1 The Gaussian isoperimetric inequality

An isoperimetric inequality gives a lower bound on the perimeter, or surface area, of a set in terms of its volume, or measure. In these notes we will be studying an isoperimetric inequality for the Gaussian measure; we will denote the Gaussian measure on \( \mathbb{R}^n \) by \( \gamma_n \) (or just \( \gamma \) sometimes). We will start with a definition of “Gaussian surface area” that comes with a natural geometric intuition: the Gaussian surface area is the infinitesimal gain in measure that comes from growing the set slightly.

**Definition 1.1.** The **gaussian Minkowski content** of a set \( A \subset \mathbb{R}^n \) is

\[
\gamma^+(A) = \liminf_{h \downarrow 0} \frac{\gamma(A + hB^n_2) - \gamma(A)}{h},
\]

where \( B^n_2 = \{x \in \mathbb{R}^n : |x| \leq 1\} \) is the Euclidean unit ball and \( A + B = \{x + y : x \in A, y \in B\} \) denotes Minkowski addition.

To be precise, the quantity we have just defined is often called the outer, lower Minkowski content: “outer” because we are expanding the set instead of contracting it, and “lower” because of the \( \lim inf \).

It isn’t too hard to show that for sufficiently nice sets, our definition of \( \gamma^+(A) \) coincides with another natural definition of Gaussian surface area.

**Exercise 1.** Define

\[
\tilde{\gamma}^+(A) = \int_{\partial A} \frac{d\gamma_n}{dx} d\mathcal{H}^{n-1},
\]

where \( \mathcal{H}^{n-1} \) denotes the \((n-1)\)-dimensional Hausdorff measure (or the classical surface area measure if \( \partial A \) is nice enough). Show that if \( A \) has Lipschitz boundary then \( \tilde{\gamma}^+(A) = \gamma^+(A) \).

Here is the Gaussian isoperimetric inequality, originally due independently to Borell [2] and to Sudakov and Tsirelson [3].

**Theorem 1.2.** For any measurable \( A \subset \mathbb{R}^n \),

\[
\gamma^+(A) \geq I(\gamma(A)),
\]

where

\[
I(x) = \phi(\Phi^{-1}(x))
\]

\[
\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

\[
\Phi(x) = \int_{-\infty}^{x} \phi(y) dy
\]

(and we extend \( I \) continuously to \([0, 1]\) by setting \( I(0) = I(1) = 0 \)).
The inequality in Theorem 1.2 is sharp, as can be seen by the example $A = \{ x \in \mathbb{R}^n : x_1 \leq b \}$ (for some $a \in \mathbb{R}$). For this set $A$ we have $\gamma(A) = \Phi(b)$ and $\gamma^+(A) = \phi(b)$. Of course, the rotational invariance of the Gaussian measure means that rotations of this set are also extremizers of Theorem 1.2: every set of the form $A = \{ x \in \mathbb{R}^n : \langle x, a \rangle \leq b \}$ for $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ satisfies $\gamma^+(A) = \phi(\Phi^{-1}(\gamma(A)))$.

1.1 Concentration

The Gaussian isoperimetric inequality was originally studied because of its applications to concentration inequalities for Gaussian processes: it implies that Lipschitz functions of standard Gaussian variables satisfy dimension-independent concentration bounds.

**Corollary 1.3.** For any 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, if $Z \sim \gamma$ and $M$ is a median of $f(Z)$ then for any $t > 0$,

$$\Pr(f(Z) \geq M + t) \leq \Phi(-t) \leq \exp(-t^2/2) \quad \Pr(f(Z) \leq M - t) \leq \Phi(-t) \leq \exp(-t^2/2)$$

The proof of Corollary 1.3 goes through an “integrated” form of the Gaussian isoperimetric inequality, which applies to non-infinitesimal expansions.

**Corollary 1.4.** For any measurable $A \subset \mathbb{R}^n$ and any $h > 0$, $\gamma(A + hB_2^n) \geq \Phi(\Phi^{-1}(\gamma(A)) + h)$.

