \le a\}$$

where $\mu(\mathrm{d}x) = 2\pi x \, \mathrm{d}x$.

(c) Formulate the one-dimensional counterpart of (b) and prove it (via (2g6)).

Hints: (a) you may construct g such that $\gamma^n\{x : f(x) \le g(t)\} = \Phi(t)$, provided that the distribution of f has no atoms, that is, $\gamma^n\{x : f(x) = a\} = 0$ for all a; think what to do if there are atoms; (b) similarly, think about g such that $\operatorname{mes}_2\{x : f(x) \le g(t)\} = \pi t^2$.

We see that 2g4 is equivalent to Theorem 2b1.

For any measure μ on $\mathbb{R}^n$ (we really need only $\mu = \gamma^n$, and sometimes $\mu = \operatorname{mes}_n$ for comparison), the number

$$\mu^+(A) = \liminf_{\varepsilon \to 0+} \frac{1}{\varepsilon} \big(\mu(A_{+\varepsilon}) - \mu(A) \big) \tag{2g9}$$

is called the boundary μ-measure of A (or μ-perimeter, or Minkowski's surface measure with respect to μ). The '$\liminf$' is stipulated for strange sets; we really need only A such that the limit exists. Especially, for a domain $A \subset \mathbb{R}^2$ with a (piecewise) smooth boundary ∂A,

$$\begin{aligned} \operatorname{mes}_2(A) &\quad \text{is the area of } A\,, \\ \operatorname{mes}_2^+(A) &\quad \text{is the length of } \partial A\,, \\ \gamma^2(A) &\quad \text{is the integral of } \tfrac{1}{2\pi} \mathrm{e}^{-|x|^2/2} \text{ over } A\,, \\ \gamma_+^2(A) &\quad \text{is the integral of } \tfrac{1}{2\pi} \mathrm{e}^{-|x|^2/2} \text{ over } \partial A\,, \end{aligned} \tag{2g10}$$

and '$\liminf$' may be replaced with '$\lim$'. Of course, γ_+^2 means $(\gamma^2)^+$.

2g11 Exercise. (a) Deduce from (2g5) the inequality

$$L \ge 2\pi\sqrt{\tfrac{1}{\pi}S}\,, \qquad \text{(that is, } L^2 \ge 4\pi S\text{)}$$

where $S = \operatorname{mes}_2(A)$, $L = \operatorname{mes}_2^+(A)$ and A is a compact set in $\mathbb{R}^2$. Explain the meaning of the inequality in terms of elementary geometry. (Why is it called 'isoperimetric'?)

(b) Deduce from (2g4) the *Gaussian isoperimetric inequality*

$$\gamma^n_+(A) \ge \Phi'\big(\Phi^{-1}(\gamma^n(A))\big)$$

for every closed set $A \subset \mathbb{R}^n$.

Hint (to (a)): first, think in terms of $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\big|_{\varepsilon=0}\sqrt{\frac{1}{\pi}\operatorname{mes}_2(A_{+\varepsilon})}$; second, think how to make it rigorous.

Again, half-spaces are extremal for 2g11(b), and disks — for 2g11(a). (Be careful with uniqueness...)

2g12 Exercise. (a) Deduce (2g5) from 2g11(a).

(b) Deduce 2g4 from 2g11(b).

Hint: first, think in terms of $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}(\dots)$ (not only at $\varepsilon = 0$). Second, think how to make it rigorous; to this end, forget integration and recall the first-year analysis, namely, *proofs* of theorems that relate $f(b) - f(a)$ to $f'(x)$ for $x \in (a, b)$.

We see that the Gaussian isoperimetric inequality 2g11(b) is equivalent to (2g4, and therefore to) Theorem 2b1.

