THE CONCENTRATION OF MEASURE

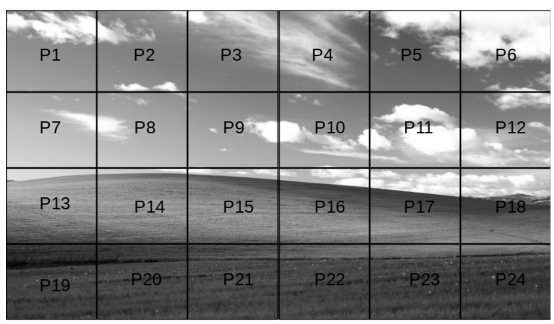
MISHA SWEENEY

INTRODUCTION

Concentration of measure is the study of how some measures concentrate most of their "mass" on a particular set of large measure. The simplest way to understand this is to observe that in some space X with a measure μ that exhibits concentration around a set A, choosing a point randomly in X means that with high probability, the point is going to be very close to, if not in, A. This is counter-intuitive to the notion of randomness and is a very powerful result with many applications.

I was introduced to the idea of concentration of measure by Dr. Liviu Nicolaescu via one such application, when we were discussing potential senior thesis topics. He directed me to a post about compressed sensing in Terence Tao's blog [12], and this post sparked my interest in the field.

To hopefully spark your interest in this field as well, I'll outline the post now. Consider a digital camera, taking a picture of some landscape. The camera has a digital sensor, of some amount of pixels, which records the light and color of a scene. Suppose for simplicity that we are dealing without color, so all the sensor is doing is just recording light. Each pixel is a little square, and the sensor grids out the captured image by a collection of pixels. For example, the following image could be taken by a $4 \times 6 = 24$ pixel sensor, where each pixel is labeled $P\cdot$.



For each image captured, an $N = n \times m$-pixel sensor maps a certain bounded light measurement to each pixel,

$$X : [0, 1] \to (P1, ..., PN),$$

giving each image a $n \times m$ $[0, 1]$-valued matrix representation, or equivalently an nm length vector representation. For N large, this is a lot of information to store, so the solution is to compress the image using a map $\Phi : \mathbb{R}^N \to \mathbb{R}^M$, for $0 < M < N$. This is traditionally done by choosing a map that groups similarly measured pixels together into a block with one value, reducing the number of stored measurements. However, this is very lossy and still not very computationally successful, so other methods are wanted. Here, concentration of measure comes into play for the first time. Namely, it turns out that if you choose a $M \times N$ matrix Φ with each entry an independent, identically distributed standard Gaussian random variable, then with high probability, Φ will preserve distances

Date: Started November 15, 2019. Completed April 24, 2020 . Last revised May 15, 2020.

and thus keep the quality of the image high while drastically reducing the amount of information necessary to store. Additionally, this method is dramatically less lossy than compression by grouping!

The punchline is that Gaussian measure γ_n on $\mathbb{R}^n$ *concentrates* on some set of large measure, say $A \subset \mathbb{R}^n$. Thus, if you choose a point x randomly in $\mathbb{R}^n$, in the point of view of the Gaussian measure, with very large probability x will be in or near A. As you will see, we can translate this to a result for "sufficiently regular", i.e. Lipschitz, maps $F : \mathbb{R}^n \to \mathbb{R}$. Then, concentration of measure states that such a function F differs from some average value only on a set of very small measure. For example, in the point of view of the Gaussian measure γ_n on $\mathbb{R}^n$, the standard Euclidean norm concentrates around the value $\sqrt{n}$. Thus, if you choose a vector $x \in \mathbb{R}^n$ randomly with respect to the Gaussian measure and compute its norm, it will be, with high probability, in some small $\epsilon > 0$ neighborhood of $\sqrt{n}$.

As you can imagine, this is a very powerful result, and it has many consequences and applications, such as the high-fidelity compression I mentioned previously. However, as I researched concentration of measure, what I found more interesting than its applications was its prevalence in drastically different areas of mathematics; the same concentration of measure results can be arrived at through geometric, functional analytic, and probabilistic methods (among others which we do not discuss).

We are most interested in the concentration of Gaussian measure, although we prove concentration of measure results for a much larger class of measures. To demonstrate a few of the most important methods to arrive at concentration of measure, this thesis will loosely follow the structure of Michel Ledoux's monograph *The Concentration of Measure Phenomenon* [6], although discussing a number of other sources, most notably Bakry, Gentil, and Ledoux's book *The Analysis and Geometry of Markov Diffusion Operators* [3] for semigroup and spectral related results.

We begin as Ledoux does and define a *concentration function*. We then define *normal* and *exponential* concentration of measure by bounding the concentration function by either Ce^{-cr^2} or Ce^{-er} for some positive $C, c > 0$. The remainder of Section 1 describes a way to arrive at normal concentration: Laplace bounds, and a way to arrive at exponential concentration: expansion coefficients.

Section 2 will show how geometric methods can imply concentration of measure, and will ultimately arrive at proving normal concentration for the Gaussian measure on $\mathbb{R}^n$. We define *boundary measure* for spaces and use these measures to discuss *isoperimetry*, which looks for sets with minimal boundary measure for a set volume. We finally show how a lower bound on isoperimetry implies concentration of measure.

Continuing, Section 3 turns instead to probabilistic methods of proving concentration of measure. There, we define *semigroups* of bounded linear operators on Banach spaces, along with their generators. We develop a theory of *Markov diffusions* and show how curvature and uniform convextity assumptions on Markov diffusions and their semigroups imply concentration of measure.

Section 4 generalizes the results of Section 3 by expanding the class of measures which exhibit concentration to *log-concave* measures. However, here we go through not semigroups, but rather functional-analytic and spectral methods, making heavy use of the *Poincaré inequality*. Although these methods are applicable for a much larger class of measures than the previous sections, they can only lead us to exponential concentration, whereas the previous sections arrive at normal concentration of measure.

Finally, Section 5 proves the *Johnson-Lindenstrauss flattening lemma*, which is essentially equivalent to the motivating example of image compression discussed above. This result is one of many powerful applications of concentration of measure, and its proof will make heavy use of the normal concentration of measure for the Gaussian measure, which we arrive at through various methods throughout our thesis.

I would like to give a special thanks to my advisor Dr. Liviu Nicolaescu. Ever since I meet him in my sophomore year, Dr. Nicolaescu has been a source of unbounded support, advice, and encouragement. I have grown so much as a mathematician because of his guidance; I genuinely do not know where I would be otherwise. I owe thanks as well to the entire Notre Dame math department, whose dedication to their students and prowess in mentoring continues to astound and impress me.

CONTENTS

1. CONCENTRATION OF MEASURE

We seek to establish concentration of measure results, so it will be good to first define what this is, and second, to establish when we know we have it. In a brief answer to the first question, concentration of measure, when it is present, states that for a certain space and measure, most of the“mass” of the space, in the sense of the equipped measure, is contained in a small neighborhood of a certain “average” set. Concentration of measure is a property of a large number of variables, much like the central limit theorem or the law of large numbers, so the results strengthen in large dimensional spaces.

To quantify the first question, “What is concentration of measure?”, we will develop a concentration function, upon which certain bounds will define the presence of certain types of concentration of measure phenomenon. To answer the second question, “When do we have concentration of measure?”, the remainder of the section will be devoted to developing Laplace bounds and expansion coefficients, which are easily computable ways to prove existence of concentration of measure. Unless otherwise mentioned, proofs follow the approach of [6, Chapter 1].

1.1. **Concentration Functions and Deviation Inequalities.** We first define the space on which we will work to develop concentration of measure, along with a convenient technical definition.

Definition 1.1. For us a *metric measured space* or (m.m.s for brevity) is a triplet (X, d, μ) where (X, d) is a metric space and μ is a Borel *probability* measure on X. Given a subset A of a metric space (X, d) we define its *radius r tube* to be the set

$$A_r := \{ x \in X; \operatorname{dist}(x, A) < r \}$$ □

Now we define a concentration function, which is how we will detect and quantify concentration of measure.

Definition 1.2. The *concentration function* α_μ of a m.m.s. (X, d, μ) is the function $\alpha_\mu : [0, \infty) \to [0, 1]$ defined by

$$\begin{aligned} \alpha_\mu(r) := &\sup \Big\{ 1 - \mu(A_r);\ A \subset X, \mu(A) \geq \frac{1}{2} \Big\}, \\ = &\sup \Big\{ \mu(X \setminus A_r);\ A \subset X, \mu(A) \geq \frac{1}{2} \Big\}. \end{aligned}$$

□

Concentration functions rely two things: a reference measure, μ, and a notion of enlargement, $r > 0$. They are clearly bounded above by $1/2$ and below by 0. As $r \to \infty$, they decrease down to their lower bound of 0. Concentration of measure phenomenon quantifies this decay.

Definition 1.3. Let (X, d, μ) be a m.m.s. with concentration function $\alpha_\mu(r)$.

(i) We say that μ has *normal concentration* if $\exists\, C, c > 0$ such that

$$\alpha_\mu(r) \leq Ce^{-cr^2}.$$

(ii) We say that μ has *expontential concentration* if $\exists\, C, c > 0$ such that

$$\alpha_\mu(r) \leq Ce^{-cr}.$$ □

If μ exhibits one of these modes of concentration, then $\alpha_\mu(r)$ decreases to 0 *rapidly* as $r \to \infty$. Thus, for some not very large r, there is a set $A \subset X$ such that A has large measure ($\mu(A) \geq 1/2$) and the measure of the complement of A's r-tube, A_r^c, is very small. Thus measure is concentrated on some certain large-measure set A. Note, concentration functions do not indicate on which sets the measure concentrates on, only that, if bounded, measure concentrates.

A fascinating and critical result concerning concentration of measure is if a measure μ on a space (X, d) exhibits concentration, then "sufficiently regular" functions $F : X \to \mathbb{R}$ concentrate around a constant value, typically the mean/median. We formalize this discussion below.

Definition 1.4. If (X, d, μ) is a m.m.s., and $F : X \to \mathbb{R}$ is μ-measurable, then we say that $m_F \in \mathbb{R}$ is a *median* of F with respect to μ if

$$\mu\big(\{F \leq m_F\} \big) \geq \frac{1}{2} \text{ and } \mu\big(\{F \geq m_F\} \big) \geq \frac{1}{2}.$$

If F is continuous, then its *modulus of continuity* is the function

$$w_F : (0, \infty) \to [0, \infty),$$

defined by

$$w_F(\eta) := \sup \{ |F(x) - F(y)| \,:\, d(x, y) < \eta \}, \quad \eta > 0.$$ □

Proposition 1.5. *Let (X, d, μ) be a m.m.s. and let $F : X \to \mathbb{R}$ be a continuous function with modulus of continuity w_F. Then, for any median of F, m_F, and $\forall\, \eta > 0$, we have*

$$\mu(\{|F - m_F| > w_F(\eta)\}) \leq 2\alpha_\mu(\eta).$$

Proof. Let $A := \{F \leq m_F\}$ and fix $y \in A$. Observe that for $x \in X$, if $d(x, y) < \eta$, then $F(x) - F(y) \leq w_F(\eta)$, and so

$$F(x) \leq F(y) + w_F(\eta) \leq m_F + w_F(\eta),$$

as $y \in A$. By the definition of the median, $\mu(A) \geq 1/2$, so we can consider A in the supremum used in the definition of α_μ. Thus, if

$$x \in \{\, F > m_F + w_F(\eta) \,\},$$

then

$$d(x, y) > \eta \ \ \forall y \in A,$$

so that $x \in A_\eta^c$. It follows that

$$\mu(\{F > m_F + w_F(\eta)\}) \leq \mu(A_\eta^c) \leq \alpha_\mu(\eta).$$

Performing the same argument using $A := \{F \geq m_F\}$ yields the similar result

$$\mu(\{F < m_F - w_F(\eta)\}) \leq \alpha_\mu(\eta),$$

so combining the two results yields

$$\mu(\{|F - m_F| > w_F(\eta)\}) \leq 2\alpha_\mu(\eta),$$

as desired. □

We call these concentration results for functions *deviation inequalities*. In practice, however, we will be more interested in deviation inequalities for Lipschitz functions rather than simply continuous functions, so we now recall the definition of a Lipschitz function and reformulate the previous result for these functions.

Definition 1.6. A function $F : X \to \mathbb{R}$ on a metric space (X, d) is said to be *K-Lipschitz* if

$$||F||_{\text{Lip}} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} \leq K < \infty.$$

We denote by $\text{Lip}_K = \text{Lip}_K(X)$ the space of K-Lipschitz functions on X and by Lip_K^b the subset of Lip_K consisting of bounded functions. □

As stated above, we can prove a similar result to Proposition 1.5 for Lipschitz functions, which we will use much more often in our later examples.

Proposition 1.7. *Let (X, d, μ) be a m.m.s., and let $F : X \to \mathbb{R}$ be a K-Lipschitz function. Then, $\forall\, r > 0$ and for any median of F, m_F, we have*

$$\mu(\{|F - m_F| > r\}) \leq 2\alpha_\mu(r/K).$$

Proof. Define $A := \{F \leq m_F\}$. For $x \in X$, whenever $d(x, y) < \eta$, for some $y \in A$, we have, by the definition of a K-Lipschitz function,

$$F(x) - F(y) \leq Kd(x, y) \leq K\eta.$$

Thus,

$$F(x) \leq F(y) + K\eta \leq m_F + K\eta.$$

Just like in the proof of Proposition 1.5, by passing to complements and appealing to the concentration function, we attain

$$\mu(\{F - m_F \geq K\eta\}) \leq \alpha_\mu(\eta).$$

Performing the substitution $r = K\eta$ yields

$$\mu(\{F - m_F \geq r\}) \leq \alpha_\mu(r/K),$$

and a similar argument for $A := \{F \geq m_F\}$ concludes the proof. □

While the previous proposition showed that concentration of measure implies deviation inequalities, the following proposition shows that deviation inequalities for *bounded* 1-Lipschitz functions around their means lead to estimates on the concentration function $\alpha_\mu(r)$. Essentially, finding a deviation inequality allows us to establish concentration of measure. This is our first result which shows how to establish concentration of measure, and we will use it repeatedly.

Proposition 1.8. *Let (X, d, μ) be a m.m.s. Assume there exists a nonincreasing function*

$$\phi : [0, \infty) \to [0, \infty)$$

such that for any bounded 1-Lipschitz function $F : X \to \mathbb{R}$, we have

$$\mu\left(\left\{ F \geq \int F d\mu + r \right\}\right) \leq \phi(r) \quad \forall\, r \geq 0. \tag{1.1}$$

Then, for every Borel set $A \subset X$ such that $\mu(A) > 0$, and for every $r > 0$,

$$1 - \mu(A_r) \leq \phi(\mu(A)r).$$

In particular,

$$\alpha_\mu(r) \leq \phi(r/2), \quad r > 0.$$

Proof. Let $A \subset X$ with $\mu(A) > 0$ and fix $r > 0$. Define

$$F : X \to \mathbb{R}, \;\; F(x) := \min(d(x, A), r),$$

Observing that the functions

$$x \mapsto d_A(x) = d(x, A) \text{ and } \; t \mapsto u(t) = \min(t, r)$$

are 1-Lipschitz we deduce that their composition $F = u \circ d_A$ is also 1-Lipschitz. Further,

$$\int_X F d\mu = \underbrace{\int_A F d\mu}_{=0} + \int_{A^c} F d\mu \leq \mu(A^c) r = \big(1 - \mu(A)\big) r. \tag{1.2}$$

Then, by the definition of F,

$$1 - \mu(A_r) = \mu\left(\{F \geq r\}\right) = \mu\left(\{F \geq \mu(A)r + (1 - \mu(A)r)\}\right)$$
$$\overset{(1.2)}{\leq} \mu\left(\left\{F \geq \int F d\mu + \mu(A) r\right\}\right) \overset{(1.1)}{\leq} \phi(\mu(A)r).$$

This proves the first result. Observe next that if $\mu(A) \geq 1/2$, then

$$\phi\big(\,\mu(A)r\,\big) \leq \alpha(r/2),$$

since ϕ is nonincreasing. We deduce that

$$\alpha_\mu(r) \leq \phi(r/2).$$

This proves the second claim. □

The next result generalizes the previous proposition to un-bounded 1-Lipschitz functions. Both results extend to general K-Lipschitz functions by homogeneity: if F is K-Lipschitz, then $(1/K)F$ is a 1-Lipschitz function.

Proposition 1.9. *Let (X, d, μ) be a m.m.s. Assume that ϕ is a continuous, non-negative function on $\mathbb{R}_+$ such that*

$$\lim_{r\to\infty} \phi(r) = 0,$$

and for any bounded 1-Lipschitz function $F : X \to \mathbb{R}$,

$$\mu\left(\left\{ F \geq \int F d\mu + r \right\}\right) \leq \phi(r) \quad \forall\, r \geq 0. \tag{1.3}$$

.

Then, any 1-Lipschitz function $G : X \to \mathbb{R}$, bounded or not, is integrable with respect to μ. If additionally the function $\phi(r)$ is continuous, then G satisfies (1.1) as well.