**Exercise 2.**

1. Prove Corollary 1.4.
2. For every $a \in (0,1)$ and every $h > 0$, give an example of a set $A$ with $\gamma(A) = a$ and $\gamma(A + hB_2^n) = \Phi(\Phi^{-1}(\gamma(A)) + h)$. (This shows that Corollary 1.4 cannot be improved.)

**Exercise 3.**

1. In Corollary 1.3, prove that $f(Z)$ has a unique median.
2. Prove Corollary 1.3.

One important consequence of the dimension-independence of Theorem 1.2 is that it can be leveraged to prove results for infinite-dimensional Gaussian measures. Here is one example (a special case of a more general inequality due to Borell):

**Corollary 1.5.** Let $B_t$ be a Brownian motion on $[0,1]$, and let $M$ be the median of $\sup_{t \in [0,1]} B_t$. Then for any $s > 0$,

$$\Pr \left( \sup_{t \in [0,1]} B_t > M + s \right) \leq \exp(-s^2/2).$$

**Exercise 4.** Prove Corollary 1.5. (Hint: first figure out what happens in Theorem 1.2 when $\gamma$ has a different covariance.)
1.2 Functional inequalities and gradient bounds

Theorem 1.2 is, at first glance, a statement about sets and their boundaries. However, there is a well-established connection between isoperimetric-type inequalities for sets and gradient lower bounds for functions. The “usual” machinery for this connection (which we will see a little of later) goes through the co-area formula, but we will start by exploring a Gaussian-specific tensorization trick.

Theorem 1.6. Define $I : [0, 1] \to \mathbb{R}^+$ by

$$I(x) = \varphi(\Phi^{-1}(x)) \quad \text{extended continuously to } I(0) = I(1) = 0.$$  
For every Lipschitz function $f : \mathbb{R}^n \to [0, 1]$,

$$I(\mathbb{E}f) \leq \mathbb{E}\sqrt{I^2(f) + |\nabla f|^2}. \quad (2)$$

Theorem 1.6 is due to Bobkov [1], who gave a direct proof of Theorem 1.6 and used it to derive Theorem 1.2. The two theorems turn out to be equivalent, and we will prove Theorem 1.6 assuming Theorem 1.2.

By the way, we have stated Theorem 1.6 for Lipschitz functions, but by straightforward approximation arguments it also holds for weakly differentiable functions.

Proof. Fix a Lipschitz function $f : \mathbb{R}^n \to (0, 1)$. We will define a set $A \subset \mathbb{R}^{n+1}$ by

$$A = \{(x, x_{n+1}) \in \mathbb{R}^n \times \mathbb{R} : x_{n+1} \leq \Phi^{-1}(f(x))\}.$$  
By Fubini’s theorem,

$$\gamma_{n+1}(A) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^1} 1_{\{(x,x_{n+1}) \in A\}} \, d\gamma_1(x_{n+1}) \, d\gamma_n(x)$$
$$= \int_{\mathbb{R}^n} \int_{-\Phi^{-1}(f(x))}^{\Phi^{-1}(f(x))} 1 \, d\gamma_1(x_{n+1}) \, d\gamma_n(x)$$
$$= \int_{\mathbb{R}^n} f(x) \, d\gamma_1(x_{n+1}) \, d\gamma_n(x)$$
$$= \mathbb{E}f. \quad (3)$$

Moreover, we can compute the Gaussian surface area of $A$ using the area formula: recall that if the hypersurface $M \subset \mathbb{R}^{n+1}$ is the graph of a $C^1$ function $h : \mathbb{R}^n \to \mathbb{R}$ then

$$\int_M u(p) \, d\text{Vol}_M(p) = \int_{\mathbb{R}^n} u(x, h(x)) \sqrt{1 + |\nabla h(x)|^2} \, dx.$$  
In particular, taking $u$ to be the density of $\gamma_{n+1}$ leads to