2h Gaussian isoperimetry: a functional form

2h1 Exercise. (a) Let $f : \mathbb{R} \to \mathbb{R}$ be a function with bounded continuous first and second derivatives (f', f''), and

$$A = \big\{(x, y) \in \mathbb{R}^2 : y \le f(x)\big\}\,;$$

then

$$\gamma^2(A) = \int_{-\infty}^{+\infty} \Phi\big(f(x)\big)\,\gamma^1(\mathrm{d}x)\,,$$

$$\gamma^2_+(A) = \int_{-\infty}^{+\infty} \Phi'\big(f(x)\big)\sqrt{1 + f'^2(x)}\,\gamma^1(\mathrm{d}x)\,.$$

(b) Let $u : \mathbb{R} \to (0,1)$ be a function, bounded away from 0 and 1, with bounded continuous first and second derivatives (u', u''), and

$$A = \{(x,y) \in \mathbb{R}^2 : \Phi(y) \le u(x)\} ;$$

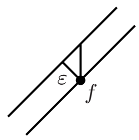
then

$$\gamma^2(A) = \int_{-\infty}^{+\infty} u(x)\,\gamma^1(\mathrm{d}x) ,$$
$$\gamma^2_+(A) = \int_{-\infty}^{+\infty} \sqrt{\Phi'^2(\Phi^{-1}(u(x))) + u'^2(x)}\,\gamma^1(\mathrm{d}x) .$$

Prove it.

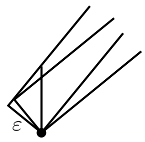
Hints: (a) roughly,

$$A_{+\varepsilon} \approx \{(x,y) : y \le f(x) + \varepsilon\sqrt{1+f'^2(x)}\} ;$$



more exactly,

$$\{(x,y) : y \le f(x) + \varepsilon\sqrt{1+f'^2(x)} - o(\varepsilon)\} \subset A_{+\varepsilon} \subset$$
$$\subset \{(x,y) : y \le f(x) + \varepsilon\sqrt{1+f'^2(x)} + o(\varepsilon)\} ;$$

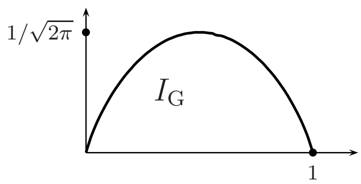


(b) try $f(x) = \Phi^{-1}\big(u(x)\big)$.

The function $I_{\mathrm{G}} : [0,1] \to \mathbb{R}$,

$$I_{\mathrm{G}}(p) = \Phi'\big(\Phi^{-1}(p)\big) , \qquad I_{\mathrm{G}}(0) = 0, \;\; I_{\mathrm{G}}(1) = 0 \tag{2h2}$$

is called the Gaussian isoperimetric function.



By the way, it satisfies a nice differential equation.

2h3 Exercise. $I_{\mathrm{G}}(p) I''_{\mathrm{G}}(p) = -1$ for all $p \in (0,1)$.

Prove it.

Hint: $I'_{\mathrm{G}}(p) = -\Phi^{-1}(p)$.

Using I_{G}, we may write the Gaussian isoperimetric inequality 2g11(b) as

$$\gamma^n_+(A) \ge I_{\mathrm{G}}\big(\gamma^n(A)\big)\,. \tag{2h4}$$

Similarly, $\operatorname{mes}_2^+(A) \ge I_2\big(\operatorname{mes}_2(A)\big)$, where $I_2(a) = 2\sqrt{\pi a}$; and $\operatorname{mes}_1^+(A) \ge I_1\big(\operatorname{mes}_1(A)\big)$, where $I_1(a) = 2$ for all a. Note that the classical isoperimetric function is dimension-specific, in contrast to Gaussian isoperimetric function.

Returning to 2h1(b) we see that

$$\gamma^2_+(A) = \int_{-\infty}^{+\infty} \sqrt{I^2_{\mathrm{G}}(u(x)) + u'^2(x)}\,\gamma^1(\mathrm{d}x)\,. \tag{2h5}$$

The Gaussian isoperimetric inequality (2h4) for $n = 2$ gives

$$\int \sqrt{I^2_{\mathrm{G}}(u(x)) + u'^2(x)}\,\gamma^1(\mathrm{d}x) \ge I_{\mathrm{G}}\bigg(\int u(x)\,\gamma^1(\mathrm{d}x)\bigg)\,; \tag{2h6}$$

this is the functional form of Gaussian isoperimetry. (Note the evident equality for $u(\cdot) = \mathrm{const.}$) Similarly, the Gaussian isoperimetric inequality (2h4) in $\mathbb{R}^{n+1}$ gives

$$\int \sqrt{I^2_{\mathrm{G}}(u(x)) + |\nabla u(x)|^2}\,\gamma^n(\mathrm{d}x) \ge I_{\mathrm{G}}\bigg(\int u(x)\,\gamma^n(\mathrm{d}x)\bigg) \tag{2h7}$$

for $u : \mathbb{R}^n \to (0,1)$ (satisfying the conditions...)