Proof. Let $G : X \to \mathbb{R}$ be a potentially unbounded 1-Lipschitz function. For any $n \geq 0$, define a function

$$G_n : X \to \mathbb{R}, \;\; G_n(x) = \min(|G(x)|, n).$$

The function G_n is 1-Lipschitz as a composition of 1-Lipschitz maps. So, by our assumptions applied to $-G_n$, $\forall\, r > 0$,

$$\mu\left(\left\{G_n \leq \int G_n\, d\mu - r\right\}\right) \leq \phi(r) \quad \forall\, r \geq 0. \tag{1.4}$$

Next, choose an $M \in \mathbb{R}$ such that $\mu\big(\, |G| \leq M \,\big) \geq 1/2$. This is possible as G is Lipschitz and thus continuous, and thus cannot be infinite everywhere. Further, choose an r_0 such that $\phi(r_0) < 1/2$. This is possible as α is decreasing to 0 as $r \to \infty$. Then

$$\forall\, n, \;\; \mu(\{G_n \leq M\}) \geq \frac{1}{2}.$$

Indeed, $G_n \leq |G|$, so

$$\big\{|G| \leq M\,\big\} \subset \big\{\, G_n \leq M\,\big\}.$$

Plugging $r = r_0$ into (1.1) gets us the system of inequalities:

$$\mu(\{G_n \leq M\}) \geq \frac{1}{2}, \tag{1.5a}$$

$$\mu\left(\left\{G_n \leq \int G_n\, d\mu - r_0\right\}\right) \leq \phi(r_0) < \frac{1}{2}. \tag{1.5b}$$

We deduce that

$$\int G_n d\mu \leq M + r_0. \tag{1.6}$$

Indeed, note that

$$\mu\big(\,\{G_n < c_1\}\,\big) < \mu\big(\,\{G_n < c_2\}\,\big) \implies c_1 < c_2.$$

Since

$$\mu\left(\left\{G_n \leq \int G_n\, d\mu - r_0\right\}\right) \overset{(1.5b)}{<} \frac{1}{2} \leq \mu(\{G_n \leq M\})$$

we deduce that

$$\int G_n\, d\mu - r_0 < M.$$

As $G_n \nearrow |G|$ as $n \to \infty$, each G_n is pointwise non decreasing, so we can use the monotone convergence theorem on G_n to obtain

$$\int G d\mu = \int \lim_{n\to\infty} G_n d\mu = \lim_{n\to\infty} \int G_n d\mu \overset{(1.6)}{\leq} M + r_0.$$

Hence G is integrable. Next, define

$$H_n := \min(\max(G, -n), n).$$

Each H_n is 1-Lipschitz as it is the composition of 1-Lipschitz functions and is clearly bounded. Thus we apply (1.1) to H_n to obtain

$$\mu\left(\left\{H_n \geq \int H_n d\mu + r\right\}\right) \leq \phi(r) \quad \forall\, r > 0,\ n \in \mathbb{N}.$$

Set

$$\bar{H}_n := \int H_n d\mu,\ \ \bar{G} := \int G d\mu.$$

Since $|H_n| \leq |G|$ and $H_n(x) \to G(x), \forall x \in X$, we deduce from the dominated convergence theorem that $\bar{H}_n \to G$. Fix $\varepsilon > 0$. We define the sets

$$A_n := \big\{H_n > \bar{H}_n + r - \varepsilon\big\}, \quad A = \big\{G > \bar{G} + r - \varepsilon\big\}.$$

The sets A_n and A are open. Since $\lim_{n\to\infty} H_n = G$, we conclude that for any $x \in A$ there exists $N = N(x)$ such that $x \in A_n, \forall n > N$. Thus

$$\lim_{n\to\infty} \boldsymbol{I}_{A_n}(x) = \boldsymbol{I}_A(x) = 1,\ \ \forall x \in A.$$

We deduce that

$$\boldsymbol{I}_A(x) \leq \liminf_{n\to\infty} \boldsymbol{I}_{A_n}(x),\ \ \forall x \in X.$$

We conclude that

$$\mu\big(\,A\,\big) = \int_X \boldsymbol{I}_A(x) d\mu \leq \int_X \liminf_{n\to\infty} \boldsymbol{I}_{A_n}(x) d\mu$$

(Fatou's lemma)

$$\leq \liminf_{n\to\infty} \int_X \boldsymbol{I}_{A_n}(x) d\mu = \liminf_{n\to\infty} \mu(A_n) \leq \phi(r+\varepsilon).$$

In other words,

$$\mu\big(\,\big\{G > \bar{G} + r - \varepsilon\big\}\,\big) \leq \phi(r-\varepsilon),\ \ \forall \varepsilon > 0. \tag{1.7}$$

The family $\big\{G > \bar{G} + r - \varepsilon\big\}$ decreases as $\varepsilon \searrow 0$ and

$$\bigcap_{\varepsilon>0} \big\{G > \bar{G} + r - \varepsilon\big\} = \big\{G \geq \bar{G} + r\big\}.$$

Letting $\varepsilon \searrow 0$ in (1.7) and using the continuity of $\phi(r)$ we deduce

$$\mu\big(\,\big\{G \geq \bar{G} + r\big\}\,\big) \leq \phi(r).$$

□

Motivated by the above results, to any m.m.s. (X, d, μ) we associate the space Φ_μ consisting of non-increasing functions $\phi : (0, \infty) \to (0, \infty)$ such that

$$\lim_{r\to\infty} \phi(r) = 0,$$

and

$$\sup_{F \in \mathrm{Lip}_1^b} \mu\big(\{ F \geq \bar{F} + r \} \big) \leq \phi(r), \quad \bar{F} := \int_X F d\mu, \quad \forall r > 0.$$

We denote by Φ_μ^c the subset of Φ_μ consisting of continuous functions. Proposition 1.8 and 1.9 imply that

$$\alpha_\mu(r) \leq \phi(r/2), \quad \forall r > 0, \quad \forall \phi \in \Phi_\mu, \tag{1.8a}$$

$$\mu\big(\{ F \geq \bar{F} + r \} \big) \leq \phi(r), \quad \forall F \in \mathrm{Lip}_1, \quad \forall \phi \in \Phi_\mu^c. \tag{1.8b}$$

This concludes our discussion of concentration functions and deviation inequalities. As we have seen, concentration functions are useful for quantifying the concentration of measure phenomenon and showing the regularity of nice functions, while deviation inequalities are intrinsically tied to concentration of measure. In the next few subsections, we demonstrate a few different ways to arrive at exponential and normal concentration for a m.m.s. by bounding the concentration function.

1.2. **Expansion Coefficients.** An important method for arriving at concentration of measure is through what are called expansion coefficients. A limitation of this method is that it can only show exponential concentration, which we noted was a weaker result than normal concentration. Let us first define an expansion coefficient.

Definition 1.10. Let (X, d, μ) be a m.m.s. Then we define the *expansion coefficient* of μ on (X, d) of order $\epsilon > 0$ as

$$\mathrm{Exp}_\mu(\epsilon) := \inf\{\eta \geq 1 \, : \, \mu(B_\epsilon) \geq \eta \, \mu(B), \, B \subset X, \, \mu(B_\epsilon) \leq 1/2\}. \qquad \square$$

In particular, if $\mathrm{Exp}_\mu(\epsilon) > 1$, then any set $B \subset X$ with $\mu(B_\epsilon) \leq 1/2$ has *very* small measure, as the expansion coefficient shows that just a small ϵ enlargement of B results in much larger measure. If we consider B to be the complement of a set with large measure, then we can picture how expansion coefficients lead to concentration.

The following proposition shows how finding an expansion coefficient allows one to demonstrate exponential concentration.

Proposition 1.11. *Let (X, d, μ) be a m.m.s. If, for some $\epsilon > 0$, $\mathrm{Exp}_\mu(\epsilon) \geq \eta > 1$, then μ has exponential concentration*

$$\alpha_\mu(r) \leq \frac{\eta}{2} e^{-r \log(\eta)/\epsilon}, \qquad r > 0.$$

Proof. Let $A \subset X$ a Borel set such that $\mu(A) \geq 1/2$. Let $B = (A_{k\epsilon})^c$. Then, clearly, $B_{k\epsilon} \subset A^c$ and so $\mu(B_{k\epsilon}) \leq \frac{1}{2}$. Thus we can invoke expansion coefficients. Observe,

$$\mu(B) \leq \frac{\mu(B_{k\epsilon})}{\mathrm{Exp}_\mu(\epsilon)^k} \leq \frac{1}{2} \frac{1}{\mathrm{Exp}_\mu(\epsilon)^k} \leq \frac{1}{2} \eta^{-k}.$$

Thus,

$$1 - \mu(A_{k\epsilon}) \leq \frac{1}{2\eta^k} = \frac{1}{2} e^{-k \log(\eta)}.$$

So if $r = k\epsilon$ for some $k \geq 1$, then

$$\alpha_\mu(r) \leq 1 - \mu(A_r) = 1 - \mu(A_{k\epsilon}) \leq \frac{1}{2} e^{-k \log(\eta)} \leq \frac{\eta}{2} e^{-k\epsilon \log(\eta)/\epsilon},$$

as desired. Recall, $\eta > 1$. Now, assume that $r \in (k\epsilon, (k+1)\epsilon)$. Then clearly $r - k\epsilon \leq 1$ and $\mu(A_r) \geq \mu(A_{k\epsilon})$. So,

$$\alpha_\mu(r) \leq 1 - \mu(A_r) \leq 1 - \mu(A_{k\epsilon}) \leq \frac{1}{2} e^{-k \log(\eta)} = \frac{1}{2} e^{-k\epsilon \log(\eta)/\epsilon}$$

$$= \frac{1}{2}\exp\left[\frac{-r\log(\eta)+(r-k\epsilon)\log(\eta)}{\epsilon}\right]$$

$((r-k\epsilon)/\epsilon < 1)$

$$\leq \frac{1}{2}\exp\left[-r\log(\eta)/\epsilon\right]\exp\left(\log\eta\right) = \frac{\eta}{2}e^{-r\log(\eta)/\epsilon}.$$

as desired. □

1.3. **Laplace Bounds.** In this section, we look at another method of establishing concentration of measure, this time through a functional approach. The results in this section will allow us to arrive at normal concentration, a much stronger result than exponential concentration. We will use this method later on as well. First, however, we define what we will bound the concentration function with.

Definition 1.12. Let (X, d, μ) be a m.m.s. Then, for $\lambda \geq 0$, we define the *Laplace functional* of μ on (X, d) as

$$E_{(X,d,\mu)}(\lambda) = E_\mu(\lambda) = \sup_F \int e^{\lambda F} d\mu,$$

where the supremum is taken over $F \in \mathrm{Lip}_1$ with mean $\bar{F} = 0$. □

The following proposition demonstrates how the Laplace functional bounds the concentration function of a m.m.s. and thus implies normal concentration.

Proposition 1.13. *Let (X, d, μ) be a m.m.s. Then,*

$$\alpha_\mu(r) \leq \inf_{\lambda\geq 0} e^{-\lambda r/2}E_\mu(\lambda), \quad r > 0.$$

In particular, if we have a normal bound in λ on E_μ, then we get normal concentration of μ. More precisely, if

$$E_\mu(\lambda) \leq e^{\lambda^2/2c}, \quad \lambda \geq 0,$$

then $e^{-cr^2/2} \in \Phi_\mu^c$ so that $\mathrm{Lip}_1 \subset L^1(X,\mu)$ *and* $\forall\, r \geq 0$, *we have*

$$\mu\left(\left\{F \geq \int F d\mu + r\right\}\right) \leq e^{-cr^2/2},$$

$$\alpha_\mu(r) \leq e^{-cr^2/8}, \quad r > 0.$$

Proof. Fix $\lambda > 0$. Then for any 1-Lipschitz function $F : X \to \mathbb{R}$,

$$F(x) \geq r \implies \lambda F \geq \lambda r \implies e^{\lambda F} \geq e^{\lambda r}.$$

Observe that $e^{\lambda F}$ is a bounded, positive function on X and $e^{\lambda r}$ is a positive number in $\mathbb{R}$. Thus $e^{\lambda F}$ is integrable, and

$$\int_X e^{\lambda F} d\mu \geq \int_{x:F(x)\geq r} e^{\lambda F} d\mu \geq \mu\left(\{F(x) \geq r\}\right) e^{\lambda r}.$$

And multiplying through by $e^{-\lambda r}$ yields

$$\mu\left(\{F(x) \geq r\}\right) \leq e^{-\lambda r}\int_X e^{\lambda F} d\mu \leq e^{-\lambda r}E_\mu(\lambda) \quad \forall \lambda \geq 0. \tag{1.9}$$

If we define

$$\phi(r) := \inf_{\lambda\geq 0} e^{-\lambda r}E_\mu(\lambda),$$

and consider our above F to be zero-mean as well, then by (1.9),

$$\mu\left(\{F(x) \geq \bar{F} + r\}\right) = \mu\left(\{F \geq r\}\right) \leq e^{-\lambda r} \int_X e^{\lambda F} d\mu \ (\forall \lambda \geq 0) \leq \phi(r).$$

And so, by Proposition 1.8,

$$\alpha_\mu(r) \leq \phi(r/2) = \inf_{\lambda \geq 0} e^{-\lambda r/2} E_\mu(\lambda),$$

which yields the first result. If we assume that

$$E_\mu(\lambda) \leq e^{\lambda^2/2c}, \quad \lambda \geq 0,$$

then $e^{-\lambda r + \lambda^2/2c} \in \Phi_\mu^c$, $\forall \lambda \geq 0$. We deduce that from (1.8b) that $\forall G \in \mathrm{Lip}_1$, we have $G \in L^1(X, \mu)$ and

$$\mu\big(\,\{G \geq \bar{G} + r\}\,\big) \leq e^{-\lambda r + \lambda^2/2c}.$$

To conclude the proof, we optimize in λ. Indeed, $e^{-\lambda r + \lambda^2/2c}$ is smallest when $l(\lambda) := -\lambda r + \lambda^2/2c$ is smallest, which occurs when $l'(\lambda) = 0$, or when $\lambda = rc$. Thus,

$$\mu\big(\,\{G \geq \bar{G} + r\}\,\big) \leq \inf_{\lambda \geq 0} e^{-\lambda r + \lambda^2/2c} = e^{-cr^2/2}.$$

Finally, using our optimized $\phi(r) = e^{-cr^2/2}$, we deduce from (1.8a) that $\alpha_\mu(r) \leq \phi(r/2)$, and so

$$\alpha_\mu(r) \leq \alpha(r/2) \leq e^{-c(r/2)^2/2} = e^{-cr^2/8},$$

as desired. □

1.3.1. *An Example.* We now illustrate the utility of the Laplace functional through an example. The below proposition essentially shows that in the point-of-view of the Gaussian measure, $\mathbb{R}^n$ for large n "looks like" the sphere of radius $\sqrt{n}$; Laplace bounds allow us to arrive at a deviation inequality for the Euclidean norm around the average value $\sqrt{n}$.

Lemma 1.14. *We follow the approach of [11, Prop 2.2]. For any $\delta \geq 0$,*

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq n + \delta\right\}\right) \leq \left(\frac{n}{n+\delta}\right)^{-n/2} e^{-\delta/2},$$

and

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \leq n - \delta\right\}\right) \leq \left(\frac{n}{n-\delta}\right)^{-n/2} e^{\delta/2}.$$

Proof. First fix a $\lambda \in (0, 1)$. Then,

$$||x||^2 \geq n + \delta \implies \lambda||x||^2/2 \geq \lambda(n+\delta)/2,$$

and so

$$e^{\lambda||x||^2/2} \geq e^{\lambda(n+\delta)/2}.$$

Reasoning identically as in (1.9) with $\lambda/2$, we observe that

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:||x||^2 \geq n + \delta\right\}\right) \leq e^{-\lambda(n+\delta)/2} \int_{\mathbb{R}^n} e^{\lambda||x||^2/2} d\gamma_n. \tag{1.10}$$

By a change of variables,

$$e^{-\lambda(n+\delta)/2} \int_{\mathbb{R}^n} e^{\lambda||x||^2/2} d\gamma_n = e^{-\lambda(n+\delta)/2} \int_{\mathbb{R}^n} e^{\frac{\lambda||x||^2}{2}} \frac{e^{-||x||^2/2}}{(2\pi)^{n/2}} dx$$

$$= e^{-\lambda(n+\delta)/2}(2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{\frac{(\lambda-1)||x||^2}{2}}dx.$$

The above integral splits into the product of one dimensional integrals and we deduce that

$$(2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{\frac{(\lambda-1)||x||^2}{2}}dx = \left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{(\lambda-1)x^2/2}dx\right)^n.$$

To calculate the latter integral, we perform the substitution

$$x = y/\sqrt{1-\lambda}, \quad dx = (1-\lambda)^{-1/2}dy.$$

The integral then becomes

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{(\lambda-1)x^2/2}dx = \frac{1}{\sqrt{1-\lambda}}\cdot\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-y^2/2}dy = (1-\lambda)^{-1/2},$$

and so

$$(2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{(\lambda-1)||x||^2/2}dx = (1-\lambda)^{-n/2}.$$

Thus, returning to (1.10),

$$\begin{aligned}\gamma_n\left(\left\{x\in\mathbb{R}^n \,:\, ||x||^2 \geq n+\delta\right\}\right) &\leq e^{-\lambda(n+\delta)/2}(2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{(\lambda-1)||x||^2/2}dx \\ &= e^{-\lambda(n+\delta)/2}(1-\lambda)^{-n/2}.\end{aligned}$$

Choosing $\lambda = \delta/(n+\delta)$ yields the first conclusion. The second conclusion is easily attained by a matter of sign changes in a similar manner to this proof. □

To summarize, we have defined concentration of measure through concentration functions, shown three different methods of arriving at concentration of measure, and used one of these methods to show an example of a deviation inequality.

2. Isoperimetric and Functional Examples

Our first results about concentration of measure for specific measures will come from isoperimetric inequalities, which are statements about extremal sets and surface measure. This approach is inherently geometric in nature, and will actually in some cases answer the question, "On which particular sets does measure concentrate?" Isoperimetric inequalites will lead us to establish normal concentration for both the uniform measure on the sphere and the Gaussian measure on $\mathbb{R}^n$. All results, unless otherwise mentioned, follow from [6, Chap 2].

2.1. **Isoperimetry to concentration.** We begin by defining what an isoperimetric inequality is, which in turn first requires defining a new measure.

Definition 2.1. For a m.m.s. (X, d, μ), the *boundary measure* of a Borel set $A \subset X$ with respect to μ is defined as

$$\mu^+(A) := \liminf_{r\to 0}\frac{1}{r}\mu(A_r \setminus A), \tag{2.1}$$

where we recall that $A_r := \{x \in X\,; d(X, A) < r\}$. □

Remark 2.2. When X is $\mathbb{R}^n$ equipped with the Euclidean distance, μ is the Lebesgue measure, and A is a bounded open set with smooth boundary, then $\mu^+(A)$ is indeed the $(n-1)$-dimensional (Hausdorff) measure of ∂A. □

This boundary measure enters the definition of the isoperimetric function.

Definition 2.3. Let (X, d, μ) be an m.m.s[1]

(i) The *isoperimetric function* of μ is the largest function
$$I_\mu : [0, 1] \to \mathbb{R}$$
such that for every Borel set $A \subset X$,
$$\mu^+(A) \geq I_\mu(\mu(A)). \tag{2.2}$$

(ii) A Borel subset $B \subset X$ is called an *extremal set* if
$$\mu^+(B) = I_\mu(\mu(B)).$$
□

Remark 2.4. (a) The extremal sets of an m.m.s. are the sets with the smallest boundary measure for a given volume. To see this, let $A,\ B \subset X$ such that $\mu(A) = \mu(B)$, but B is an extremal set. Then
$$\mu^+(A) \geq I_\mu(\mu(A)) = I_\mu(\mu(B)) = \mu^+(B),$$
and so $\mu^+(A) \geq \mu^+(B)$. The search for explicit descriptions of extremal sets is one of the main focuses of isoperimetric inequalities, but these descriptions are rarely attainable. However, even in situations where we cannot find explicit descriptions for these extremal sets, we can still say a lot about them using concentration results.

(b) Very few explicit isoperimetric functions are known. Among those that are known, the most notable is example is the sphere $X = \mathbb{S}^n$, equipped with its geodesic metric $d(x, y) = \arccos\langle x, y\rangle$ (assuming the usual inner product on $\mathbb{R}^n$), and the unique rotation invariant probability measure μ. If we write $v(r) :=$ the volume of a geodesic ball of radius r on $\mathbb{S}^n$, then the isoperimetric function for the triple $(\mathbb{S}^n, d, \mu)$ is
$$I_\mu = v' \circ v^{-1}.$$
Further, the extremal sets for this space are geodesic balls, i.e. spherical caps. We will not prove this statement, but will rather prove the concentration implications of this result. □

Let us now define another technical condition upon which we will work.

Definition 2.5. We say that a m.m.s. (X, d, μ) is *convenient* if the following hold.

(i) If A is a finite union of balls, then the $\liminf$ in (2.1) is actually a limit, i.e.,
$$\mu^+(A) = \lim_{r\to 0} \frac{1}{r}\mu(A_r \setminus A) = \lim_{r\to 0} \frac{1}{r}\big(\mu(A_r) - \mu(A)\big).$$

(ii) The measure μ is uniquely determined by its values on finite union of balls. □

For example, a sphere in $\mathbb{R}^n$ equipped with a geodesic distance metric and uniform area measure is a convenient m.m.s. The next result is a comparison result that is integral to establishing a bridge from isoperimetry to concentration of measure.

Proposition 2.6. *Suppose that (X, d, μ) is a convenient m.m.s. and*
$$I_\mu \geq v' \circ v^{-1} \tag{2.3}$$
for some strictly increasing differentiable function[2]
$$v : I \subset \mathbb{R} \to [0, 1].$$

[1]Recall, according to our definition of an m.m.s., μ is a probability measure.

[2]It helps to think of $v(r)$ as the volume of ball of radius r.