$$\gamma^+(A) = \int_{\partial A} u(p) \, d\text{Vol}_{\partial A}(p)$$
$$= \int_{\mathbb{R}^n} u(x, \Phi^{-1}(f(x))) \sqrt{1 + |\nabla(\Phi^{-1} \circ f)(x)|^2} \, dx. \quad (3)$$
Thanks to the product structure of the Gaussian measure, we have

\[ u(x, y) = \frac{d\gamma_n(x)}{dx} \phi(y), \]

and thanks to the chain rule, we have \( \nabla (\Phi^{-1} \circ f) = \frac{\nabla f}{\sqrt{\Phi^{-1}(f)}} = \frac{\nabla f}{\gamma_n}. \) Applying these two observations to (3) yields

\[ \gamma^+(A) = \int_{\mathbb{R}^n} I(f(x)) \sqrt{1 + \frac{|\nabla f(x)|^2}{I^2(f(x))}} \, d\gamma_n(x) \]

\[ = \mathbb{E} \sqrt{I^2(f) + |\nabla f|^2}. \]

We complete the proof (for \( f : \mathbb{R}^n \to (0, 1) \)) by applying Theorem 1.2 (in \( \mathbb{R}^{n+1} \)) to \( A \):

\[ I(\mathbb{E}f) = I(\gamma(A)) \leq \gamma^+(A) = \mathbb{E} \sqrt{I^2(f) + |\nabla f|^2}. \]

It only remains to consider functions \( f \) that can take values on the closed interval \([0, 1]\). For such a function \( f \), take \( \epsilon > 0 \) and consider \( f_\epsilon = (1 - 2\epsilon) f + \epsilon. \) Then \( f_\epsilon \) takes values in \((0, 1)\), and so

\[ I(\mathbb{E}f_\epsilon) \leq \mathbb{E} \sqrt{I^2(f_\epsilon) + |\nabla f_\epsilon|^2}. \]

Now the point is just to take \( \epsilon \to 0 \) on both sides. On the left, \( I \) is continuous and so \( I(\mathbb{E}f_\epsilon) = I(\epsilon + (1 - 2\epsilon)\mathbb{E}f) \to I(\mathbb{E}f) \) as \( \epsilon \to 0 \). On the right, \( I(f_\epsilon) \to I(f) \) pointwise, and \( \nabla f_\epsilon = (1 - 2\epsilon) \nabla f \to \nabla f \) pointwise. Since \( f \) is Lipschitz and \( I \) is bounded, the dominated convergence theorem implies that

\[ \mathbb{E} \sqrt{I^2(f_\epsilon) + |\nabla f_\epsilon|^2} \to \mathbb{E} \sqrt{I^2(f) + |\nabla f|^2} \]

as \( \epsilon \to 0. \)

We saw how to derive Theorem 1.6 in \( \mathbb{R}^n \) from Theorem 1.2 in \( \mathbb{R}^{n+1} \). On the other hand, by taking functions to approximate the indicator function of a set, it is straightforward to show that Theorem 1.6 in \( \mathbb{R}^n \) implies Theorem 1.2 in \( \mathbb{R}^n \). Indeed, fix a set \( A \subset \mathbb{R}^n \) and (for \( \epsilon > 0 \)) consider the Lipschitz function

\[ f_\epsilon(x) = \begin{cases} 
1 & \text{if } x \in A \\
1 - \frac{1}{\epsilon}d(x, A) & \text{if } d(x, A) < \epsilon \\
0 & \text{otherwise.}
\end{cases} \]

Note that \( f_\epsilon \) converges monotonically to \( 1_A \) (with \( A \) denoting the closure of \( A \)), and hence \( \mathbb{E}f_\epsilon \to \gamma(\overline{A}) \) as \( \epsilon \to 0 \). Assuming that \( \gamma^+(A) < \infty \) (if not, (1) is trivial), \( \gamma(\overline{A}) = \gamma(A) \) and so \( \mathbb{E}f_\epsilon \to \gamma(A) \). On the other hand, the dominated convergence theorem implies that \( \mathbb{E}I(f_\epsilon) \to 0 \) (because \( I(f_\epsilon) \to 0 \) pointwise).
Finally, \(|\nabla f| = 0\) a.e. on both \(A\) and \((A + \epsilon B_2^n)^c\) and is bounded a.e. by \(\frac{1}{\epsilon}\) elsewhere, and so

\[
\mathbb{E}|\nabla f| \leq \mathbb{E}[1_{(A + \epsilon B_2^n)\setminus A}] |\nabla f|_n \leq \frac{1}{\epsilon}(\gamma(A + \epsilon B_2^n) - \gamma(A)).
\]