You may think that the functional form is weaker (than the original, geometric form), since sets of the form $\{(x,y) : y \le f(x)\}$ (or $\Phi(y) \le u(x)\dots$) are much more special than arbitrary closed sets (or even arbitrary domains with smooth boundaries). That is reasonable in each dimension separately. However, the functional form (2h7) implies Gaussian isoperimetry (2h4) in $\mathbb{R}^n$ (rather than $\mathbb{R}^{n+1}$); a surprise! Sets of the form $A \times \mathbb{R}$ in $\mathbb{R}^{n+1}$ (where A is a domain in $\mathbb{R}^n$) can be approximated by sets of the form $\{(x,y) : y \le f(x)\}$ (take $f_n(x) \to \pm\infty$), thus, (2h4) for $A \times \mathbb{R}$ follows from (2h7). On the other hand, (2h4) for $A \times \mathbb{R} \subset \mathbb{R}^{n+1}$ is equivalent to (2h4) for $A \subset \mathbb{R}^n$, since $(A \times \mathbb{R})_{+\varepsilon} = (A_{+\varepsilon}) \times \mathbb{R}$. We see that

$$\begin{aligned} &\big((2\text{h}4)\text{ for } n=1\big) \Longleftarrow \big((2\text{h}7)\text{ for } n=1\big) \Longleftarrow \\ &\qquad \Longleftarrow \big((2\text{h}4)\text{ for } n=2\big) \Longleftarrow \big((2\text{h}7)\text{ for } n=2\big) \Longleftarrow \dots \end{aligned}$$

therefore

$$\big((2\text{h}4)\text{ for all } n\big) \quad\Longleftrightarrow\quad \big((2\text{h}7)\text{ for all } n\big)\,.$$

A wonder: an n-dimensional body will be gradually turned into a half-space via a continuum of $(n+1)$-dimensional bodies! Nothing like that happens for classical isoperimetry.

2i Gaussian isoperimetry: a proof

2i1 Exercise. Deduce from (2h7) that

$$\sqrt{I_{\mathrm{G}}^2(u) + |\nabla u|^2} * \gamma^n \ge I_{\mathrm{G}}(u * \gamma^n)$$

and, more generally,

$$\sqrt{I_{\mathrm{G}}^2(u) + t|\nabla u|^2} * \gamma_t \ge I_{\mathrm{G}}(u * \gamma_t) \,.$$

Hint: similar to 2f6, 2f7.

Similarly to 2f, we try 'interpolation',

$$\sqrt{I_{\mathrm{G}}^2(u*\gamma_{1-t}) + t|\nabla (u*\gamma_{1-t})|^2} * \gamma_t \Big|_{t=0}^{t=1} = \sqrt{I_{\mathrm{G}}^2(u) + |\nabla u|^2} * \gamma_1 - I_{\mathrm{G}}(u * \gamma_1) \,;$$

in order to prove (2h7) (and therefore Theorem 2b1), it is sufficient to prove that

$$\frac{\mathrm{d}}{\mathrm{d}t} \Big(\sqrt{I_{\mathrm{G}}^2(u*\gamma_{1-t}) + t|\nabla (u*\gamma_{1-t})|^2} * \gamma_t \Big) \ge 0$$

for every function $u : \mathbb{R}^n \to (0,1)$, bounded away from 0 and 1, with bounded continuous first and second derivatives. Existence of the needed derivative $\frac{\mathrm{d}}{\mathrm{d}t}(\dots)$ and other derivatives that will be used is rather evident. Similarly to (2f9),