Then, for every $r > 0$*,*

$$v^{-1}(\mu(A_r)) \geq v^{-1}(\mu(A)) + r.$$

Proof. Because X is convenient, it suffices to assume that $A \subset X$ is given by a finite union of open balls in X. Fix such an $A \subset X$. We introduce the function $h(r) = v^{-1}(\mu(A_r))$.[3]

Then, as v is strictly increasing and differentiable, by the inverse function theorem,

$$h'(r) = \big(v^{-1}(\mu(A_r))\big)' = \frac{(\mu(A_r)')}{v' \circ v^{-1}(\mu(A_r))}$$
$$= \frac{\mu^+(A_r)}{v' \circ v^{-1}(\mu(A_r))} \overset{(2.2)}{\geq} \frac{I_\mu\big(\mu(A_r)\big)}{v' \circ v^{-1}(\mu(A_r))} \overset{(2.3)}{\geq} 1.$$

Then as $h'(r) \geq 1$,

$$h(r) = h(0) + \int_0^r h'(s)ds \geq h(0) + r.$$

□

Remark 2.7. Conversely, if we assume

$$v^{-1}(\mu(A_r)) \geq v^{-1}(\mu(A)) + r, \quad r > 0,$$

then, for any Borel subset $A \subset X$, we have

$$\mu^+(A) = \liminf_{r\to 0} \frac{1}{r}\mu(A_r \setminus A) = \lim_{r\to 0} \frac{1}{r}\left[v(v^{-1}(\mu(A_r))) - \mu(A)\right]$$
$$\geq \liminf_{r\to 0} \frac{1}{r}\left[v(v^{-1}(\mu(A)) + r) - \mu(A)\right] = v' \circ v^{-1}(\mu(A)).$$ □

Remark 2.8. In constant curvature spaces (i.e. $\mathbb{R}^n$ and $\mathbb{S}^n \subset \mathbb{R}^{n+1}$) equality in Proposition 2.6 is achieved clearly on geodesic balls. Thus, in constant curvature spaces, Proposition 2.6 is equivalent to the statement

$$\mu(A_r) \geq \mu(B_r), \quad r > 0, \tag{2.4}$$

for every $A \subset X$ such that $\mu(A) = \mu(B)$ and $B \subset X$ is a geodesic ball. This is the so-called *isoperimetric inequality* for μ on a constant-curvature space. To see this, observe that under these conditions, and under the assumptions of Proposition 2.6),

$$\mu(A_r) \geq v(v^{-1}(\mu(A)) + r) = v(v^{-1}(\mu(B)) + r) = \mu(B_r), \quad r > 0.$$ □

As a corollary to these results, we can establish the bridge from isoperimetry to concentration of measure. Namely, the result below states that if we have a lower bound on the isoperimetric function of a convenient m.m.s., then the measure on that space has a bound on its concentration function, proving concentration of measure.

Corollary 2.9. *Let* (X, d, μ) *be a convenient m.m.s. Assume that*

$$I_\mu \geq v' \circ v^{-1}$$

for some strictly increasing differentiable function

$$v : I \subset \mathbb{R} \to [0, 1].$$

Then,

$$\alpha_\mu(r) \leq 1 - v\big(v^{-1}(1/2) + r\big), \quad r > 0.$$

[3] In many applications $h(r)$ is the radius of a ball of volume $\mu(A_r)$.

Proof. If $A \subset X$ is a Borel such that $\mu(A) \geq 1/2$, then by Proposition 2.6, we know that

$$v^{-1}(\mu(A_r)) \geq v^{-1}(\mu(A)) + r,$$

and so $\mu(A_r) \geq v\big(v^{-1}(\mu(A)) + r \big)$. Then, by our choice of A,

$$1 - \mu(A_r) \leq 1 - v(v^{-1}(1/2) + r \big).$$

Hence,

$$\alpha_\mu(r) = \sup \big\{ 1 - \mu(A_r);\ A \subset X, \mu(A) \geq 1/2 \big\} \leq 1 - v\big(v^{-1}(1/2) + r \big).$$

□

2.2. **Concentration on the unit sphere.** As an example of this previous result, we establish normal concentration of measure for the uniform measure on $\mathbb{S}^n$.

Theorem 2.10. *For the standard n-sphere $\mathbb{S}^n$, $n \geq 2$, equipped with the geodesic metric d and normalized uniform measure μ,*

$$\alpha_{(\mathbb{S}^n,d,\mu)} \leq \frac{1}{\pi} e^{-(n-1)r^2/2}, \quad 0 < r \leq \pi.$$

Proof. On the unit sphere $\mathbb{S}^n$ the geodesic distance between two points P_1, P_2 is the angle $\measuredangle(P_1 O P_2)$. Using spherical coordinates see [10, Eq.(15.50)] we deduce that volume of the unit ball $B^{n+1} \in \mathbb{R}^{n+1}$ is (see [10, Ex. 15.53 and (15.66)])

$$\boldsymbol{\omega}_{n+1} = \frac{2\pi}{n+1} s_1 \cdots s_n = \frac{\pi^{n/2}}{\Gamma(1 + n/2)},$$

where

$$s_n = \int_0^\pi \sin^{n-1}\theta \, d\theta = 2 \int_0^{\pi/2} \sin^{n-1}\theta \, d\theta.$$

The “area” of the unit sphere $\mathbb{S}^n$ is

$$\boldsymbol{\sigma}_n = (n+1)\boldsymbol{\omega}_{n+1} = 2\pi s_1 \cdots s_n = \frac{2\pi^{(n+1)/2}}{\Gamma\big((n+1)/2 \big)},$$

where $\Gamma(x)$ is Euler’s Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt, \quad x > 0.$$

In particular

$$s_n = \frac{\boldsymbol{\sigma}_n}{\boldsymbol{\sigma}_{n-1}} = \frac{\pi^{1/2}\Gamma(n/2)}{\Gamma\big((n+1)/2 \big)}. \tag{2.5}$$

The volume of the ball of radius $r \in (0, \pi)$ centered at the North Pole is

$$2\pi s_1 \cdots s_{n-1} \int_0^r \sin^{n-1}\theta d\theta.$$

If we use the normalized uniform measure on $\mathbb{S}^n$ we deduce that for every $0 < r \leq \pi$, the normalized volume of the geodesic ball of radius r is

$$v(r) = \frac{2\pi s_1 \cdots s_{n-1}}{\boldsymbol{\sigma}_n} \int_0^r \sin^{n-1}\theta d\theta = s_n^{-1} \int_0^r \sin^{n-1}\theta d\theta.$$

We know that $v^{-1}(1/2) = \pi/2$, and so we evaluate $1 - v(\pi/2 + r)$ for $0 < r \leq \pi/2$. We have

$$1 - v(\pi/2 + r) = 1 - s_n^{-1} \int_0^{(\pi/2)+r} \sin^{n-1}\theta d\theta = s_n^{-1} \int_{(\pi/2)+r}^{\pi} \sin^{n-1}\theta d\theta$$

$$= s_n^{-1} \int_r^{\pi/2} \cos^{n-1}\theta d\theta.$$

We use the inequality

$$\cos(u) \leq 1 - u^2/2 \leq e^{-u^2/2}, \quad 0 \leq u \leq \pi/2,$$

and perform the change of variables $\theta = \tau/\sqrt{n-1}$ to obtain

$$\int_r^{\pi/2} \cos^{n-1}\theta d\theta \leq \int_r^{\pi/2} e^{-(n-1)\theta^2/2} d\theta$$

$$= \frac{1}{\sqrt{n-1}} \int_{r\sqrt{n-1}}^{(\pi/2)\sqrt{n-1}} e^{-\tau^2/2}\, d\tau \leq \frac{1}{\sqrt{n-1}} \int_{r\sqrt{n-1}}^{\infty} e^{-\tau^2/2} d\tau.$$

Let us observe that

$$\int_r^{\infty} e^{-t^2/2} dt \leq \frac{2}{\pi} e^{-r^2/2}, \ \forall r \in (0, \pi/2). \tag{2.6}$$

Indeed, set

$$F(r) = \int_r^{\infty} e^{-t^2/2} dt, \quad \phi(r) = \frac{2}{\pi} e^{-r^2/2}.$$

Indeed

$$F(0) = \sqrt{\frac{\pi}{2}} > \frac{2}{\pi} = \phi(0),$$

and

$$F'(r) - \phi'(r) = e^{-r^2/2}\left(1 - \frac{2r}{\pi}\right) > 0, \quad \forall r \in (0, \pi/2).$$

Hence

$$1 - v(\pi/2 + r) \leq \frac{2}{\pi s_n \sqrt{n-1}} e^{-(n-1)r^2/2}.$$

From (2.5) and the classical identity $\Gamma(x+1) = x\Gamma(x)$, $x > 0$, we deduce that s_n exhibits the recursion formula (see also [10, Eq. (9.43)])

$$s_n = \frac{n-1}{n} s_{n-2}.$$

Hence

$$\sqrt{n-1}s_n = \frac{(n-1)^{3/2}}{n} s_{n-2}.$$

Now observe that

$$\frac{(n-1)^{3/2}}{n} > (n-3)^{1/2}, \quad \forall n \geq 3.$$

We deduce

$$\sqrt{n-1}s_n \geq \sqrt{n-3}s_{n-2}$$

which implies that

$$\sqrt{n-1}s_n \geq 2, \ \forall n \geq 2.$$

Putting all these together we deduce that

$$\alpha_\mu(r) \leq 1 - v(\pi/2 + r) \leq \frac{1}{\pi} e^{-(n-1)r^2/2}.$$

□

2.3. **From spherical concentration to Gaussian concentration.** Now that we have established concentration of measure for the uniform measure on the unit sphere of high dimension, we are able to establish normal concentration for the Gaussian measure on $\mathbb{R}^n$ using the Poincaré Lemma. However, first we must build up some technical machinery and so briefly discuss Beta distributions and some of their properties.

Definition 2.11. The *Gamma function* is the function

$$\Gamma : (0,\infty) \to \mathbb{R}, \quad \Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt,$$

and the *Beta function* is the ratio

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \qquad \square$$

Definition 2.12. We say a random variable X has a *Beta Distribution* with parameters a and b, $X \sim \beta(a,b)$, if its probability distribution is given by

$$\mathbb{P}_X = \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}\, \mathbb{I}_{(0,1)} dx,$$

where dx is the Lebesgue measure on $\mathbb{R}$ and $B(a,b)$ is the above Beta function. $\square$

Lemma 2.13. *Let $\alpha, \beta \in \mathbb{R}$. Then, for all $x > 0$,*

$$\frac{\Gamma(x+\beta)}{\Gamma(x+\alpha)} \sim x^{\beta-\alpha}, \quad \text{as } x \to \infty.$$

Proof. Let $\gamma = \beta - \alpha$. By Stirling's Approximation, as $x \to \infty$ we have

$$\frac{\Gamma(x+1+\beta)}{\Gamma(x+1+\alpha)} \sim \frac{\sqrt{2\pi(x+\beta)}\left(\frac{x+\beta}{e}\right)^{x+\beta}}{\sqrt{2\pi(x+\alpha)}\left(\frac{x+\alpha}{e}\right)^{x+\alpha}}$$

$$= \left(1 + \frac{\gamma}{x+\alpha}\right)^{x+\alpha+1/2} (1+\beta/x)^{\gamma}(x/e)^{\gamma} \sim e^{\gamma}(x/e)^{\gamma} = x^{\gamma}.$$

$\square$

Lemma 2.14. *Let (X_i) be a sequence of independent standard normal random variables. Let $N \in \mathbb{N}$ and define*

$$R_N^2 := X_1^2 + ... + X_N^2.$$

Then

$$\frac{R_n^2}{R_{N+1}^2} \sim \beta(n/2, (N+1-n/2)),$$

where $\beta(a,b)$ is the Beta distribution.

Proof. If $X \sim N(0,1)$, then $X^2 \sim \Gamma(1/2, 1/2)$. It follows then that

$$X_1^2 + ... + X_N^2 \sim \underbrace{\Gamma(1/2,1/2) * \cdots * \Gamma(1/2,1/2)}_{n} \sim \Gamma(n/2, 1/2).$$

If we set

$$U := R_n^2,\ V := \sum_{k=n+1}^{N} X_k,$$

then U and V are independent, with

$$U \sim \Gamma(n/2, 1/2),\ V \sim \Gamma(m/2, 1/2), \quad m = N - n.$$

Write F_U and F_V for the cumulative distribution functions of U and V respectively. Then,

$$\mathbb{P}_U[du] = F_U' du = \rho_U(u)du = \frac{(1/2)^{n/2}}{\Gamma(n/2)} u^{n/2-1} e^{-u/2} du, \quad u > 0,$$
$$\mathbb{P}_V[dv] = F_V' dv = \rho_V(v)dv = \frac{(1/2)^{m/2}}{\Gamma(m/2)} v^{n/2-1} e^{-m/2} du.$$

Set

$$X := \frac{R_n^2}{R_N^2} = \frac{U}{U+V}, \quad r = r(x) := \frac{x}{1-x},$$

so that

$$F_X(x) = \mathbb{P}[X \le x] = \mathbb{P}[U \le x(U+V)] = \mathbb{P}[U \le rV]$$
$$= \int_{\mathbb{R}} \int_{-\infty}^{rv} \mathbb{P}_U[du]\mathbb{P}_V[dv] = \int_{\mathbb{R}} F_U(rv)\rho_V(v)dv.$$

For $x \in (0,1)$, we have

$$\frac{d}{dx} F_X(x) = \int_{\mathbb{R}} vr'(x)\rho_U(rv)\frac{d}{dx}\rho_V(v)[dv]$$

$$= r'(x)\frac{(1/2)^{n/2}(1/2)^{m/2}}{\Gamma(n/2)\Gamma(m/2)} \int_{\mathbb{R}} v(rv)^{n/2-1} e^{-rv/2} v^{m/2-1} e^{-v/2} dv$$

$(N = n + m)$

$$= r'(x) r^{n/2-1} \frac{(1/2)^{N/2}}{\Gamma(n/2)\Gamma(m/2)} \int_{\mathbb{R}} v^{N/2-1} e^{-(r+1)v/2} dv$$

$(t = (r+1)v/2, v = 2t/(r+1))$

$$= r'(x) r^{n/2-1} \frac{(1/2)^{N/2}}{\Gamma(n/2)\Gamma(m/2)} (2/(r+1))^{N/2} \underbrace{\int_{\mathbb{R}} t^{N/2-1} e^{-t} dt}_{=\Gamma(N/2)}$$

$((r'(x)) = 1/(1-x)^2, r + 1 = 1/(1-x))$

$$= \frac{1}{(1-x)^2} \frac{x^{n/2-1}(1-x)^{N/2}}{(1-x)^{n/2-1}} \frac{\Gamma(N/2)}{\Gamma(n/2)\Gamma(m/2)}$$

$$= \frac{\Gamma(N/2)}{\Gamma(n/2)\Gamma(m/2)} x^{n/2-1}(1-x)^{m/2-1},$$

and so

$$X \sim \beta(n/2, m/2).$$

It follows immediately that

$$\frac{R_n^2}{R_{N+1}^2} \sim \beta(n/2, (N + 1 - n/2)),$$

as desired. ∎

We are now able to state Poincaré's lemma, which as we previously mentioned will allow us to establish concentration of measure for the Gaussian measure on $\mathbb{R}^n$. Roughly, the lemma states that uniform measures on n-dimensional spheres of radius $\sqrt{n}$ approximate Gaussian measures. This result lets us use the isoperimetric inequality on $\mathbb{S}^n$ to obtain a Gaussian isoperimetric inequality, which will then provide us with a normal concentration of measure result through Corollary 2.9. To simplify the proof, we first introduce some notation.

- We denote by $\Pi_{N+1,n}$ be the canonical projection from $\mathbb{R}^{N+1}$ to $\mathbb{R}^n$.
- We denote by Σ^N be the N-dimensional sphere of radius $\sqrt{N}$ with uniform probability measure σ^N.
- Let γ_n be the canonical Gaussian measure on $\mathbb{R}^n$

$$\gamma_n(dx) = (2\pi)^{-n/2} \exp(-|x|^2/2)dx.$$

- For every Borel set $A \subset \mathbb{R}^n$ we set

$$A^N := \Pi_{N+1,n}^{-1}(A) \cap \Sigma^N.$$

  In other words, A^N consists of all points in Σ^N that project in A.
- We denote by $\mu_{N,n}$ the pushforward of σ^N via $\Pi_{N+1,n}$. In other words, for every Borel set $A \subset \mathbb{R}^n$ we have

$$\mu_{N,n}(A) = \sigma^N(A^N).$$

Lemma 2.15 (Poincaré). *For any Borel set $A \subset \mathbb{R}^n$ we have*

$$\lim_{N\to\infty} \sigma^N\big(A^N\big) = \gamma_n(A),$$

i.e.,

$$\lim_{N\to\infty} \mu_{N,n}(A) = \gamma_n(A).$$

Proof. We follow the approach of the proof to Lemma 1.2 in [7, p 9]. Let $(X_i)_{i\geq 1}$ be a sequence of independent standard normal random variables. Fix an $N \geq 1$. For any $1 \leq k \leq N$ we define

$$R_k := \sqrt{X_1^2 + ... + X_k^2}.$$

Then the random vector

$$V_N := \frac{\sqrt{N}}{R_{N+1}}(X_1, ..., X_{N+1})$$

is distributed uniformly on Σ^N. In particular, the distribution of the random vector

$$U_N := \frac{\sqrt{N}}{R_{N+1}}(X_1, ..., X_n)$$

is $\mu_{N,n}$.

By our assumptions, the expectation of each X_i^2 is 1, so R_N^2/N converges to 1 almost surely by the strong law of large numbers. Hence,

$$U_N := \frac{\sqrt{N}}{R_{N+1}}(X_1, ..., X_n) \xrightarrow{\text{a.s}} U_\infty := (X_1, ..., X_n) \sim \gamma_n.$$

In particular, U_N converges in distribution to U_∞, i.e., the probability measure $\mu_{N,n}$ converges weakly to γ_n as $N \to \infty$. The lemma however claims strong convergence, namely

$$\lim_{N\to\infty} \mu_{N,n}(A) = \gamma_n(A)$$

for any Borel set $A \subset \mathbb{R}^n$.

To prove this stronger result, we first define the following random variables.

$$D := R_n^2,\ Y := R_{N+1}^2 - R_n^2,\ \Theta := (X_1, ..., X_n)/R_n,\ Q = \frac{R_n^2}{R_N^2}.$$

It is immediately clear that D and Y are independent, as X_i and X_l for $i \leq n$, $l > n$ are independent and so all functions of X_i and X_l are independent, including D and Y. Similarly, it is clear that Y and Θ are independent. Additionally, D and T are independent.

To see this, observe that for $(X_1, ..., X_n)$ a standard normal vector, its joint distribution is given by

$$\mathbb{P}_{X_1,...,X_n}(dx) = \frac{1}{(2\pi)^{n/2}} e^{-||x||^2/2} |dx|,$$

which, when converted to spherical coordinates [10] yields

$$\mathbb{P}_{X_1,...,X_n}(dx) = \frac{r}{(2\pi)^n} e^{-r^2/2} = \frac{1}{2\pi} r^{n-1} e^{-r^2/2} dr d\theta,$$

where $d\theta$ is the area element on the unit sphere $\{\|x\| = 1\} \subset \mathbb{R}^n$. The product form of this joint density demonstrates that R_n and T are independent.