Applying Theorem 1.6, we have

\[
I(\gamma(A)) = \liminf_{\epsilon \to 0} I(\mathbb{E} f) \leq \liminf_{\epsilon \to 0} \mathbb{E} \sqrt{T^2(f, \epsilon) + |\nabla f|^2}
\]

\[
\leq \liminf_{\epsilon \to 0} (\mathbb{E} I(f, \epsilon) + \mathbb{E} |\nabla f|) = \gamma^+(A).
\]

We have shown that Theorem 1.6 in \(\mathbb{R}^n\) implies Theorem 1.2 in \(\mathbb{R}^n\) which in turn implies Theorem 1.6 in \(\mathbb{R}^{n-1}\). To close this cycle of implications, we will show a remarkable tensorization property satisfied by (2): if it holds in \(\mathbb{R}\) and in \(\mathbb{R}^{n-1}\) then, it holds in \(\mathbb{R}^n\) (and so by induction, it suffices to prove Theorem 1.6 in \(\mathbb{R}\)). To see this, take a Lipschitz function \(f : \mathbb{R}^n \to [0, 1]\). Let \(Z \sim \gamma_n\), and write \(Z_i\) for the \(i\)th component of \(Z\). By first applying Bobkov’s inequality to the function \((\text{of one variable}) \mathbb{E}[f \mid Z_n]\) and then applying it again for fixed \(Z_n\) to the function \((Z_1, \ldots, Z_{n-1}) \mapsto f(Z)\), we obtain

\[
I(\mathbb{E} f) = I(\mathbb{E}[f \mid Z_n])
\]

\[
\leq \mathbb{E} \sqrt{T^2(\mathbb{E} f \mid Z_n)^2 + |\nabla \mathbb{E} f \mid Z_n|^2}
\]

\[
\leq \mathbb{E} \sqrt{\left[ T^2(\mathbb{E} f(Z)) + |\Pi_{\mathbb{R}^{n-1}} \nabla f(Z)|^2 \right] + |\nabla \mathbb{E} f \mid Z_n|^2}.
\]

Applying Jensen’s inequality in the form \(\sqrt{(\mathbb{E} X)^2 + (\mathbb{E} Y)^2} \leq \mathbb{E} \sqrt{X^2 + Y^2}\), we obtain

\[
I(\mathbb{E} f) \leq \mathbb{E} \sqrt{T^2(f) + |\Pi_{\mathbb{R}^{n-1}} \nabla f|^2 + |\partial_n f|^2} = \mathbb{E} \sqrt{T^2(f(X)) + |\nabla f(X)|^2},
\]

which is Bobkov’s inequality in \(\mathbb{R}^n\).

### 1.3 Gradient bounds

There is a slight resemblance between the inequality of Theorem 1.6 and the family of Gagliardo-Nirenberg-Sobolev inequalities that are so useful in PDE and analysis: for every compactly supported and differentiable \(f : \mathbb{R}^n \to \mathbb{R}\),

\[
||f||_{L^{(n/\lambda_n)}(\mathbb{R}^n, \lambda^n)} \leq C(n) ||\nabla f||_{L^1(\mathbb{R}^n, \lambda^n)}, \tag{4}
\]

where \(C(n)\) is a constant depending only on \(n\) and \(\lambda^n\) is the Lebesgue measure on \(\mathbb{R}^n\). Like (4), Theorem 1.6 gives some sort of an upper bound on the values that a function can take in terms of the size of its gradient (and if you’re worried about the fact that the values of \(f\) appear in both sides of Theorem 1.6, apply the inequality \(\sqrt{a^2 + b^2} \leq a + b\) and rearrange to obtain \(I(\mathbb{E} f) - \mathbb{E} I(f) \leq \mathbb{E} |\nabla f|\)).