$$\begin{aligned} \sqrt{I_{\mathrm{G}}^2(u*\gamma_{1-s}) + s|\nabla (u*\gamma_{1-s})|^2} * \gamma_s \Big|_{s=t-\varepsilon}^{s=t} = \\ = \sqrt{I_{\mathrm{G}}^2(u*\gamma_{1-t}) + t|\nabla (u*\gamma_{1-t})|^2} * \gamma_t - \\ - \sqrt{I_{\mathrm{G}}^2(u*\gamma_{1-t+\varepsilon}) + (t-\varepsilon)|\nabla (u*\gamma_{1-t+\varepsilon})|^2} * \gamma_{t-\varepsilon} = \\ = \sqrt{I_{\mathrm{G}}^2(v) + t|\nabla v|^2} * \gamma_\varepsilon - \sqrt{I_{\mathrm{G}}^2(v*\gamma_\varepsilon) + (t-\varepsilon)|\nabla (v*\gamma_\varepsilon)|^2} \,, \end{aligned}$$

where $v = u * \gamma_{1-t}$ satisfies (at least) the same conditions as u; renaming it back to u we see[1] that it is sufficient to prove the following:

$$\text{(2i2)} \quad \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \Big(\sqrt{I_{\mathrm{G}}^2(u) + t|\nabla u|^2} * \gamma_\varepsilon - \sqrt{I_{\mathrm{G}}^2(u*\gamma_\varepsilon) + (t-\varepsilon)|\nabla (u*\gamma_\varepsilon)|^2} \Big) \ge 0 \,.$$

Now all convolutions are infinitesimal, and may be eliminated.

[1] Some additional effort is needed, as in 2f...

2i3 Exercise.
$$f * \gamma_\varepsilon = f + \frac{\varepsilon}{2}\Delta f + o(\varepsilon)$$
for every $f : \mathbb{R}^n \to \mathbb{R}$ with bounded continuous first and second derivatives; here
$$\Delta f(x) = \frac{\partial^2}{\partial x_1^2} f(x) + \dots + \frac{\partial^2}{\partial x_n^2} f(x) \,.$$
Prove it.

Hint: $f(x+\sqrt{\varepsilon}\, y) = f(x) + \sqrt{\varepsilon}\langle \nabla f(x), y\rangle + \frac{\varepsilon}{2}\sum_{k,l} \frac{\partial^2 f(x)}{\partial x_k \partial x_l} y_k y_l + o(\varepsilon)$; integrate it (and do not forget about an integrable majorant).

Now we stop thinking and start calculating. To simplify the calculations, we restrict ourselves to $n = 1$; the general case will be treated afterwards. For convenience, we use abbreviations
$$I = I_{\mathrm{G}}(u), \quad I_1 = I'_{\mathrm{G}}(u), \quad I_2 = I''_{\mathrm{G}}(u), \\ R = \sqrt{I^2 + tu'^2}\,;$$
note that $I' = I_1 u'$, $I'_1 = I_2 u'$.

2i4 Exercise. Check that
$$R'' = \frac{(I^2 + tu'^2)''}{2R} - \frac{(I^2+tu'^2)'^2}{4R^3} \,.$$

2i5 Exercise. Check that
$$\sqrt{I_{\mathrm{G}}^2(u*\gamma_\varepsilon) + (t-\varepsilon)(u*\gamma_\varepsilon)'^2} = R + \frac{\varepsilon}{2R}(II_1u'' + tu'u''' - u'^2) + o(\varepsilon) \,.$$

2i6 Exercise. Check that the limit (2i2) is equal to
$$\frac{(I^2+tu'^2)''}{4R} - \frac{(I^2+tu'^2)'^2}{8R^3} - \frac{II_1u'' + tu'u''' - u'^2}{2R} \,.$$

2i7 Exercise. Check that
$$(I^2+tu'^2)' = 2u'(II_1 + tu'')\,; \\ (I^2+tu'^2)'' = 2(I_1^2u'^2 + II_2u'^2 + II_1u'' + tu''^2 + tu'u''') \,.$$

2i8 Exercise. Check that
$$\frac{(I^2+tu'^2)''}{4R} - \frac{II_1u'' + tu'u''' - u'^2}{2R} = \frac{1}{2R}(I_1^2u'^2 + \underbrace{(II_2+1)}_{=0}u'^2 + tu''^2) \,.$$

Hint: use 2h3.

2i9 Exercise. Check that the limit (2i2) is equal to

$$\frac{t}{2R^3}(I_1 u'^2 - Iu'')^2 \,.$$

Hint: $R^2 = I^2 + tu'^2$.

We see that the limit cannot be negative, which completes the proof of (2h7) for $n = 1$.