Thus, $Q = \frac{R_n^2}{R_{N+1}^2}$, a function of D and Y, is independent of Θ, and is distributed according to $\beta(n/2, (N+1-n)/2)$ by Lemma 2.14. Observe that

$$\begin{aligned}
\sigma^N(A^N) &= \mathbb{P}\Big[\frac{\sqrt{N}}{R_{N+1}} (X_1, ..., X_n) \in A \Big] \\
&= \mathbb{P}\Big[\big(N R_n^2 / R_{N+1}^2 \big)^{1/2} \cdot \frac{1}{R_n} (X_1, ..., X_n) \in A \Big] \\
&= \mathbb{P}\big[(NQ)^{1/2} \cdot \Theta \in A \big] = \mathbb{E}\big[\boldsymbol{I}_A\big((NQ)^{1/2} \cdot \Theta \big) \big] \\
&= \int_{S^{n-1} \times [0,1]} \boldsymbol{I}_A\big(\sqrt{Nq}\theta \big) \mathbb{P}_\Theta[d\theta] \mathbb{P}_Q[dq]
\end{aligned}$$

$$= \frac{1}{B\big(n/2, \frac{N+1-n}{2}\big)} \int_0^1 q^{n/2-1} (1-q)^{\frac{N+1-n}{2}-1} dt \cdot \int_{S_1^{n-1}} \boldsymbol{I}_A\big(\sqrt{Nq}\theta \big) \sigma_1^{n-1}[d\theta]$$

$$= \frac{1}{B\big(n/2, \frac{N+1-n}{2}\big)} \int_{S_1^{n-1}} \int_0^1 \boldsymbol{I}_A\big(\sqrt{N\theta} q^{n/2-1} (1-q)^{\frac{N+1-n}{2}-1} dt d\sigma_1^{n-1}(\theta)$$

($q = u^2/N$, $dq = 2u/N du$)

$$= \frac{1}{B\big((n/2, (N+1-n)/2 \big)} \frac{2}{N} \times$$

$$\times \int_{S_1^{n-1}} \int_0^{\sqrt{N}} I_A(u\theta)(u)(u^2/N)^{n/2-1} \left(1 - \frac{u^2}{N} \right)^{N+1-n/2} du\, \sigma_1^{n-1}[d\theta]$$

$$= \frac{1}{B\big((n/2, (N+1-n)/2 \big)} \frac{2}{N^{n/2}} \times$$

$$\times \int_{S_1^{n-1}} \int_0^{\sqrt{N}} \boldsymbol{I}_A(ux) u^{n-1} (1 - \frac{u^2}{N})^{\frac{N+1-n}{2}-1} du\, \sigma_1^{n-1}[d\theta].$$

Recall that

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

so

$$\frac{2}{(N^{n/2})B\big((n/2,(N+1-n)/2\big)} = \frac{2}{N^{n/2}} \cdot \frac{\Gamma(\frac{N+1}{2})}{\Gamma(n/2)\Gamma(\frac{N+1-n}{2})}.$$

By Lemma 2.13,

$$\frac{\Gamma(\frac{N+1}{2})}{\Gamma(\frac{N+1-n}{2})} \sim \left(\frac{N+1}{2}\right)^{n/2},$$

and so

$$\lim_{N\to\infty} \frac{2}{N^{n/2}} \cdot \frac{\Gamma(\frac{N+1}{2})}{\Gamma(n/2)\Gamma(\frac{N+1-n}{2})} = \frac{2}{\Gamma(n/2)} \lim_{N\to\infty} \frac{\Gamma(\frac{N+1}{2})}{(N^{n/2})\Gamma(\frac{N+1-n}{2})}$$

$$= \frac{2}{\Gamma(n/2)} \lim_{N\to\infty} \left(\frac{N+1}{2}\right)^{n/2} \left(\frac{1}{N}\right)^{n/2} = \frac{2}{\Gamma(n/2)} \lim_{N\to\infty} \left(\frac{1}{2} + \frac{1}{2N}\right)^{n/2} = \frac{2}{\Gamma(n/2)\cdot 2^{n/2}}.$$

Likewise,

$$\lim_{N\to\infty} \left(1 - \frac{u^2}{N}\right)^{\frac{N+1-n}{2}-1} = e^{-u^2/2},$$

and so in the limit as $N \to \infty$,

$$\lim_{N\to\infty} \frac{1}{B\big((n/2,(N+1-n)/2\big)} \frac{2}{N^{n/2}} \times$$

$$\int_{S_1^{n-1}} \int_0^{\sqrt{N}} \boldsymbol{I}_A(ux) u^{n-1} (1 - \frac{u^2}{N})^{\frac{N+1-n}{2}-1} du\, \sigma_1^{n-1}[d\theta]$$

$$= \frac{2}{\Gamma(n/2)\cdot 2^{n/2}} \int_{S_1^{n-1}} \int_0^{\infty} \boldsymbol{I}_A(ux) u^{n-1} e^{-u^2/2} du\, \sigma_1^{n-1}[d\theta].$$

But this is precisely $\gamma_n(A)$ in polar coordinates, so

$$\lim_{N\to\infty} \sigma^N\big(A^N\big) = \gamma_n(A),$$

for any Borel set $A \subset \mathbb{R}^n$. □

An immediate corollary of this lemma is the isoperimetric inequality for the Gaussian measure.

Corollary 2.16. *Let $A \subset \mathbb{R}^n$ be a Borel set, and let $H := \{x \in \mathbb{R}^n : \langle x, a\rangle \leq a\} \subset \mathbb{R}^n$ be a half space such that $\gamma_n(A) = \gamma_n(H)$.*

Then, $\forall\, r \geq 0$,

$$\gamma_n(A_r) \geq \gamma_n(H_r).$$

Proof. To see this, recall that the extremal sets for the isoperimetric inequality (2.4) for the uniform measure on $\mathbb{S}^n$ were geodesic balls, or spherical caps. In the notation of the proof of the Poincaré Lemma, for a half space H, $H^N = \Pi^{-1}_{N+1,n}(H) \cap \Sigma^N$ is a spherical cap $B \subset \mathbb{S}^n$. Thus, as $N \to \infty$, by the Poincaré Lemma,

$$\sigma^N(B) = \gamma_n(H),$$

giving us the extremal sets for the Gaussian measure on $\mathbb{R}^n$. □

Remark 2.17. Because γ_n is rotation invariant and a product measure, the measure of a half space is actually computed in one dimension, using the distribution of γ_1, Φ, where

$$\Phi(t) = \int_{-\infty}^{t} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}, \quad t \in \mathbb{R}.$$

Then, for $H = \{x \in \mathbb{R}^n : \langle x, a\rangle \leq a\}$, we can write $\gamma_n(H) = \Phi(a)$ and Corollary 2.16 expresses equivalently that

$$\gamma_n(A_r) \geq \Phi(a + r) \quad \forall\, r \geq 0.$$

□

Thus we can now identify the isoperimetric function for the Gaussian measure.

Theorem 2.18. *For every $A \subset \mathbb{R}^n$ Borel, and for every $r \geq 0$,*

$$\Phi^{-1}(\gamma_n(A_r)) \geq \Phi^{-1}(\gamma_n(A)) + r,$$

and consequently

$$I_\gamma = \Phi' \circ \Phi^{-1}.$$

Further, the equality $\gamma_n^+(A) = I_{\gamma_n}(\gamma_n(A))$ holds if and only if A is a half space in $\mathbb{R}^n$.

Proof. The first claim follows immediately from the preceding remark. Indeed,

$$\gamma_n(A_r) \geq \Phi(a + r) \implies \Phi^{-1}(\gamma_n(A_r)) \geq a + r = \Phi^{-1}(\gamma_n(A)) + r,$$

where the last equality is due to our assumption that $\gamma_n(A) = \Phi(a)$. The Gaussian isoperimetric function follows then similarly to Remark 2.7.

$$\gamma_n^+(A) = \liminf_{r\to 0} \frac{1}{r}\left[\gamma_n(A_r) - \gamma_n(A)\right] = \liminf_{r\to 0} \frac{1}{r}\left[\Phi(\Phi^{-1}(\gamma_n(A_r))) - \gamma_n(A)\right]$$
$$\geq \liminf_{r\to 0} \frac{1}{r}\left[\Phi(\Phi^{-1}(\gamma_n(A)) + r) - \gamma_n(A)\right] = \liminf_{r\to 0} \frac{1}{r}\left[\Phi(\Phi^{-1}(\gamma_n(A)) + r) - \Phi(\Phi^{-1}(\gamma_n(A)))\right]$$
$$= \Phi' \circ \Phi^{-1}(\gamma_n(A)).$$

Thus, as desired, $\gamma_n^+ \geq \Phi' \circ \Phi^{-1}$, with equality on half spaces. □

Finally, we are able to establish normal concentration for the Gaussian measure on $\mathbb{R}^n$, via isoperimetry.

Theorem 2.19. *For the convenient m.m.s. $(\mathbb{R}^n, |\cdot|, \gamma_n)$, where $|\cdot|$ is the standard Euclidean distance and γ_n is the n-dimensional Gaussian measure,*

$$\alpha_{\gamma_n}(r) \leq e^{-r^2/2}.$$

Proof. By Theorem 2.18, the isoperimetric function for γ_n is $I_\gamma = \Phi' \circ \Phi^{-1}$, where

$$\Phi(t) = \int_{-\infty}^{t} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}, \quad t \in \mathbb{R}.$$

Thus, $\Phi(0) = 1/2$, and so $0 = \Phi^{-1}(1/2)$. By Corollary 2.9, taking $v = \Phi$, it follows that

$$\alpha_{\gamma_n}(r) \leq 1 - \Phi(\Phi^{-1}(\tfrac{1}{2}) + r), \quad r > 0,$$

and so

$$\alpha_{\gamma_n}(r) \leq 1 - \Phi(r) \leq e^{-r^2/2}, \quad r > 0,$$

where the second inequality follows from

$$1 - \Phi(r) = \int_{r}^{\infty} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} \leq e^{-r^2/2}.$$

□

We have now established Gaussian concentration through isoperimetric methods. In the next section, we will again prove normal concentration for the Gaussian measure on $\mathbb{R}^n$, but through very different techniques.

3. Semigroup Methods

We now develop semigroup theory and use it to demonstrate concentration of measure for the Gaussian measure on $\mathbb{R}^n$, among other measures that satisfy a strong convexity assumption. We begin by defining semigroups and their generators. After, we discuss the Hille-Yosida theory and when an operator can generate a semigroup. From there we present a class of operators that do in fact generate good semigroups, and finally discuss how some curvature, or convexity, assumptions imply concentration of measure. Results in this section follow from Chapters 1 and 3 in [3] unless otherwise noted. General inspiration was also attained from [14].

3.1. Semigroups and generators.

Definition 3.1. Let X be a Banach space. A family $(P_t)_{t\geq 0}$ of bounded linear operators from $X \to X$ is called a *semigroup of bounded linear operators* on X if

(i) $P_0 = I$, where I is the identity operator on X.
(ii) $P_{t+s} = P_t \circ P_s \quad \forall\, t, s \geq 0$. (semigroup property)

The semigroup is called a *semigroup of contractions* if

$$\|P_t\| \leq 1, \;\; \forall t \geq 0.$$

The semigroup is called *strongly continuous* of C_0-*semigroup* if

$$\lim_{t\to 0} P_t x = x \quad \forall\, x \in X.$$

□

Semigroups of bounded linear operators have convenient, quantifiable bounds on their operator norms as well, which we now discuss.

Proposition 3.2. *This result is inspired by the proof of Theorem 2.2 in [13, p 4]. Let P_t be a C_0-semigroup on a Banach space X. Then*

$$\exists\, w \geq 0,\; M \geq 1: \quad ||P_t|| \leq Me^{wt} \quad \forall\, 0 \leq t < \infty.$$

If $w = 0$, then P_t is uniformly bounded, and if $M = 1$, then P_t is called a semigroup of contractions.

Proof. Fix $T > 0$. We first show that

$$M := \sup_{t\in[0,T]} ||P_t|| < \infty. \tag{3.1}$$

Indeed, if this were not true, then we could find a sequence $(\tau_n) \in [0, T]$ such that

$$\|P_{\tau_n}\| > 2^{n^2}.$$

Set $t_n := \tau_n/n$ so that $t_n \searrow 0$. From the semigroup property we deduce

$$2^{n^2} < \|P_{\tau_n}\| \leq \|P_{t_n}\|^n \implies 2^n < \|P_{t_n}\|, \;\; \forall n \geq 1.$$

Then, by the Uniform Boundedness Theorem, $\exists\, x \in X$ such that $||P_{t_n}x||$ is unbounded, which contradicts the strong continuity property of our semigroup. This proves (3.1).

Since$||P_0|| = 1$, we must have $M \geq 1$. Set

$$w := \frac{\log M}{T} \geq 0.$$

Observe that $\forall\, t \geq 0$, we can write $t = nT + q$, where $0 \leq q < T$. By the semigroup property,

$$||P_t|| = ||P_{nT+q}|| = ||P_{nT}P_q|| = ||P_T^n P_q||$$

$$\leq MM^n = MM^{t-q/T} \leq MM^{t/T} = Me^{wt}.$$

as desired. □

The above notions of continuity are essential for guaranteeing the existence of a generator, which is essentially the time-derivative of the semigroup at $t = 0$.

Definition 3.3. Given a semigroup of bounded linear operators P_t on a Banach space X, we define the linear operator $L : D(L) \subset X \to X$

$$Lx = \lim_{t\to 0} \frac{P_t x - x}{t} = \frac{d^+ P_t x}{dt}\bigg|_{t=0} \quad x \in D(L),$$

where

$$D(L) := \big\{\, x \in X \;:\; Lx \text{ exists} \,\big\}.$$

We call L the *generator* of P_t. □

Now, before we proceed, let us recall some definitions from functional analysis.

Definition 3.4 (Resolvent)**.** Let X be a Banach space and let

$$L : D(L) \subset X \to X$$

be a potentially unbounded linear operator. Then the *resolvent set* of L is defined as

$$\rho(L) := \{\lambda \in \mathbb{C} \;:\; \lambda I - L \text{ is invertible}\}.$$

Note, $\lambda I - L$ being invertible is equivalent to $(\lambda I - L)^{-1}$ being a bounded linear operator on X. For the family of such operators, we write

$$R(\lambda : L) := (\lambda I - L)^{-1}, \quad \lambda \in \rho(L).$$ □

Definition 3.5 (Closed Operator)**.** On a Banach space X, a linear operator

$$L : D(L) \subset X \to X$$

is *closed* if $\forall\, (x_n) \subset D(L)$ such that $x_n \to x \in X$ and $Lx_n \to y$, we have that $x \in D(L)$ and $Lx = y$. □

If X is the the space $\mathbb{R}^n$ equipped with the standard Euclidean norm, then any C_0-semigroup of contractions on X has as the form e^{tA} where the eigenvalues of the generator A have nonpositive real part. The famous Hille-Yosida theorem shows that the situation is essentially the same in infinite dimensions. We do not prove this result, but refer the interested reader to [13, Thm 3.1, p 8] for the proof.

Theorem 3.6 (Hille-Yosida)**.** *Given a Banach space X, A linear unbounded operator $A : X \to X$ is the generator of a C_0-semigroup of contractions P_t if and only if*

(i) *A is closed and $\overline{D(A)} = X$.*
(ii) *$\mathbb{R}^+ \subset \rho(A)$ and $\forall\, \lambda > 0$,*

$$||R(\lambda : A)|| \leq \frac{1}{\lambda}.$$

3.2. **Markov semigroups and their generators.** We now transition into a more probabilistic concept, the Markov semigroup. However, before we can introduce this semigroup, we must discuss specific measure-theoretic assumptions on the underlying space which we will be working with.

Definition 3.7. A *kernel* on a measurable space $(E, \mathcal{F})$ is a function

$$p : E \times \mathcal{F} \to [0, 1]$$

such that

(i) $\forall\, x \in E, p(x, \cdot)$ is a probability measure on $(E, \mathcal{F})$, and
(ii) $\forall\, A \in \mathcal{F}, p(\cdot, A)$ is a measurable function from $E \to [0, 1]$.

Further, we say that the space $(E, \mathcal{F})$ exhibits the *measure decomposition property* if every probability measure μ on $(E \times E, \mathcal{F} \otimes \mathcal{F})$ can be decomposed as, for some kernel k,

$$\mu(dx, dy) = k(x, dy)\mu_1(dx),$$

where μ_1 is the projection of μ onto the first coordinate. □

One can show (see e.g. [5, Th. IV2.10]) that if E is a Polish space[4] and $\mathcal{F}$ is its Borel σ-algebra, then $(E, \mathcal{F})$ exhibits the measure decomposition property.

Definition 3.8. Suppose that $(E, \mathcal{F})$ is a measurable space.

(i) A measure μ on $(E, \mathcal{F})$ is called *σ-finite* if E is the countable union of μ-measurable sets of finite measure.
(ii) We say that $(E, \mathcal{F})$is *good* if it satisfies the measure-decomposition property defined above, and the σ-algebra $\mathcal{F}$ is generated by a countable family.
(iii) The measured space $(E, \mathcal{F}, \mu)$ is called *good* if $(E, \mathcal{F})$ is a good measurable space and μ is a σ-finite measure on $\mathcal{F}$. □

Now, let us attach some additional demands on the semigroups with which we will work.

Definition 3.9. A semigroup *over* a good measured space $(E, \mathcal{F}, \mu)$ is a semigroup $(P_t)_{t\geq 0}$ of bounded linear operators on $\mathbb{L}^\infty(E, \mu)$ such that

(i) **Identity**
$$P_0 f = f, \ \ \forall f \in \mathbb{L}^\infty(E, \mu).$$
(ii) **Mass conservation**
$$P_t(\boldsymbol{I}_E) = \boldsymbol{I}_E, \ \ \forall t \geq 0.$$
(iii) **Positivity**
$$\forall f \in \mathbb{L}^\infty(E, \mu) : \ \ f \geq 0 \implies P_t f \geq 0.$$

□

Semigroups extend to contractions via Jensen's inequality. Suppose that $(P_t)_{t\geq 0}$ is a semigroup over the good measured space $(E, \mathcal{F}, \mu)$.

Proposition 3.10 (Jensen's inequality)**.** *For any convex function $\varphi : \mathbb{R} \to \mathbb{R}$, any $f \in \mathbb{L}^\infty(E, \mu)$ and any $t > 0$ we have*

$$\varphi(P_t f) \leq P_t(\phi(f)). \tag{3.2}$$

[4]We recall that a Polish space is a seprable complete metric space.

Proof. We need to use a lesser known property of convex functions, [1, Thm.6.3.4]. More precisely, we use the fact that there exist sequences of real numbers $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such that

$$\varphi(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n).$$

Set $\ell_n(x) = a_n x + b_n$. Clearly,

$$\ell_n\big(\, P_t f\,\big) = a_n P_t f + b = P_t(a_m f + b) = P_t(\ell_n(f)) \leq P_t\varphi(f).$$

Hence

$$\varphi\big(\, P_t f\,\big) = \sup_{n\in\mathbb{N}} \ell_n\big(\, P_t f\,\big) = \sup_{n\in\mathbb{N}} P_t \ell_n(f) \leq P_t\varphi(f).$$

□

As a consequence of Jensen's inequality we deduce that for any $p \in [1,\infty)$ and any $f \in \mathbb{L}^\infty(E,\mu)$ we have

$$\|P_t f\|_{\mathbb{L}^p} \leq \|f\|_{\mathbb{L}^p}.$$

Since $\mathbb{L}^\infty$ is dense in $\mathbb{L}^p$ we deduce that P_t extends to a contraction $\mathbb{L}^p(E,\mu) \to \mathbb{L}^p(E,\mu)$ for any $p \in [1,\infty]$.

Now we can state the final version of a semigroup that we will be working with.

Definition 3.11 (Markov Semigroup)**.** A semigroup $(P_t)_{t\geq 0}$ of bounded linear operators over a good measured space (E,μ) is called a *Markov semigroup* the semigroup it induces on $\mathbb{L}^2(E,\mu)$ is a C_0-semigroup, i.e., for any $f \in \mathbb{L}^2(E,\mu)$

$$\lim_{t\searrow 0} P_t f = f, \;\; \text{in } \mathbb{L}^2(E,\mu).$$

For such a semigroup, we write $(P_t)_{t\geq 0}$ or (P_t) or, where the context is clear, P_t, equivalently. □

By Hille-Yosida, any Markov semigroup has an infinitesimal generator

$$L : D(L) \subset \mathbb{L}^2(E,\mu) \to \mathbb{L}^2(E,\mu)$$

determined by

$$Lf = \lim_{t\to 0} \frac{1}{t}\big(\, P_t f - f\,\big) \text{ in } \mathbb{L}^2(E,\mu), \;\; f \in D(L).$$

Moreover $P_t \in D(L)$ and

$$\partial_t P_t f = L P_t f = P_t L f, \;\; \forall t \geq 0, \;\; f \in D(L).$$

For such a Markov semigroup, there are two important measures to consider.