There is, however, one nice interpretation of (4) that doesn’t apply to Theorem 1.6: (4) says that integrality of the gradient implies more integrability of over the whole space.
the function, but in Theorem 1.6 we already assumed boundedness of \( f \). It turns out, however, that there is a corollary of Theorem 1.6 having an interpretation like this. It’s known as the Gaussian log-Sobolev inequality:

**Theorem 1.7.** For any Lipschitz \( f : \mathbb{R}^n \to \mathbb{R} \),

\[
E[f^2 \log f^2] - E[f^2] \log E[f^2] \leq E[|\nabla f|^2].
\]

In particular, square-integrability of \( \nabla f \) under the Gaussian measure does imply slightly-better-than-square-integrability of \( f \). It is interesting to note, particularly in comparison to (4), that Theorem 1.7 cannot be improved very much.

**Exercise 5.**

1. Find a sequence of compactly supported Lipschitz functions \( f_i : \mathbb{R} \to \mathbb{R} \) such that for every \( 1 \leq p \leq \infty \),

\[
\frac{\|f_i\|_{L^p(\mathbb{R}, \gamma)}}{\|\nabla f_i\|_{L^2(\mathbb{R}, \gamma)}} \to \infty.
\]

(The point here is to realize that having compact support in \((\mathbb{R}^n, \gamma^n)\) is not nearly as restrictive as having compact support in \((\mathbb{R}^n, \lambda^n)\).)

2. For every \( 2 < p \leq \infty \), find a sequence of Lipschitz functions \( f_i : \mathbb{R} \to \mathbb{R} \) such that

\[
\frac{(E[f_i]^p - (E[f_i])^p)^{1/p}}{E[|\nabla f_i|^2]} \to \infty.
\]

(The point here is to realize that we can’t hope for a polynomial increase in integrability.)

**Exercise 6.** Use Bobkov’s inequality to prove Theorem 1.7. (Hint: first, consider a bounded function \( f \), and apply Bobkov’s inequality to \( \epsilon f^2 \) for sufficiently small \( \epsilon > 0 \). The following Lemma has a potentially helpful asymptotic estimate for \( I \) near zero.)

**Lemma 1.8.** As \( \epsilon \to 0 \), \( I(\epsilon) = (1 + o(1))\epsilon \sqrt{2 \log(1/\epsilon)} \).

**Proof.** We first recall the classical inequality: for \( x > 0 \),

\[
\left(\frac{1}{x} - \frac{1}{x^3}\right)\phi(x) \leq \Phi(-x) \leq \frac{1}{x}\phi(x)
\]

or, in other words, \( \Phi(-x) = (1+o(1))\phi(x)/x \) as \( x \to \infty \). To find the asymptotics of \( \Phi^{-1} \), note that if we set \( x = \sqrt{2(1 + \eta) \log \frac{1}{\epsilon}} \) in \( \Phi(-x) = (1 + o(1))\phi(x)/x \), we obtain

\[
\Phi(-x) = (1 + o(1)) \frac{y^{1+\eta}}{\sqrt{4\pi (1 + \eta) \log \frac{1}{\epsilon}}} = (1 + o(1))y^{1+\eta+o(1)}.\]

6
In particular, as $y \to 0$ there is some $\eta(y) \to 0$ such that $\Phi(\sqrt{2(1 + \eta) \log \frac{1}{y}}) = y.

Or in other words,

$$\Phi^{-1}(y) = (1 + o(1)) \sqrt{2 \log y}.$$  

as $y \to 0$.

On the other hand, if we set $x = -\Phi^{-1}(y)$ and rearrange $\Phi(-x)(1 + o(1))\phi(x)/x$, we obtain

$$\phi(\Phi^{-1}(y)) = -(1 + o(1))y\Phi^{-1}(y)$$

as $y \to 0$. Plugging in our asymptotics for $\Phi^{-1}(y)$ completes the proof. \qed

References