I do not know, whether the arguments above can work for all n, or not. (Generalization is rather straightforward for 2i4–2i8 but not 2i9.) Fortunately, we do not need it, due to a striking property of (2h7), called *tensorization;* the n-dimensional inequality follows (rather easily) from its 1-dimensional special case!

2i10 Exercise. For all $f, g \in L_2(\mathbb{R}, \gamma^1)$,

$$\int \sqrt{f^2(x) + g^2(x)}\,\gamma^1(\mathrm{d}x) \ge \sqrt{\Big(\int f(x)\,\gamma^1(\mathrm{d}x) \Big)^2 + \Big(\int g(x)\,\gamma^1(\mathrm{d}x) \Big)^2} \,.$$

Prove it.

Hint: the function $(a,b) \mapsto \sqrt{a^2+b^2}$ is convex on $\mathbb{R}^2$.

(The same holds for any probability measure, Gaussian or not.)

For $n = 2$, the inequality (2h7) becomes

$$\iint \sqrt{I_{\mathrm{G}}^2\big(u(x,y)\big) + \Big(\frac{\partial}{\partial x} u(x,y) \Big)^2 + \Big(\frac{\partial}{\partial y} u(x,y) \Big)^2}\, \gamma^1(\mathrm{d}x)\gamma^1(\mathrm{d}y) \ge \\ \ge I_{\mathrm{G}} \bigg(\iint u(x,y)\,\gamma^1(\mathrm{d}x)\gamma^1(\mathrm{d}y) \bigg) \,.$$

In order to prove it, we keep y constant (for a while), integrate in x, apply 2i10 to $f(x) = \sqrt{I_{\mathrm{G}}^2\big(u(x,y)\big) + \big(\frac{\partial}{\partial x}u(x,y)\big)^2}$, $g(x) = \frac{\partial}{\partial y}u(x,y)$, and then apply the one-dimensional case of (2h7):

$$\int \sqrt{I_{\mathrm{G}}^2(u) + \Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial u}{\partial y}\Big)^2}\,\gamma^1(\mathrm{d}x) \ge \\ \ge \sqrt{ \Big(\underbrace{\int \sqrt{I_{\mathrm{G}}^2(u) + (\tfrac{\partial u}{\partial x})^2}\,\gamma^1(\mathrm{d}x)}_{\ge I_{\mathrm{G}}\left(\int u\,\gamma^1(\mathrm{d}x)\right)} \Big)^2 + \Big(\int \frac{\partial u}{\partial y}\,\gamma^1(\mathrm{d}x) \Big)^2 } \ge$$

$$\geq \sqrt{I_{\mathrm{G}}^2\Big(\int u\,\gamma^1(\mathrm{d}x)\Big) + \Big(\int \frac{\partial u}{\partial y}\,\gamma^1(\mathrm{d}x)\Big)^2} = \sqrt{I_{\mathrm{G}}^2\big(v(y)\big) + v'^2(y)}\,,$$

where

$$v(y) = \int u(x,y)\,\gamma^1(\mathrm{d}x)\,.$$

It remains to integrate in y and apply the one-dimensional case of (2h7) once again (this time to v). The two-dimensional case of (2h7) is proven.

2i11 Exercise. Prove (2h7) for all n.
Hint: by induction.

The functional form of Gaussian isoperimetry (for all n) implies its original, geometric form (for all n), as explained in 2h.

References

The proof given in 2i is a 'continuous' (semigroup) counterpart of a wonderful 'discrete' proof [2]; there, amazingly, $\mathbb{R}^1$ is factorized into two-point spaces! A shorter semigroup proof is available [1]. See also [3] about 2h, [4, Sect. 1.7] about 2e, and [5, Lemma 1.2(b)] about (2a7).

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

[2] S.G. Bobkov, *An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space,* Ann. Probab. **25**:1, 206–214 (1997).

[3] S. Bobkov, *A functional form of the isoperimetric inequality for the Gaussian measure,* J. Funct. Anal. **135**:1, 39–49 (1996).

[4] V.I. Bogachev, *Gaussian measures,* AMS 1998.

[5] A. Guionnet, O. Zeitouni, *Concentration of the spectral measure for large matrices,* Elect. Comm. in Probab. **5**, 119–136 (2000).

Index