Definition 3.12 (Invariant measure)**.** Given a Markov semigroup P_t over a good measurable space $(E,\mathcal{F},\mu)$, the measure μ is said to be *invariant* for P_t if for all bounded positive measurable functions $f : E \to \mathbb{R}$, and $\forall\, t \geq 0$,

$$\int_E P_t f d\mu = \int_E f d\mu. \tag{3.3}$$

□

Definition 3.13 (Reversible measure)**.** Given a Markov semigroup P_t over a good measurable space $(E,\mathcal{F})$, a measure μ is *reversible* for P_t, or P_t is *symmetric* with respect to μ, if

$$\forall f,g \in \mathbb{L}^2(E,\mu), \; \forall\, t \geq 0, \quad \int_E f P_t g \, d\mu = \int_E g P_t f \, d\mu. \tag{3.4}$$

□

Remark 3.14. Observe that if μ is a probability measure, then the constant function $\boldsymbol{I}_E$ is in $\mathbb{L}^2(E,\mu)$, and so taking $f=\boldsymbol{I}_E$ in (3.4) we obtain (3.3). This shows shows that if a probability measure is symmetric for a Markov semigroup P_t, then it is also invariant for P_t.

The reversibility of a measure forces the generator to be symmetric. Indeed, differentiating the identity

$$\int_E fP_tg\,d\mu=\int_E gP_tf\,d\mu,\ \ f,g\in D(L)$$

at $t=0$ yields

$$\int_E fLg\,d\mu=\int_E gLf\,d\mu,\ \ \forall f,g\in D(L).$$

In general this does not imply that μ is invariant for P_t. □

Suppose that $(E,\mathcal{F},\nu)$ is a good measured space and $P_t:\mathbb{L}^2(E,\mu)\to\mathbb{L}^2(E,\mu)$ is a C_0-semigroup. To define a Markov semigroup it needs to satisfy additional conditions.

- $P_t\mathbb{L}^\infty(E,\mu)\subset\mathbb{L}^\infty(E,\mu)$.
- $P_tf\geq 0$, $\forall f\in\mathbb{L}^\infty(E,\mu)$, $f\geq 0$, $t\geq 0$.
- $P_t\boldsymbol{I}_E=\boldsymbol{I}_E$, $\forall t\geq 0$.

This is by no means obvious, nor guaranteed. However, we present below a large class of Markov semigroups.

Example 3.15. Let $n\in\mathbb{N}$ and suppose that E is the Euclidean vector space $\mathbb{R}^n$ equipped with the canonical inner product $(\cdot,\cdot)$. Let $\mathcal{F}$ denote the Borel sigma-algebra of E. Suppose that $W\in C^2(E)$ is such that

$$\int_{\mathbb{R}^n}e^{-W(x)}dx=1.$$

Then, the space $(E,\mathcal{F},\mu)$ is then a good measured space. Consider the differential operator

$$L=L_W:C_0^\infty(E)\to C_0^\infty(E),\ \ Lf=\Delta f-(\nabla W,\nabla f).$$

It satisfies the symmetry condition

$$\int_E g(Lf)d\mu=-\int_E(\nabla f,\nabla g)d\mu=\int_E f(Lg)d\mu,\ \ \forall f,g\in C_0^\infty(E).$$

We obtain an unbounded, symmetric operator

$$L:C_0^\infty(E)\subset\mathbb{L}^2(E,\mu)\to\mathbb{L}^2(E,\mu).$$

Once can show (see [4, Thm. 4.6]) that its closure in $\mathbb{L}^2(E,\mu)$ is a nonpositive selfadjoint operator and thus generates a symmetric C_0-semigroup of contractions

$$P_t:\mathbb{L}^2(E,\mu)\to\mathbb{L}^2(E,\mu).$$

This semigroup satisfies the additional conditions [4, Thm. 4.25], [4, Thm. 4.27]

$$P_tf\geq 0,\ \ \forall f\geq 0,\ \ t\geq 0. \tag{3.5a}$$

$$\|P_tf\|_\infty\leq\|f\|_\infty,\ \ \forall t\geq 0,\ \ \forall f\in\mathbb{L}^\infty(E,\mu) \tag{3.5b}$$

Moreover μ is also invariant for P_t; see [8, Sec. 8.1.2]. In other words P_t is a Markov semigroup on (E,μ).

Suppose additionally that $W\in C^2(\mathbb{R}^n)$ satisfies the uniform convexity requirement

$$\exists c:0:\ \ \nabla^2(W(x)-c|x|^2/2)\geq 0,\ \ \forall x\in\mathbb{R}^n. \tag{3.6}$$

Then P_t also satisfies the mass conservation condition; see [2] or [3, Thm. 3.26] as thus it is a Markov semigroup on (E, μ). As shown in [3, Prop. 3.1.13], this semigroup also satisfies the following $\mathbb{L}^2$-ergodicity result

$$\forall f \in \mathbb{L}^2(E, \mu) : \lim_{t \to \infty} \left\| P_t f - \int_E f d\mu \right\|_{\mathbb{L}^2} = 0. \tag{3.7}$$

□

3.3. **Diffusions and the Carré-du-Champ operator.** To summarize, we have now developed the concept of a Markov semigroup and its generator, along with some of their fundamental properties. Additionally, using Hille-Yosida, we can go from a semigroup to its generator. However, in practice, it is much more common for one to be given a generator and asked to use the properties of its semigroup, if it exists. This is a very nuanced problem, and it comes down to finding a dense algebra of functions, $\mathcal{A} \subset D(L)$, in the topology of the domain, meaning that

$$\forall\, f \in D(L),\, \exists\, (f_k) \subset \mathcal{A} \,:\, f_k \to f, \text{ and } Lf_k \to Lf \text{ as } k \to \infty.$$

We then work with this known algebra of functions and extend results to $D(L)$ by density. In generality, this is almost impossible, so we restrict ourselves to a large and useful class of generators called diffusion operators, to which we can associate Markov semigroups. We begin by recalling the definition of a derivation.

Definition 3.16. On an algebra of functions $\mathcal{A}$, a *derivation* is a linear map $D : \mathcal{A} \to \mathcal{A}$ such that for all $f, g \in \mathcal{A}$,

$$D(fg) = fDg + gDf.$$

□

With this definition in mind, we define the Carré-du-champ operator, which intuitively measures how much a generator differs from being a derivation.

Definition 3.17 (Carré-du-champ)**.** Consider a Markov semigroup over a good measured space $(E, \mathcal{F}, \mu)$ with generator L. Let $\mathcal{A}$ be an algebra of functions contained in the domain of L. Then, the bilinear map $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$,

$$\Gamma(f, g) = \frac{1}{2}[L(fg) - fLg - gLf], \quad \forall\, (f, g) \in \mathcal{A} \times \mathcal{A},$$

is called the *Carré-du-champ operator* of the Markov generator L. For $\Gamma(f, f)$, we simply write $\Gamma(f)$. □

We now list some properties of Carré-du-champ operators.

Proposition 3.18. *Let P_t be a Markov semigroup over a good measured space $(E, \mathcal{F}, \mu)$, with generator L and symmetric, invariant measure μ. Let $\mathcal{A} \subset D(L)$ be an algebra of functions. Then, for all $f, g \in \mathcal{A}$,*

(i) $\Gamma(f) \geq 0$.
(ii) $\Gamma(f, g) \leq \sqrt{\Gamma(f)\Gamma(g)}$.
(iii) $\int_E \Gamma(f, g) d\mu = -\int_E fLg d\mu$.

Proof. First, (iii) follows from μ being invariant and symmetric. Namely,

$$2\Gamma(f, g) = L(fg) - fLg - gLf$$

$$2\int_E \Gamma(f, g) = \int_E L(fg) d\mu - \int_E fLg d\mu - \int_E gLf\, d\mu$$

$$2\int_E \Gamma(f, g) = 0 - 2\int_E fLg\, d\mu, \quad \int_E \Gamma(f, g) = -\int_E fLg\, d\mu.$$

Property (ii) follows from property (i) applied to $f+g$ and the bilinearity of Γ:

$$\forall a \in \mathbb{R} \;\; 0 \leq \Gamma(f+ag, f+ag) = \Gamma(f,f) + 2a\Gamma(f,g) + a^2\Gamma(g,g).$$

Choosing $a = -\Gamma(f,g)/\Gamma(g,g)$ yields the desired result. Finally, to see property (i), observe that $2\Gamma(f) = Lf^2 - 2fLf$, so if suffices to show that $Lf^2 \geq 2fLf$. By Proposition 3.10 applied to the convex function $\phi(r) = r^2$, we have for every bounded measurable $f : E \to \mathbb{R}$,

$$P_t(f^2) \geq (P_t f)^2.$$

By the product rule applied to $d(P_t f)^2/dt$,

$$\frac{d(P_t f)^2}{dt} = 2f\frac{dP_t f}{dt} = 2fLP_t f.$$

And finally by the definition of the generator,

$$Lf^2 = \lim_{t\to 0} \frac{P_t f^2 - f^2}{t}$$
$$\geq \lim_{t\to 0} \frac{(P_t f)^2 - f^2}{t} = \frac{d(P_t f)^2}{dt}\bigg|_{t=0} = 2fLP_t f\bigg|_{t=0} = 2fLf,$$

yielding the desired property. □

Remark 3.19. The above property (iii) is very important because, if $\mathcal{A}$ is dense in $D(L)$, it states that a measured space and a given Carré-du-champ operator completely characterize a Markov generator and its respective Markov semigroup.

Indeed, in the literature it is common to work with a *Markov triple* (E, μ, Γ) and deduce the properties of the corresponding generator and semigroup from this triple. We later will work with a different tuple, specifying not a Carré-du-champ operator, but rather a diffusion operator, for which the conditions of the Hille-Yosida theorem will be satisfied, making the operator the generator of a Markov semigroup. □

Example 3.20. Consider the good measured space $(E, \mathcal{F}, \mu)$ discussed in Example 3.15, where $n \in \mathbb{N}$, E is the Euclidean vector space $\mathbb{R}^n$ equipped with the canonical inner product $(\cdot,\cdot)$, $\mathcal{F}$ is the Borel sigma-algebra of E, and μ is the Borel probability measure $\mu(dx) = e^{-W(x)}dx$, such that $W \in C^2(E)$ and

$$\int_E e^{-W} dx = 1.$$

Consider the symmetric differential operator

$$L : C_0^\infty(E) \subset \mathbb{L}^2(E,\mu) \to \mathbb{L}^2(E,\mu), \quad Lf = \Delta f - (\nabla W, \nabla f),$$

whose closure generates a Markov semigroup P_t. Then, the associated Carré-du-champ operator to L is of the form

$$\Gamma(f,g) = \nabla f \cdot \nabla g.$$

Proof. Observe, taking $L = \Delta - \nabla W \cdot \nabla$, then, for any $f, g \in C_0^\infty(E)$ we have

$$\int_E fLg = \int_E f(\Delta g - \nabla W \nabla g) d\mu$$
$$= \int_E f\Delta g d\mu + \int_E f\nabla g(-\nabla W)e^{-W}dx = \int_E f\Delta g d\mu + \int_E f\nabla g \nabla(e^{-W})dx$$
$$= \int_E f\Delta g d\mu - \int_E \nabla(f\nabla g)e^{-W}dx = \int_E f\Delta g d\mu - \int_E \nabla f \nabla g + f\Delta g d\mu$$

$$= -\int_E \nabla f \nabla g \, d\mu = -\int_E \Gamma(f,g) d\mu,$$

as desired. Note, the fourth equality follows from integration by parts in $\mathbb{R}^n$, which is possible as f and g are compactly supported, and we can integrate equivalently on an open ball in $\mathbb{R}^n$ containing the support of f and g. □

Now we transition into diffusion operators, first defining some notation.

Definition 3.21. Let Ψ_0 be the union over $n \geq 1$ of the spaces of smooth functions from $\psi : \mathbb{R}^n \to \mathbb{R}$ such that $\psi(0) = 0$, i.e.

$$\Psi_0 = \bigcup_{n \geq 1} \left\{ \psi \in C^\infty(\mathbb{R}^n, \mathbb{R}) \, : \, \psi(0) = 0 \right\}.$$ □

Definition 3.22 (Diffusion)**.** Consider a symmetric Markov semigroup P_t over a good measured space $(E, \mathcal{F}, \mu)$ with generator L and Carré-du-champ operator Γ, defined on $\mathcal{A} \times \mathcal{A}$, where $\mathcal{A} \subset D(L)$ is an algebra of bounded measurable functions $f : E \to \mathbb{R}$ dense in all $\mathbb{L}^p(E, \mu)$ spaces and such that $\forall k \in \mathbb{N}$, $\forall f_1, \ldots, f_k \in \mathcal{A}$ and any $\psi \in \Psi_0$, we have

$$\psi(f_1, \ldots, f_k) \in \mathcal{A}.$$

The Carré-du-champ operator Γ is called a *diffusion* if for all $\forall k \in \mathbb{N}$, $\forall f_1, \ldots, f_k \in \mathcal{A}$ and any $\psi \in \Psi_0$ we have

$$\Gamma(\psi(f_1, ..., f_k), g) = \sum_{i=1}^{k} \partial_i \psi(f_1, ..., f_k) \Gamma(f_i, g).$$

Equivalently this means that

$$L\psi(f_1, ..., f_k) = \sum_{i=1}^{k} \partial_i \psi(f_1, ..., f_k) L f_i + \sum_{i,j=1}^{k} \partial_i \partial_j \psi(f_1, ..., f_k) \Gamma(f_i, f_j),$$

in which case we call L a *diffusion generator* and P_t a *Markov diffusion semigroup*. □

Remark 3.23. The above definition can be viewed as the requirement that Γ satisfies a sort of chain rule. Indeed, for $k = 1$, the above identities reduces to the equalites

$$\Gamma(\psi(f), g) = \psi'(f) \Gamma(f, g),$$

and

$$L\psi(f) = \psi'(f) L f + \psi''(f) \Gamma(f).$$

For $\psi(f, g) = fg$, a polynomial function, the definition for Γ simplifies to the condition

$$\Gamma(fg, h) = f\Gamma(g, h) + g\Gamma(f, h), \quad \forall \, f, g, h \in \mathcal{A},$$

which appears much like a chain rule. We can transfer the diffusion properties of generators and Carré-du-champ operators back and forth through the integration by parts formula. □

Example 3.24. Consider again the good measured space $(E, \mathcal{F}, \mu)$ discussed in Example 3.15, where $n \in \mathbb{N}$, E is the Euclidean vector space $\mathbb{R}^n$ equipped with the canonical inner product $(\cdot, \cdot)$, $\mathcal{F}$ is the Borel sigma-algebra of E, and μ is the Borel probability measure $\mu(dx) = e^{-W(x)} dx$, where $W \in C^2(E)$ satisfies the uniform convexity condition (3.6).

As explained in Example 3.15, the symmetric operator

$$L : C_0^\infty(E) \subset \mathbb{L}^2(E, \mu) \to \mathbb{L}^2(E, \mu), \quad Lf = \Delta f - (\nabla W, \nabla f),$$

is essentially selfadjoint and its closure closure generates a Markov semigroup P_t. Then, L is a diffusion generator.

To see this, recall that, by Definition 3.22, it suffices to show that for $\forall\, \psi \in \Psi_0$,

$$L\psi(f_1, ..., f_k) = \sum_{i=1}^{k} \partial_i \psi(f_1, ..., f_k) L f_i + \sum_{i,j=1}^{k} \partial_i \partial_j \psi(f_1, ..., f_k) \Gamma(f_i, f_j).$$

We will prove this result for $k = 1$. The computations for $k > 1$ follows similarly. Let $\psi \in \Psi_0$ and let $f \in \mathcal{A} = C_0^\infty(E)$. Then,

$$\begin{aligned} L\psi(f) &= \Delta(\psi(f)) - \nabla W \cdot \nabla(\psi(f)) \\ &= \nabla(\nabla\psi(f)\nabla f) - \nabla W \cdot (\nabla\psi(f)\nabla f) \\ &= \Delta\psi(f)\nabla f \nabla f + \nabla\psi(f)\Delta f - \nabla W \nabla\psi(f)\nabla f \\ &= \nabla\psi(f)\Big(\Delta f - \nabla W \nabla f\Big) + \Delta\psi(f)\nabla f \nabla f \\ &= \nabla\psi(f) L f + \Delta\psi(f)\Gamma(f), \end{aligned}$$

as desired. It is easy to check as well that the associated Carré-du-champ operator $\Gamma(f, g) = \nabla f \cdot \nabla g$ is a diffusion as well. □

The reason we discuss diffusions is because generally it is much easier to check that a Carré-du-champ operator or a generator is a diffusion than it is to check that the generator satisfies the conditions of the Hille-Yosida theorem. Thus, diffusions, under some additional technical conditions, guarantee Markov semigroups

Definition 3.25. For a good measured space $(E, \mathcal{F}, \mu)$, we will call an algebra of functions $\mathcal{A}$ *nice* if $\mathcal{A}$ is dense in $\mathbb{L}^p(E, \mu)$ for all $1 \le p \le \infty$ and $\mathcal{A}$ is stable by Ψ_0. □

.

Definition 3.26 (Good Markov tuples)**.** The tuple $(E, \mathcal{F}, \mu, L, \mathcal{A})$ is called a *good Markov tuple* if the following hold.

(i) $(E, \mathcal{F}, \mu)$ is a good measured space with μ a probability measure.
(ii) $\mathcal{A} \subset \mathbb{L}^2(E, \mu)$ is nice algebra of functions.
(iii) $L : \mathcal{A} \subset \mathbb{L}^2(E, \mu) \to \mathbb{L}^2(E, \mu)$ is a symmetric diffusion operator, its closure is self-adjoint, and it generates a Markov diffusion semigroup.

□

Remark 3.27. The preceding definition is a very restricted case of the discussion developed throughout [3, Chap 3]. There, the authors define the general structure of a Markov tuple not in terms of a operator L, but in terms of a good measured space $(E, \mathcal{F}, \mu)$ and an abstract notion of a Carré-du-champ operator Γ, calling their construction (E, μ, Γ) a *Markov triple*, implicitly assuming an underlying algebra $\mathcal{A}$. They define a number of various Markov triples, beginning with the assumption that Γ is a diffusion, [3, Def 3.1.1, p 121], and adding assumptions successively, culminating in the definition of a *compact* Markov triple, [3, 3.4.4, p 171], which is equivalent to our assumptions on a good Markov tuple. □

Example 3.28. Consider again the good measured space $(E, \mathcal{F}, \mu)$ discussed in Example 3.15, where $n \in \mathbb{N}$, E is the Euclidean vector space $\mathbb{R}^n$ equipped with the canonical inner product $(\cdot, \cdot)$, $\mathcal{F}$ is the

Borel sigma-algebra of E, and μ is the Borel probability measure

$$\mu(dx) = e^{-W(x)}dx,$$

where $W \in C^2(E)$ satisfies the uniform convexity condition (3.6). Recall the symmetric operator

$$L : C_0^\infty(E) \subset \mathbb{L}^2(E,\mu) \to \mathbb{L}^2(E,\mu), \quad Lf = \Delta f - (\nabla W, \nabla f),$$

and let $\mathcal{A} = C_0^\infty(E)$. Then $(E, \mathcal{F}, \mu, L, \mathcal{A})$ is a good Markov tuple, as we have shown the closure of L generates a Markov semigroup, and by (3.24), L is a diffusion. For brevity, we will refer to this tuple as the *good W-tuple* from now on. □

Remark 3.29. Our later results will be proven in the context of Definition 3.26, but it is important to state that this setting is not actually sufficient for all of the results we prove. Indeed, the issue is that later we will deal with results that implicitly assume that for $f \in \mathcal{A}$, $P_t f \in \mathcal{A}$, which is generally not the case.

To remedy this, [3] introduces an exterior algebra $\mathcal{A}_e$, [3, Section 3.3, p 151], with the following properties.

(i) $\mathcal{A} \subset \mathcal{A}_e$.
(ii) $\forall f \in \mathcal{A}_e, g \in \mathcal{A}, gf \in \mathcal{A}$ (ideal property).
(iii) $\mathcal{A}_e$ is stable by smooth functions.
(iv) If $f \in \mathcal{A}_e$: $\int_E gf \, d\mu \geq 0 \, \forall \, g \in \mathcal{A}$, then $f \geq 0$.
(v) There are no integrability assumptions on $\mathcal{A}_e$. Namely, in most cases $\mathcal{A}_e \not\subset \mathbb{L}^2(E,\mu)$.

Then, [3] demands that in addition to generating a Markov diffusion semigroup, L also extends to a diffusion generator on $\mathcal{A}_e$, and that Γ is then defined on $\mathcal{A}_e \times \mathcal{A}_e$. Additionally, and most importantly, [3] also assumes that if P_t is the Markov diffusion semigroup generated by the diffusion operator L in a good Markov tuple $(E, \mathcal{F}, \mu, L, \mathcal{A})$, then $P_t\mathcal{A} \subset \mathcal{A}_e$. This assumption allows later curvature computations to be made rigorous.

For our good W-tuple (Ex. 3.28), $\mathcal{A}$ will be $C_0^\infty(E)$, and $\mathcal{A}_e$ is then $C^\infty(E)$. We define our later results in terms of Definition 3.26, leaving out the added analysis of the exterior algebra for notational convenience. The proofs are similar if not identical, and we will provide explicit citations whenever an exterior algebra is needed. □

Now we will discuss the property of *ergodicity* for good Markov tuples. In a good Markov tuple $(E, \mathcal{F}, \mu, L, \mathcal{A})$, μ is a probability measure and is invariant for P_t. Moreover, $P_t\boldsymbol{I}_E = \boldsymbol{I}_E$. As explained, see [3, Prop. 3.1.3] for a more nuanced discussion.

Proposition 3.30 (Ergodicity)**.** *Let P_t be the Markov diffusion semigroup generated by a good Markov tuple $(E, \mathcal{F}, \mu, L, \mathcal{A})$. If* $\ker L$ *consists only of constant functions, then, for all $f \in \mathbb{L}^2(E,\mu)$,*

$$\lim_{t\to\infty} P_t f = \int_E f \, d\mu \text{ in } \mathbb{L}^2(E,\mu). \qquad \square$$

3.4. **Curvature and Dimension.** For our purposes we have now developed a sufficient theory of Markov semigroups and their associated operators. We turn now to some comments on curvature and dimension that will provide some necessary identities to establish concentration of measure. We begin by defining a new bilinear form out of the Carré-du-champ operator, and afterwards define the necessary curvature conditions from conditions on this new operator.

Definition 3.31. Let $(E, \mathcal{F}, \mu, L, \mathcal{A})$ be a good Markov tuple. Let $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ be the Carré-du-champ operator defined from L. Then, $\forall \, f, g \in \mathcal{A}$, we define the *iterated Carré-du-champ operator*

$\Gamma_2 : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$,

$$\Gamma_2(f,g) = \frac{1}{2}[L\Gamma(f,g) - \Gamma(f,Lg) - \Gamma(Lf,g)], \quad \forall\, f,g \in \mathcal{A}.$$

Again, we write $\Gamma_2(f,f) := \Gamma_2(f)$. □

This operator also enjoys an integration by parts formula, much like the Carré-du-champ operator it is defined from.

Proposition 3.32. *Let $(E,\mathcal{F},\mu,L,\mathcal{A})$ be a good Markov tuple. Then, the iterated Carré-du-champ operator enjoys an integration by parts formula. Namely, $\forall\, f,g \in \mathcal{A}$,*

$$\int_E \Gamma_2(f,g)d\mu = \int_E (Lf)(Lg)d\mu.$$

Proof. By the definition of Γ_2,

$$\Gamma_2(f,g) = \frac{1}{2}[L\Gamma(f,g) - \Gamma(f,Lg) - \Gamma(Lf,g)],$$

so that,

$$2\int_E \Gamma_2(f,g)d\mu = \underbrace{\int_E L\Gamma(f,g)d\mu}_{=0} - \int_E \Gamma(f,Lg)d\mu - \int_E \Gamma(Lf,g)d\mu$$
$$= \int_E fL(Lg)d\mu + \int_E LfLgd\mu = 2\int_E (Lf)(Lg)d\mu,$$

where the third and fourth inequalities follow from μ being invariant and symmetric with respect to L. Diving by 2 yields the desired result. □

Example 3.33. Consider the good W-tuple from Example 3.28, where

$$L = \Delta - \nabla W \cdot \nabla, \quad \mu(dx) = e^{-W}dx,$$

and $\mathcal{A} = C_0^\infty(E)$. Recall, by Example 3.20, the associated Carré-du-champ operator to this tuple is $\Gamma(f,g) = \nabla f \cdot \nabla g$. Then, $\forall\, f \in \mathcal{A}$,

$$\Gamma_2(f) = |\nabla^2 f|^2 + \nabla^2 W(\nabla f, \nabla f).$$ □

This follows from simple but tedious computation, so we omit the proof.

The iterated Carré-du-champ operator is important because of its presence in curvature results for semigroups. Indeed, the curvature condition we will be dealing with is defined in terms of Γ_2.

Definition 3.34. For a good Markov tuple $(E,\mathcal{F},\mu,L,\mathcal{A})$ with associated Carré-du-champ operator $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$, the diffusion operator L satisfies the *curvature dimension condition $CD(R,n)$* of curvature $R \in \mathbb{R}$ and dimension $n \geq 1$ if $\forall\, f \in \mathcal{A}$,

$$\Gamma_2(f) \geq R\,\Gamma(f) + \frac{1}{n}(Lf)^2.$$

We say the operator L is of curvature R if L satisfies condition $CD(R,\infty)$, i.e. if

$$\Gamma_2(f) \geq R\,\Gamma(f).$$ □

We can alternatively express this curvature-dimension condition in terms of commutation of the semigroup with its associated Carré-du-champ operator.

Lemma 3.35. *Let $(E, \mathcal{F}, \mu, L, \mathcal{A})$ be a good Markov tuple with Carré-du-champ operator Γ generating the Markov diffusion semigroup P_t, symmetric with respect to μ. If L satisfies the curvature condition $CD(R, \infty)$ for some $R \in \mathbb{R}$, then $\forall\, f \in \mathcal{A}$ and for every $t \geq 0$,*

$$\Gamma(P_t f) \leq e^{-2Rt} P_t(\Gamma(f)).$$

Proof. Fix $f \in \mathcal{A}$ and $t > 0$. Then define the function

$$F(s) = e^{-2Rs}\Lambda(s), \quad s \in [0,t],$$

where

$$\Lambda(s) = P_s(\Gamma(P_{t-s}f)), \quad s \in [0,t].$$

As we define Γ on $\mathcal{A} \times \mathcal{A}$, the definition of $\Lambda(s)$ inherently assumes that $P_t\mathcal{A} \subset \mathcal{A}$, which like we mentioned is generally not true. The proof of this result in full detail, using exterior algebras, can be found in [3, Corollary 3.3.19, p 163]. However, we continue the proof only assuming the interior algebra $\mathcal{A}$, per [3, Theorem 3.2.3, p 144], as the proofs are similar barring technical details.

Observe that

$$\Lambda'(s) = LP_s(\Gamma(P_{t-s}f)) - 2P_s\big(\Gamma(P_{t-s}f, P_{t-s})\big),$$

where the first term comes from $\partial_t P_t = LP_t$ and the second term comes from

$$(\Gamma(f,f))' = \Gamma(f', f) + \Gamma(f, f') = 2\Gamma(f, f').$$

If we write $g(s) := P_{t-s}f$, then we can rewrite the above as

$$\Lambda'(s) = P_s(L\Gamma(g) - 2\Gamma(g, Lg)) = 2P_s\Gamma_2(g),$$

as the semigroup commutes with its generator. Thus,

$$\begin{aligned} F'(s) = -2Re^{-2Rs}\Lambda(s) + e^{-2Rs}\Lambda'(s) &= 2e^{-2Rs}\big(-RP_s(\Gamma(P_{t-sf})) + P_s\Gamma_2(P_{t-s}f)\big) \\ &\geq 2e^{-2Rs}\big(-RP_s(\Gamma(P_{t-sf})) + RP_s\Gamma(P_{t-s}f)\big) = 0. \end{aligned}$$

The inequality follows from our curvature assumptions. This shows that F is non-decreasing on $[0,t]$, and as all of the above is true $\forall s \in [0,t]$, we can compare $s = 0$ and $s = t$ to get the desired result. Indeed, as F is non-decreasing,

$$\begin{aligned} F(0) \leq F(t) &\implies P_0(\Gamma(P_t f)) \leq e^{-2Rt}P_t(\Gamma(P_0 f)) \\ &\implies \Gamma(P_t f) \leq e^{-2Rt}P_t(\Gamma(f)). \end{aligned}$$

□

Example 3.36. Consider again the good W-tuple $(E, \mathcal{F}, \mu, L, \mathcal{A})$, where

$$E = \mathbb{R}^n, L = \Delta - \nabla W \cdot \nabla, \;\; \mu(dx) = e^{-W}dx, \, \mathcal{A} = C_0^\infty(E).$$

Suppose that $W \in C^2(E)$ satisfies the uniform convexity requirement

$$\exists c : 0 : \;\; \nabla^2(W(x) - c|x|^2/2) \geq 0, \;\; \forall x \in E. \tag{3.8}$$

Above $\nabla^2 W$ denotes the Hessian of W. Upon modifying it by an additive constant we can assume that so that $\mu(dx) = e^{-W(x)}dx$ is a Borel probability measure on $\mathbb{R}^n$. Recall that the associated Carré-du-champ operator to the tuple is $\Gamma(f,g) = \nabla f \cdot \nabla g$. Then, L is of curvature c and satisfies the curvature-dimension condition $CD(c, \infty)$ and thus

$$|\nabla P_t f|^2 \leq e^{-2ct}P_t(|\nabla f|^2).$$

Proof. Recall, by the definition of Γ_2 in the general case,

$$\Gamma_2(f,g) = \frac{1}{2}[L\Gamma(f,g) - \Gamma(f,Lg) - \Gamma(Lf,g)],$$

and so for our triple,

$$\Gamma_2(f) = \frac{1}{2}[L(\nabla f)^2 - \nabla f \cdot \nabla(Lf)].$$

Likewise by Example 3.33, for our triple,

$$\Gamma_2(f) = |\nabla^2 f|^2 + \nabla^2 W(\nabla f, \nabla f),$$

so we have the equality

$$\frac{1}{2}[L(\nabla f)^2 - \nabla f \cdot \nabla(Lf)] = |\nabla^2 f|^2 + \nabla^2 W(\nabla f, \nabla f).$$

But $|\nabla^2 f|^2 \geq 0$ and by assumption $\nabla^2 W \geq c|x|^2$, so the identity

$$\Gamma_2(f) = \frac{1}{2}[L(\nabla f)^2 - \nabla f \cdot \nabla(Lf)] \geq c(\nabla f)^2 = c\Gamma(f)$$

holds $\forall f \in \mathcal{A}$. Thus L satisfies curvature-dimension condition $CD(c,\infty)$. By Lemma 3.35, it follows that

$$|\nabla P_t f|^2 \leq e^{-2ct} P_t(|\nabla f|^2).$$

□

3.5. **Concentration from semigroup methods.** For our concentration results we are interested in the good W-tuple introduced in Example 3.15, $(E, \mathcal{F}, \mu, L, \mathcal{A})$, where

$$L = \Delta - \nabla W \cdot \nabla, \;\; \mu(dx) = e^{-W} dx, \;\; \mathcal{A} = C_0^\infty(E),$$

and thus $\Gamma(f,g) = \nabla f \cdot \nabla g$. We require that the potential W satisfies the uniform convexity condition (3.8) and thus μ is log-concave in a rather strong sense.

Additionally, by Example 3.36, L satisfies the curvature condition $CD(c,\infty)$.We are now ready to prove our main concentration result.

Theorem 3.37. *Let $d\mu = e^{-W}dx$ be a probability measure on the Borel sets of $\mathbb{R}^n$ such that μ is the symmetric invariant measure with respect to the generator $L = \Delta - \nabla W \cdot \nabla$. Suppose further that W is smooth and $\nabla^2(W(x) - c|x|^2/2) \geq 0$, so that L satisfies curvature dimension condition $CD(c,\infty)$. Then, for every bounded 1-Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$, and $\forall\, r \geq 0$,*

$$\mu\left(\left\{F \geq \bar{F} + r\right\}\right) \leq e^{-cr^2/2}, \;\; \text{where } \bar{F} := \int_{\mathbb{R}^n} F(x)\mu(dx).$$

Proof. This proof expands on the proof of Proposition 2.17 in [6, p 39]. Let P_t be the semigroup generated by L and let $F : \mathbb{R}^n \to \mathbb{R}$ be a bounded mean-zero 1-Lipschitz function. We suppose that F is sufficiently smooth for what follows. Let $\lambda \geq 0$. As F is sufficiently smooth and 1-Lipschitz, then $|\nabla F| \leq 1$ almost everywhere. We deduce from (3.5b) that

$$\|P_t F\|_\infty \leq \|F\|_\infty.$$

Set

$$G_t(x) := e^{\lambda P_t(x)},$$

so G_t is bounded. Let

$$\Psi(t) := \int_{\mathbb{R}^n} G_t d\mu \int_{\mathbb{R}^n}, \quad t \geq 0.$$

The ergodicity of P_t implies that

$$P_t F \to \bar{F} = 0 \text{ in } \mathbb{L}^2(\mathbb{R}^n, \mu),$$

and therefore in probability, and thus $G_t(x) = e^{\lambda G_t F(x)} \to 1$ in probability. The family $G_t(x)$ is uniformly bounded and thus uniformly integrable so $e^{\lambda P_t F(x)} \to 1$ in L^1. In particular,

$$\lim_{t\to\infty} \Psi(t) = 1.$$

Continuing, as $|\nabla F| \leq 1$ a.e., by Example 3.36,

$$|\nabla P_t F|^2 \leq e^{-2ct}.$$

So, for all $t \geq 0$,

$$\begin{aligned}
\Psi(t) = 1 - (1 - \Psi(t)) = 1 - (\Psi(\infty) - \Psi(t)) = 1 - \int_t^\infty \Psi'(s) ds \\
= 1 - \int_t^\infty \left(\int_E e^{\lambda P_s F} d\mu \right)' ds \\
= 1 - \int_t^\infty \left(\int_E \lambda L P_s F e^{\lambda P_s F} d\mu \right) ds \\
= 1 - \int_t^\infty \left(-\lambda \int_E \Gamma(P_s F, P_s F) d\mu \right) ds \\
= 1 + \lambda^2 \int_t^\infty \int_E |\nabla P_s F|^2 e^{\lambda P_s F} d\mu \, ds \\
\leq 1 + \lambda^2 \int_t^\infty e^{-2cs} \Psi(s) \, ds.
\end{aligned}$$

By Gronwall's Lemma,

$$\Psi(t) \leq 1 \cdot \exp\left[\lambda^2 \int_t^\infty e^{-2cs} ds \right] ds = e^{\lambda^2/2c}.$$

Thus,

$$\Psi(0) = \int e^{\lambda F} d\mu \leq e^{\lambda^2/2c}.$$

The result, along with the extension to non mean-zero bounded 1-Lipschitz functions, follows from Proposition 1.13. □

By Proposition 1.8, we can use the above deviation inequality to establish normal concentration of measure, $\alpha_\mu(r) \leq e^{-cr^2/8}$, with c the curvature constant.

Corollary 3.38. *The Gaussian measure γ_n on $\mathbb{R}^n$ exhibits normal concentration. Namely, $\forall\, F : \mathbb{R}^n \to \mathbb{R}$, and $\forall\, r > 0$,*

$$\gamma_n \left(\left\{ F \geq \bar{F} d\gamma_n + r \right\} \right) \leq e^{-r^2/2}.$$

Proof. Up to a normalization constant, $\gamma_n(dx) = e^{-|x|^2/2} dx$. Thus the result follows from Theorem 3.37 applied to

$$W(x) = |x|^2/2,$$

and then finally by Proposition 1.8. □

4. CONCENTRATION THROUGH SPECTRUM

We now investigate the functional inequality known as the Poincaré inequality and demonstrate when a Poincaré inequality is satisfied, and how this inequality implies exponential concentration for a large class of measures: the log-concave measures. This expands our concentration results from those in the previous section considerably, as we can drop the uniform convexity assumption. Results are generally expanded upon from [3, Chap 4], although the discussion of log-concave measures also draws from [15].

4.1. **Poincaré Inequalites.** Throughout this discussion, we will be working in the context of a good measured space $(E, \mathcal{F}, \mu)$, a nice algebra of functions $\mathcal{A}$, and a diffusion generator $L : \mathcal{A} \subset \mathbb{L}^2(E, \mu) \to \mathbb{L}^2(E, \mu)$ whose closure in $\mathbb{L}^2(E, \mu)$ extends to a self-adjoint operator with domain $D(L)$. We also assume that μ is an invariant symmetric measure for L. To this set-up, Γ will be the associated Carré-du-champ operator $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$,

$$2\Gamma(f, g) = L(fg) - fLg - gLf.$$

We refrain from considering a good Markov tuple as it will not necessarily be clear that L generates a Markov semigroup, instead of potentially a *sub-Markov* semigroup, which satisfy every condition on a Markov semigroup, except that $P_t \boldsymbol{I}_E \leq \boldsymbol{I}_E$ instead of $P_t \boldsymbol{I}_E = \boldsymbol{I}_E$. These assumptions will be implicit in the following results.

Definition 4.1. The *Dirichlet form*, or *energy*, is the operator $\mathcal{E} : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$,

$$\mathcal{E}(f, g) = \int_E \Gamma(f, g)\, d\mu = -\int_E fLg\, d\mu,$$

where the second equality is the integration by parts formula known for Γ. We write $\mathcal{E}(f, f) = \mathcal{E}(f)$. For the domain of the Dirichlet form, we write $D(\mathcal{E})$. Most results will only be stated for $\mathcal{A}$, which by construction is dense in $D(\mathcal{E})$. □

We next recall the traditional definition of the variance of a function in $\mathbb{L}^2(\nu)$ for a probability measure ν.

Definition 4.2. For a probability measure ν on a good measurable space $(E, \mathcal{F})$, we define the *variance* of a function $f \in \mathbb{L}^2(E, \nu)$ as

$$\mathrm{Var}_\nu(f) = \int_E f^2 d\nu - \left(\int_E f d\nu \right)^2.$$ □

These two definitions immediately lead us to the critical definition of a Poincaré inequality:

Definition 4.3. The probability measure μ is said to satisfy a *Poincaré inequality* $P(C)$ with respect to L for some $C > 0$, if $\forall\, f : E \to \mathbb{R} \in D(\mathcal{E})$,

$$\mathrm{Var}_\mu(f) \leq C\, \mathcal{E}(f).$$

The best constant $C > 0$ for which such an inequality holds is called the *Poincaré constant* (relative to X). We can equivalenty say that μ satisfies the Poincaré inequality $P(C)$ with respsect to the Carré-du-champ operator Γ. □

Poincaré inequalities are equivalently called *Spectral gap inequalities* for the following reason. Assume X satisfies a Poincaré inequality $P(C)$. If f is an eigenfunction of $-L$ with eigenvalue λ, then

$$\int_E \lambda f\, d\mu = \int_E -Lf\, d\mu = 0 \implies \int_E f\, d\mu = 0.$$

Likewise, by the Poincaré inequality $P(C)$,

$$\int_E f^2\,d\mu = \mathrm{Var}_\mu(f) \le C\,\mathcal{E}(f) = C\int_E f(-Lf)\,d\mu = C\lambda\int_E f^2\,d\mu.$$

Thus,

$$1 \le C\lambda \implies \lambda \ge \frac{1}{C},$$

and so every non-zero eigenvalue of $-L$ is greater than $1/C$. Consequently, even when the spectrum of $-L$ is not discrete, it is still contained in $\{0\} \cup [1/C, \infty)$. $P(C)$ then describes a *gap* in the spectrum of $-L$.

4.2. **The case of Log-concave measures.** A natural question after the above definitions is, "Which measures satisfy Poincaré, or spectral gap, inequalities?" We answer this question by addressing a large class of measures, the log-concave measures.

Definition 4.4. A probability measure μ on Borel subsets of $\mathbb{R}^n$ is called *log-concave* if it is of the form

$$\mu(dx) = e^{-W}dx,$$

where W is a lower-bounded smooth convex function. For such a function, $\nabla^2 W \ge 0$. □

Note, log-concave measures are called so because $\log(e^{-W}) = -W$ is a concave function if e^{-W} is the density of a log-concave measure. The following theorem is the main result of this section, as it will imply concentration of measure for log-concave measures.

Theorem 4.5. *Let $E = \mathbb{R}^n$, $\mathcal{F}$ be the Borel sigma-algebra on $\mathbb{R}^n$, $\mathcal{A} = C_0^\infty(\mathbb{R}^n)$, and let $\mu(dx) = e^{-W}dx$ be a log-concave Borel probability measure on $\mathbb{R}^n$. Then μ satisfies a Poincaré inequality with respect to the generator $L = \Delta - \nabla W \cdot \nabla$ (or equivalently the Carré-du-champ operator $\Gamma(f,g) = \nabla f \cdot \nabla g$).*

Before we can prove this result, we need a collection of lemmas and a few new definitions. The first of which is a local Poincaré inequality.

Definition 4.6 (local Poincaré inequality)**.** Consider the usual set up with a good probability-measured space $(E, \mathcal{F}, \mu)$, algebra $\mathcal{A}$, and generator L. Let $K \subset E$. Then we say μ satisfies a *Poincaré inequality on K* (with respect to L) if

$$\exists C_K > 0,\ \ \forall f \in \mathcal{A}:\ \int_K \left(f - \bar{f}_K\right)^2 d\mu \le C_K \int_K \Gamma(f)\,d\mu,\ \ \bar{f}_K := \frac{1}{\mu(K)}\int_K f\,d\mu. \tag{4.1}$$

□

Lemma 4.7. *Let $E = \mathbb{R}^n$, $\mathcal{F}$ be the Borel sigma-algebra on $\mathbb{R}^n$, $\mathcal{A} = C_0^\infty(\mathbb{R}^n)$, and let $\mu(dx) = e^{-W}dx$ be a log-concave Borel probability measure on $\mathbb{R}^n$. Then μ satisfies a Poincaré inequality on the open ball $B_R := B_R(0)$, with respect to the generator $L = \Delta - \nabla W \cdot \nabla$ (or equivalently the Carré-du-champ operator $\Gamma(f,g) = \nabla f \cdot \nabla g$).*

Proof. Let $p \in [1, \infty)$ and denote $L^{1,p}(B_R)$ as the closure of $C^\infty(\overline{B_R})$ in $L^p(B_R)$ with respect to the norm $||u||_{1,p} := ||u||_{L^p} + ||\nabla u||_{L^p}$. We will show that

$$\exists C_R > 0:\ \int_{B_R} |u - \bar{u}|^p\,dx \le C_R \int_{B_R} |\nabla u|^p\,dx,\ \ \forall u \in C^\infty(\mathbb{R}^n).$$

Note that it suffices to prove this inequality for the function $v = u - \bar{u}$ whose mean on B_R is 0 and satisfies $\nabla v = \nabla u$. The inequality in this special case is proved in [9, Thm. 3.65].

The function W is locally bounded, and so $\exists R > 0$ and $0 < c_1 < c_2$ such that on B_R,

$$c_1 \leq e^{-W(x)} \leq c_2.$$

Then,

$$\begin{aligned}\int_{B_R} |\nabla u|^p e^{-W} dx &> c_1 \int_{B_R} |\nabla u|^p \, dx \geq C_R c_1 \int_{B_R} |u - \bar{u}|^p \, dx \\ &\geq \frac{C_R c_1}{c_2} \int_{B_R} |u - \bar{u}|^p e^{-W} dx.\end{aligned}$$

□

We will show that a local Poincaré inequality implies a global Poincaré inequality if there exists a particular function, called a Lyapunov function, which we now define.

Definition 4.8. For a good measured space $(E, \mathcal{F}, \mu)$ with generator L, we say that a function $J : E \to [1, \infty)$ is a *Lyapunov function* if $\exists\, \lambda, b > 0$ and $K \subset E$ measurable such that μ satisfies a local Poincaré inequality on K and

$$1 \leq -\frac{LJ}{\lambda J} + b 1_K.$$

□

Using Lyapunov functions, we can define necessary conditions for extending a local Poincaré inequality to a global one.

Lemma 4.9. *Let $E = \mathbb{R}^n$, $\mathcal{F}$ be the Borel sigma-algebra on $\mathbb{R}^n$, $\mathcal{A} = C_0^\infty(\mathbb{R}^n)$, and let $\mu(dx) = e^{-W} dx$ be a Borel probability measure on $\mathbb{R}^n$, with W smooth but not necessarily convex. Let L be a diffusion generator. If there exists a Lyapunov function J on $\mathbb{R}^n$ defined with constants λ, b and measurable subset $K \subset E$, and with μ satisfying a local Poincaré inequality $P(C_K)$ on K, then μ satisfies a Poincaré inequality $P(C)$ on the whole space $\mathbb{R}^n$ with*

$$C = \frac{1}{\lambda} + bC_k.$$

Proof. This proof expands upon the proof of Theorem 4.6.2 in [3, p. 202]. We denote by $\bar{f}_E$ the mean of f on E

$$\bar{f}_E = \int_E f(x)\mu(dx).$$

Since $\bar{f}_E$ is the $\mathbb{L}^2(\mu)$-orthogonal projection on the one-dimensional space of constant functions we deduce

$$\mathrm{Var}_\mu(f) = \|f - \bar{f}_E\|^2_{\mathbb{L}^2(\mu)} \leq \|f - c\|^2_{\mathbb{L}^2(\mu)}, \quad \forall c \in \mathbb{R}. \tag{4.2}$$

As $\mathcal{A}$ is dense in the domain of the Dirichlet form

$$\mathcal{E}(f) = \int_E \Gamma(f) d\mu,$$

it suffices to show that $\forall\, f \in \mathcal{A}$,

$$\int_E (f - c)^2 d\mu \leq C \int_E \Gamma(f) d\mu$$

for any $c \in \mathbb{R}$, which we will chose later.

To prove this inequality, we multiply the Lyapunov inequality

$$1 \leq -\frac{LJ}{\lambda J} + b 1_K$$

though by $(f-c)^2$ and integrate over E with respect to μ. This yields

$$\int_E (f-c)^2 d\mu \le -\frac{1}{\lambda}\int_E \frac{LJ}{J}(f-c)^2 d\mu + b\int_K (f-c)^2 d\mu. \tag{4.3}$$

Choose $c = \bar{f}_K$ as in (4.1). We deduce

$$b\int_K (f-c)^2 d\mu \le bC_K \int_K \Gamma(f) d\mu.$$

To bound the other term on the right-hand-side of (4.3), we will show that $\forall\, g \in \mathcal{A}$,

$$-\int_E \frac{LJ}{J} g^2 d\mu \le \int_E \Gamma(g) d\mu.$$

To see this, let $g \in \mathcal{A}$ and integrate by parts (for Γ) to obtain

$$-\int_E \frac{LJ}{J} g^2 d\mu = \int_E \Gamma\left(\frac{g^2}{J}, J\right) d\mu.$$

Further, observe that

$$\Gamma\left(\frac{g^2}{J}, J\right) = \frac{2g}{J}\Gamma(g,J) - \frac{g^2}{J^2}\Gamma(J) \le \Gamma(g),$$

where the first equality is from Γ being a diffusion operator, and the inequality following from the identity

$$\Gamma(g) + \frac{g^2}{J^2}\Gamma(J) - \frac{2g}{J}\Gamma(g,J) = \Gamma(g-J) \ge 0.$$

The proof is then concluded by letting $g = f - c$ and observing that $\Gamma(f-c) = \Gamma(f)$, so

$$\mathrm{Var}_\mu(f) \overset{(4.2)}{\le} \int_E (f-c)^2 d\mu \le -\frac{1}{\lambda}\int_E \frac{LJ}{J}(f-c)^2 d\mu + b\int_K (f-c)^2 d\mu$$

$$\le \frac{1}{\lambda}\int_E \Gamma(f) d\mu + bC_k \underbrace{\int_K \Gamma(f) d\mu}_{\ge 0} \le \left(\frac{1}{\lambda} + bC_k\right)\int_E \Gamma(f) d\mu,$$

giving us a Poincaré inequality for μ with constant

$$C = \frac{1}{\lambda} + bC_k.$$

□

Thus to show that a log-concave measure satisfies a Poincaré inequality, we only have to construct a Lyapunov function for the measure. However, to do this we need the following technical lemma.

Lemma 4.10. *If $V : \mathbb{R}^n \to \mathbb{R}$ is differentiable, convex and*

$$\int_{\mathbb{R}^n} e^{-V} dx < \infty,$$

then

$$\exists\, \alpha, R > 0 \,:\, \forall\, |x| \ge R,\; x \cdot \nabla V(x) \ge \alpha |x|.$$

Proof. This conclusion is inspired by the proof of Lemma 2.2 in [15]. First, observe that $\forall x \in \mathbb{R}^n$, $x \cdot \nabla V(x) \geq V(x) - V(0)$. This is true as $t \to g(t) := V(tx)$ is convex, so

$$g'(1) \leq \frac{g(0) - g(1)}{0 - 1}.$$

Observing that $g(0) = V(0)$ and $g(1) = V(x)$, rearranging the above inequality yields

$$V(0) \geq V(x) - 1(x \cdot \nabla V(tx))\Big|_{t=1},$$

and so

$$x \cdot \nabla V(x) \geq V(x) - V(0).$$

We now show that $\exists\, \alpha, R > 0$ such that for $|x| \geq R$, $V(x) - V(0) \geq \alpha|x|$, which will conclude the proof.

To see this, choose some $K > V(0) + 1$. Let

$$A_k := \left\{ x \in \mathbb{R}^n \,:\, V(x) \leq K \right\},$$

and observe that this level set has non-empty interior as V is continuous, so it contains $B_r(0)$ for some $r > 0$. The set A_k is closed by definition, and further, $\mathrm{vol}(A_k) < \infty$, as for Lebesgue measure λ,

$$\mathrm{vol}(A_k) = \int_{A_k} e^V \mu(dx) \leq e^K \int_{A_k} \mu(dx) < \infty,$$

Without loss of generality, we can assume that $0 \in A_K$. Fix $r > 0$ such that $B_r(0) \subset A_K$. Let $a \in A_K \setminus B_r(0)$ and let C_a be the convex hull of $\{a\} \cup B_r(0)$. Observe that $\mathrm{vol}(C_a)$ depends only on $|a|$, and that $\mathrm{vol}(C_a) \to \infty$ as $a \to \infty$. However, $\mathrm{vol}(C_a) \leq \mathrm{vol}(A_k) < \infty$, so $\sup |a|_{a \in A_K} < \infty$ and thus A_K is bounded. Thus $\exists\, R > 0$ such that $A_K \subset B_{R-1}(0)$.

Now, let $u \in \mathbb{R}^n$ such that $|u| = R$. Then $u \notin A_K$, so $V(u) - V(0) \geq 1$. But as V is convex,

$$t \to \frac{V(tu) - V(0)}{t}$$

is non-decreasing. So, for $|x| \geq R$,

$$V(x) - V(0) \geq \frac{|x|}{R}.$$

Choosing $\alpha = 1/R > 0$ concludes the proof. $\square$

Theorem 4.11. *Let $\mu(dx) = e^{-W} dx$ be a log-concave Borel probability measure on $\mathbb{R}^n$. Then μ satisfies a Poincaré inequality with respect to the usual set up, where*

$$L = \Delta - \nabla W \cdot \nabla.$$

Proof. This proof follows the strategy of proof of Theorem 1.4 in [15, p 63]. By Lemma 4.9, it suffices to show that μ has a Lyapunov function. Fix $\gamma, R > 0$. Let $K = B_R(0)$. By Lemma 4.7, μ restricted to $B_R(0)$ satisfies a local Poincaré inequality with some constant C_K. So, let

$$J(x) = e^{\gamma |x|}, \quad |x| \geq R,$$

i.e., on $(B_R(0))^c$. On $B_R(0)$, let $J(x) \geq 1$ so that J is continuous and smooth on all of $\mathbb{R}^n$. We claim that J is a Lyapunov function for μ. We know that μ satisfies a local Poincaré inequality on K, so it suffices to show that $\exists\, \lambda, b > 0$ such that

$$1 \leq -\frac{LJ}{\lambda J} + b 1_K.$$

To see this, recall that $L = \Delta - \nabla W \cdot \nabla$ and observe that

$$LJ(x) = \left(\frac{\gamma(n-1)}{|x|} + \gamma^2 - \frac{x \cdot \nabla V(x)}{|x|} \right) J(x), \quad |x| \geq R,$$

where it is well known that $\Delta|x| = (n-1)/|x|$ and $\nabla|x| = x/|x|$. Then, by Lemma 4.10, $\exists\, \alpha > 0$ such that $(x \cdot \nabla V(x)) \geq \alpha/|x|$, and so

$$LJ(x) \leq \left(\frac{\lambda(n-1)}{R} + \lambda^2 - \lambda\alpha \right) J(x), \quad |x| \geq R.$$

On the bounded set $B_R(0)$, $J(x) \geq 1$ and is less than $J(y)$, $\forall\, y \in (B_R 0)^c$, as J is convex on $|x| \geq R$. Thus, we can find some $\beta > 0$ such that

$$LJ(x) \leq \beta, \quad |x| < R,$$

so that

$$LJ(x) \leq \left(\frac{\lambda(n-1)}{R} + \lambda^2 - \lambda\alpha \right) J(x) + \beta 1_K(x), \quad x \in \mathbb{R}^n,$$

and after rearranging,

$$LJ(x) \leq -\lambda \left(\frac{-(n-1)}{R} - \lambda - \alpha \right) J(x) + \beta 1_K(x), \quad x \in \mathbb{R}^n.$$

Taking $R > 0$ as large as we need, we can find some sufficiently small $\lambda > 0$ such that

$$\theta = \lambda(\alpha - \lambda - (n - 1/R)) > 0.$$

Thus,

$$LJ \leq -\theta J + \beta 1_K.$$

Finally, we can rearrange this equality to show that J is a Lyapunov function. We have

$$-LJ \geq \theta J - \beta 1_K$$

so

$$\frac{-LJ}{\theta J} \geq 1 - \frac{\beta}{\theta J} 1_K \geq 1 - \frac{\beta}{\theta} 1_K,$$

as $J \geq 1$ on K. Taking $b = \beta/\theta > 0$ yields

$$\frac{-LJ}{\theta J} + b 1_K \geq 1,$$

and so J is a suitable Lyapunov function with set $K = B_R(0)$ and constants $b, \theta > 0$. Thus by Lemma 4.9, μ satisfies a global Poincaré inequality with constant

$$C = \frac{1}{\theta} + bC_k.$$

□

4.3. **Concentration from Poincaré inequalities.** Let us now prove the main result of this section, which shows that log-concave measures exhibit (at least) exponential concentration.

Proposition 4.12. *Let $(X, \mathcal{F}, \mu)$ be a good measured space, with (X, d) being a metric space as well. Assume that μ is a probability measure that satisfies a Poincaré inequality $P(C)$ with respsect to the Carré-du-champ operator $\Gamma(f) = |\nabla f|^2$. Then,*

$$\alpha_\mu(r) \leq e^{-r/3\sqrt{C}}, \quad r \geq 0.$$

Proof. This proof expands upon the proof of Theorem 3.1 in [6, p 48]. Fix some $\epsilon > 0$, and let $A, B \subset X$ open such that $d(A, B) = \epsilon$. For convenience, we write $a := \mu(A)$ and $b := \mu(B)$. Consider the function $f : X \to \mathbb{R}$, where

$$f(x) = \left(1 - \frac{\min(d(x, A), \epsilon)}{\epsilon}\right) \frac{1}{a} - \left(\frac{\min(d(x, A), \epsilon)}{\epsilon}\right) \frac{1}{b}.$$

Essentially, f is a sort of smooth bump function, taking value $1/a$ on A, smoothly transitioning to $1/b$ on A_r as $r \to \epsilon$, and taking value $1/b$ on A_ϵ^c. f is clearly smooth, and by subtracting out $\alpha = \int_X f\, d\mu$, we can assume without loss of generality that f is mean-zero as well. f is smooth and with bounded derivative is thus Lipschitz, with $\nabla f = 0$ on $A \cup B$, and

$$|\nabla f| \leq \frac{1}{\epsilon}\left(\frac{1}{a} + \frac{1}{b}\right)$$

almost everywhere. Thus,

$$\begin{aligned} \int_X \Gamma(f)\, d\mu = \int_X |\nabla f|^2\, d\mu &= \int_{(A \cup B)^c} |\nabla f|^2\, d\mu \\ &\leq \frac{1}{\epsilon}\left(\frac{1}{a} + \frac{1}{b}\right)(1 - [\mu(A) + \mu(B)]) \\ &= \frac{1}{\epsilon}\left(\frac{1}{a} + \frac{1}{b}\right)(1 - a - b), \end{aligned}$$

and so $f \in D(\mathcal{E})$, allowing us to use f in the Poincaré inequality $P(C)$. On the other hand,

$$\begin{aligned} \mathrm{Var}_\mu(f) &= \int_X \left(f - \int f d\mu\right)^2 d\mu = \int_X f^2\, d\mu \\ &\geq \int_{A \cup B} f^2\, d\mu = \int_A f^2\, d\mu + \int_B f^2\, d\mu \\ &= \frac{\mu(A)}{a^2} + \frac{\mu(B)}{b^2} = \frac{1}{a} + \frac{1}{b}. \end{aligned}$$

Therefore, by the Poincaré inequality applied to f,

$$\frac{1}{a} + \frac{1}{b} \leq \mathrm{Var}_\mu(f) \leq C \int_X \Gamma(f)\, d\mu \leq \frac{C}{\epsilon^2}\left(\frac{1}{a} + \frac{1}{b}.\right)^2 (1 - a - b),$$

which implies that

$$\frac{\epsilon^2}{C} \leq \left(\frac{1}{a} + \frac{1}{b}\right)(1 - a - b) \leq \frac{1 - a - b}{ab}.$$

Separating a and b yields

$$\begin{aligned} \frac{\epsilon^2}{C} \leq \frac{1 - a - b}{ab} \implies \frac{ab\epsilon^2}{C} + b \leq 1 - a \implies \\ b\left(\frac{a\epsilon^2}{C} + 1\right) \leq 1 - a \implies \\ b \leq \frac{1 - a}{1 + \epsilon^2 a/C}. \end{aligned}$$

Now, we choose a specfic set A to conclude the proof. Let $A = (B_\epsilon)^c$. Assume that $\mu(B_\epsilon) = 1 - a \leq 1/2$. Note that this implies that $\mu(A) = a \geq 1/2$. Then,

$$b \leq \frac{\mu(B_\epsilon)}{1 + \epsilon^2 a/C} \implies (1 + \epsilon^2 a/C)\mu(B) \leq \mu(B_\epsilon),$$

and consequently

$$(1+\epsilon^2/2C)\mu(B) \le (1+\epsilon^2 a/C)\mu(B) \le \mu(B_\epsilon).$$

We now turn to the notation of Expansion coefficients, which we recall from Chapter 2. By the above inequality, it follows that

$$\mathrm{Exp}_\mu(\epsilon) \ge 1 + \frac{\epsilon^2}{2C} > 1.$$

Finally, let us choose a specific $\epsilon > 0$ such that $\epsilon^2/C = 2$. Then, by Proposition 1.11, for $r > 0$,

$$\alpha_\mu(r) \le \frac{(1+\epsilon^2/2C)}{2} \exp\left[-r\log\left(1+\epsilon^2/2C\right)/\epsilon\right] \tag{4.4}$$

$$\le 1\cdot\exp\left[-r\log(2)/\sqrt{2C}\right] \tag{4.5}$$

$$\le \exp\left[-r/3C\right], \tag{4.6}$$

as desired. Note, the constant 3 in the concluding inequality is far from sharp. □

5. An application: the Johnson-Lindenstrauss Flattening Lemma

The rest of this thesis is devoted to stating and proving the Johnson-Lindenstrauss lemma. In brief, the lemma states that if we randomly choose a k-dimensional subspace of $\mathbb{R}^n$ to nicely project N vectors $a_1, ..., a_N \in \mathbb{R}^n$ onto, the distances between the projected vectors will be very close to the distances between the same original vectors.

This is a fascinating result, and it has many applications in imaging and compression. Namely, if we interpret our vectors $a_1, ..., a_N$ as slices of an image, then instead of storing the information these vectors contain in the costly high-dimensional $\mathbb{R}^n$, we can instead store them in a lower dimensional subspace of $\mathbb{R}^n$, which in practice is much cheaper in terms of memory and recovery time.

The proof of the flattening lemma has at its core the concentration of measure phenomenon for the Gaussian measure on $\mathbb{R}^n$, the result which we have been developing throughout the entirety of this thesis. Consequently this result is not only fascinating, it is also a great example of the utility of the concentration of measure phenomenon, as well as the power of the Gaussian measure. This section expands upon the results on pages 5-12 of [11].

5.1. **Concentration Lemmas.** We first must develop a number of technical lemmas, which are a direct consequence of the Laplace bound method from section 1.3 and the normal concentration of the Gaussian measure on $\mathbb{R}^n$. Recall, first, the example of Laplace bounds, Lemma 1.14.

Lemma 5.1. *For any $\delta \ge 0$,*

$$\gamma_n\left(\left\{x\in\mathbb{R}^n : ||x||^2 \ge n+\delta\right\}\right) \le \left(\frac{n}{n+\delta}\right)^{-n/2} e^{-\delta/2},$$

and

$$\gamma_n\left(\left\{x\in\mathbb{R}^n : ||x||^2 \le n-\delta\right\}\right) \le \left(\frac{n}{n-\delta}\right)^{-n/2} e^{\delta/2}.$$

From this result we immediately have a more powerful corollary, which we will reference throughout the proof of the Johnson-Lindenstrauss flattening lemma.

Corollary 5.2. *For any $\epsilon \in (0,1)$,*

$$\gamma_n\left(\left\{x\in\mathbb{R}^n : ||x||^2 \ge \frac{n}{1-\epsilon}\right\}\right) \le e^{-\epsilon^2 n/4}, \text{ and}$$

$$\gamma_n\left(\left\{x\in\mathbb{R}^n : ||x||^2 \le (1-\epsilon)n\right\}\right) \le e^{-\epsilon^2 n/4}.$$

Proof. Take $\delta = n\epsilon/(1-\epsilon)$. Then $n + \delta = n + n\epsilon(1-\epsilon) = n/(1-\epsilon)$. By Lemma 5.1,

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq n + \delta\right\}\right) \leq \left(\frac{n}{n+\delta}\right)^{-n/2} e^{-\delta/2},$$

so plugging in $n + \delta = n/(1-\epsilon)$ yields

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq n + \frac{n}{1-\epsilon}\right\}\right) \leq (1-\epsilon)^{-n/2} \exp\left[\frac{-n\epsilon}{2(1-\epsilon)}\right].$$

$(1-\epsilon)^{-n/2} = \exp[-(n/2)\log(1-\epsilon)]$, so rearranging the above inequality yields

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq n + \frac{n}{1-\epsilon}\right\}\right) \leq \exp\left[\left(-\frac{n}{2}\right)\left(\frac{\epsilon}{1-\epsilon} + \log(1-\epsilon)\right)\right].$$

As $\epsilon \in (0,1)$, we can take the power series expansions of $\epsilon/(1-\epsilon)$ and $\log(1-\epsilon)$, which are as follows:

$$\frac{\epsilon}{1-\epsilon} = \sum_{k=1}^{\infty} \epsilon^k,$$

$$\log(1-\epsilon) = \sum_{k=1}^{\infty} -\frac{\epsilon^k}{k}.$$

Summing the two convergent series,

$$\begin{aligned} \frac{\epsilon}{1-\epsilon} + \log(1-\epsilon) &= \epsilon^2 - \epsilon^2/2 + \epsilon^3 - \epsilon^3/3 + -... \\ &\geq \epsilon^2/2, \end{aligned}$$

and thus

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq n + \frac{n}{1-\epsilon}\right\}\right) \leq \exp\left[\left(-\frac{n}{2}\right)\left(\frac{\epsilon^2}{2}\right)\right] \leq e^{-n\epsilon^2/4},$$

as desired. The other result follows similarly from the second conclusion of Lemma 5.1,

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \leq n - \delta\right\}\right) \leq \left(\frac{n}{n-\delta}\right)^{-n/2} e^{\delta/2},$$

using $\delta = n\epsilon$. □

5.2. **Push-forwards of the Gaussian.** Next, we show that the Gaussian measure behaves very nicely when composed with various projections. First, we consider orthogonal projections. Let us recall what an orthogonal projection is.

Definition 5.3 (Orthogonal projection)**.** On $\mathbb{R}^n$, an *orthogonal projection* is a linear map $P : \mathbb{R}^n \to \mathbb{R}^n$ such that $P^2 = P$ and, in matrix form, $P^T = P$. An orthogonal projection onto a subspace $U \subset \mathbb{R}^n$ is the projection P_U, where the image of P_U is U. □

The property that $P^2 = P$ is the definition of a general projection, and the property that $P = P^T$ indicates that orthogonal projections are the projections on $\mathbb{R}^n$ that minimize the distance between vectors and their projections. This is equivalent to the property that for an orthogonal projection P_W and a vector $x \in \mathbb{R}^n$, $P(x) - x$ is orthogonal to $y, \forall\, y \in W$. □

Example 5.4. If $U \subset \mathbb{R}^n$ is a k-dimensional subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, ..., u_k$, then the orthogonal projection onto U is given by

$$P_U = \sum_{i=1}^{k} (u_i, \cdot) u_i,$$

where $(\cdot, \cdot)$ is the canonical inner product on $\mathbb{R}^n$. For $e_1, ..., e_n$, the canonical basis of $\mathbb{R}^n$, if U is the span of $e_1, ..., e_k$ for some $k < n$, then $\forall\, (x_1, ..., x_n) \in \mathbb{R}^n$, $P_U(x_1, ..., x_n) = (x_1, ..., x_k, 0, ..., 0)$ is the orthogonal projection onto U. □

Now, let us recall what a push-forward measure is.

Definition 5.5. Let $(X, \mathcal{C}, \mu)$ be a measured space and let $(Y, \mathcal{D})$ be a measurable space. For a measurable mapping $\phi : X \to Y$, the *push-forward measure* $f_{\#}\mu$ on $(Y, \mathcal{D})$ is the measure

$$f_{\#}\mu(D) = \mu(f^{-1}(D)), \quad D \in \mathcal{D}.$$

□

Likewise, recall what invariance for a measure is.

Definition 5.6. For a measured space $(X, \mathcal{F}, \mu)$ and a measurable mapping $F : X \to X$, we say that μ is *invariant by* F if $\forall\, A \in \mathcal{F}$, $\mu(A) = \mu(F(A))$, where by $F(A)$ we mean the set $\{F(x) \,:\, x \in A\}$. □

A useful example of measure invariance comes in the following result, which is key in the proof of the flattening lemma.

Proposition 5.7. *For a k-dimensional subspace $L \subset \mathbb{R}^n$ with orthogonal projection $P : \mathbb{R}^n \to L$, the push forward of the standard Gaussian measure γ_n on $\mathbb{R}^n$ by P is the standard Gaussian measure γ_k on L, with density* $(2\pi)^{-k/2} \exp[-||x||^2/2]$, $x \in L$.

Proof. First, observe that the Gaussian measure on $\mathbb{R}^n$ is invariant under orthogonal transformations of $\mathbb{R}^n$. To see this, recall that for an orthogonal transformation $O : \mathbb{R}^n \to \mathbb{R}^n$, $||O(x)|| = ||x||$ for $x \in \mathbb{R}^n$. The Gaussian measure is computed through its density, which depends only on $||x||$, so

$$\gamma_n(O(A)) = (2\pi)^{-n/2} \int_A e^{-||O(x)||^2/2} dx$$
$$= (2\pi)^{-n/2} \int_A e^{-||x||^2/2} dx = \gamma_n(A), \quad A \subset \mathbb{R}^n.$$

Thus if necessary, we can have an orthogonal transformation act on L, so that without loss of generality,

$$L = (x_1, ..., x_k, 0, ..., 0), \quad x_k \in \mathbb{R},$$

where the last $n - k$ coordinates are 0. Let $A \subset L$ and write $B = P^{-1}(A) \subset \mathbb{R}^n$. Then,

$$P_{\#}\gamma_n = \gamma_n(B) = (2\pi)^{-n/2} \int_B \exp\left[-(x_1^2 + ... + x_n^2)/2\right] dx_1 ... dx_n$$
$$= (2\pi)^{-n/2} \int_A e^{\left(-(x_1^2 + ... + x_k^2)/2\right)} dx_1 ... dx_k \times \prod_{i=k+1}^{n} (2\pi)^{-n/2} \int_{\mathbb{R}} e^{-x_i^2/2} dx_i$$
$$= \gamma_k(A) \cdot 1 \cdot ... \cdot 1 = \gamma_k(A).$$

□

Remark 5.8. We can consider another push-forward of the Gaussian measure γ_n, this time by the *radial projection* $\phi : \mathbb{R}^n \setminus 0 \to \mathbb{S}^{n-1}$, $\phi(x) = x/||x||$. Then, the push-forward $\phi_\# \gamma_n$ is the uniform probability measure on $\mathbb{S}^{n-1}$, μ_n. Indeed, by the above proposition, the Gaussian measure is invariant by orthogonal transformations, which include rotations, so the Gaussian measure is rotation-invariant. It is easy to see that push-forward measures preserve invariance, so $\phi_\# \gamma_n$ must be rotation invariant as well. There is a unique Borel probability measure on $\mathbb{S}^{n-1}$ that is invariant by rotations, μ_n, so this measure must be the push-forward of the Gaussian by ϕ.

We will not prove that the unit sphere $\mathbb{S}^{n-1}$ has a unique rotation invariant Borel probability measure μ, but instead we describe how to sample a point with μ_n: sample $x \in \mathbb{R}^n$ randomly by γ_n, then project radially, $x \to x/||x||$. Note that $x \neq 0$ with probability one. □

Using these results, we can prove a precursor to the Johnson-Lindenstrauss flattening lemma. The next result shows that if we fix a subspace $L \subset \mathbb{R}^n$ and then randomly choose a vector $x \in \mathbb{R}^n$, the orthogonal projection of x onto L, x_L, will have length very close to the length of x, i.e. $||x_L||$ will not differ much from $||x||$. This is a sort of dual to the actual flattening lemma, in which we fix vectors and then choose a subspace randomly. However, we will show later on that the two problems are in fact equivalent.

Proposition 5.9. *Let $L \subset \mathbb{R}^n$ be a k-dimensional subspace of $\mathbb{R}^n$. For a vector $x \in \mathbb{R}^n$, let x_L be its orthogonal projection onto L. Then, $\forall\, \epsilon \in (0,1)$,*

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, \sqrt{\frac{n}{k}}||x_L|| \geq \frac{||x||}{(1-\epsilon)}\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}, \text{ and}$$

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, \sqrt{\frac{n}{k}}||x_L|| \geq (1-\epsilon)||x||\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}.$$

Proof. Let $\epsilon \in (0,1)$. Recall, by Corollary 5.2,

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \leq (1-\epsilon)n\right\}\right) \leq e^{-\epsilon^2 n/4},$$

so

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x||^2 \geq (1-\epsilon)n\right\}\right) \geq 1 - e^{-\epsilon^2 n/4},$$

and thus

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x| \geq \sqrt{(1-\epsilon)n}\right\}\right) \geq 1 - e^{-\epsilon^2 n/4}.$$

By Proposition 5.7, if $x \in \mathbb{R}^n$ has standard Gaussian distribution, then its projection x_L has standard Gaussian distribution γ_k on L. Thus by Corollary 5.2 applied to x_L,

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, ||x_L|| \leq \sqrt{\frac{k}{1-\epsilon}}\right\}\right) \geq 1 - e^{-\epsilon^2 k/4}.$$

If we let

$$A := \left\{x \in \mathbb{R}^n \,:\, ||x||^2 \leq (1-\epsilon)n\right\}$$
$$B := \left\{x \in \mathbb{R}^n \,:\, ||x_L||^2 \geq \frac{k}{1-\epsilon}\right\},$$

then

$$A^c := \left\{x \in \mathbb{R}^n \,:\, ||x| \geq \sqrt{(1-\epsilon)n}\right\}$$
$$B^c := \left\{x \in \mathbb{R}^n \,:\, ||x_L|| \leq \sqrt{\frac{k}{1-\epsilon}}\right\}.$$

Observe that

$$\gamma_n(A^c \cap B^c) = \gamma_n(A \cup B)^c = 1 - \gamma_n(A \cup B) \geq 1 - \gamma_n(A) - \gamma_n(B),$$

so by our above computations

$$\gamma_n(A^c \cap B^c) \geq 1 - e^{-k\epsilon^2/4} - e^{-n\epsilon^2/4}.$$

For $x \in A^c \cap B^c$,

$$||x_L|| \leq \sqrt{\frac{k}{1-\epsilon}} \implies \frac{\sqrt{n}}{\sqrt{k}}||x_L|| \leq \sqrt{\frac{n}{1-\epsilon}} \leq \frac{\sqrt{n(1-\epsilon)}}{1-\epsilon} \leq \frac{||x||}{1-\epsilon}.$$

Putting everything together, we get

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, \sqrt{\frac{n}{k}}||x_L|| \leq \frac{||x||}{1-\epsilon}\right\}\right) \geq 1 - e^{-n\epsilon^2/4} - e^{-k\epsilon^2/4},$$

and passing to the complement yields

$$\gamma_n\left(\left\{x \in \mathbb{R}^n \,:\, \sqrt{\frac{n}{k}}||x_L|| \geq \frac{||x||}{(1-\epsilon)}\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4},$$

as desired. The second computation follows similarly. □

As a corollary, we can attain the same result on the sphere $\mathbb{S}^{n-1}$, via the radial projection.

Corollary 5.10. *Let μ_n be the rotation-invariant Borel probability measure on the unit sphere $\mathbb{S}^{n-1}$ and let $L \subset \mathbb{R}^n$ be a k-dimensional subspace. For $x \in \mathbb{S}^{n-1}$, let x_L be the orthogonal projection of x onto L. Then, $\forall\, \epsilon \in (0,1)$,*

$$\mu\left(\left\{x \in \mathbb{S}^{n-1} \,:\, \sqrt{\frac{n}{k}}||x_L|| \geq \frac{||x||}{1-\epsilon}\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}, \text{ and}$$

$$\mu\left(\left\{x \in \mathbb{S}^{n-1} \,:\, \sqrt{\frac{n}{k}}||x_L|| \leq (1-\epsilon)||x||\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}.$$

Proof. For $x \in \mathbb{R}^n$, the ratio $||x_L||/||x||$ is unchanged under the radial projection $\phi(x) = x/||x||$, so the inequalities of Proposition 5.9 remain unchanged when x is replaced with $\phi(x) \in \mathbb{S}^{n-1}$, and the measure is replaced by the push-forward $\phi_{\#}\gamma_n = \mu_n$, as in our discussion in Remark 5.8. □

However, this result is still not enough to prove our final result. We must switch the assumptions of Corollary 5.10 to a fixed vector $x \in \mathbb{S}^{n-1}$ and a random $L \subset \mathbb{R}^n$ in order to prove the flattening lemma. However, like we previously mentioned, none of our results change in this switch. We discuss why that is the case now.

5.3. **Random subspaces, fixed vectors.**
Let us first define a space which will not only simplify our notation, but will allow us to extract deep results out of a familiar concept.

Definition 5.11 (Grassmannian)**.** Let $k \leq n$ and define the *Grassmannian* $G_k(\mathbb{R}^n)$ as the set of all k-dimensional subspaces of $\mathbb{R}^n$,

$$G_k(\mathbb{R}^n) := \{\, L \subset \mathbb{R}^n \,:\, L \text{ is a subspace},\ \dim L = k \,\}.$$ □

Additionally, let us recall some definitions from group theory.

Definition 5.12 (Orthogonal group)**.** We let O_n be the group of orthogonal transformations of $\mathbb{R}^n$. We can write

$$O_n := \left\{ M \in \text{Mat}(n,n,\mathbb{R}) : \; MM^T = 1 \right\},$$

where $\text{Mat}(n,n,\mathbb{R})$ is the space of $n \times n$ dimensional $\mathbb{R}$-valued matrices, and $\det(M)$ is the determinant of M. □

Definition 5.13. For a group G and a non-empty set X, we call the *orbit* of an element $x \in X$ as the set

$$Gx := \{g \cdot x : \, g \in G\}.$$

Additionally, we say that the group action of G acts *transitively* on X if there exists only one orbit, i.e. if $\exists\, x \in X \; : Gx = X$. □

Remark 5.14. We state without proving that both $G_k(\mathbb{R}^n)$ and O_n are compact smooth manifolds $\forall\, 0 \leq k \leq n$, $n \geq 1$. Additionally, both spaces can be equipped with a metric invariant under the action of the orthogonal group, which acts transitively on both spaces. As a result, which we also do not prove, both spaces are equipped with a unique Borel probability measure, invariant by orthogonal maps, i.e., invariant by O_n. On $G_k(\mathbb{R}^n)$, we will denote this measure $\mu_{n,k}$. To sample a random $A \in G_k(\mathbb{R}^n)$ with respect to $\mu_{n,k}$, sample k vectors $x_1, ..., x_k \in \mathbb{R}^n$ randomly from the standard Gaussian measure on $\mathbb{R}^n$ and let A be the subspace spanned by $x_1, ..., x_k$. In other words,

$$\mu_{n,k} = \text{span}_{\#}(\gamma_n \times ... \times \gamma_n).$$

Similarly, let ν_n be the unique O_n-invariant Borel probability on O_n. To sample a $M \in O_n$ randomly, sample n vectors $x_1, ..., x_n \in \mathbb{R}^n$, apply the Gram-Schmidt orthogonalization process to the vectors to obtain vectors $u_1, ..., u_n \in \mathbb{R}^n$ and let $M = [u_1 \, ... \, u_n]$ be the matrix with columns $u_1, ..., u_n$. Note that this holds as $x_1, ..., x_n$ are linearly independent with probability one. In other words,

$$\nu_n = \text{Gram-Schmidt}_{\#}(\gamma_n \times ... \times \gamma_n).$$ □

With these results, we can now state the final lemma necessary to prove the Johnson-Lindenstrauss flattening lemma. It essentially states that because of the O_n-invariance of ν_n and $\mu_{n,k}$, fixing a subspace and choosing a vector randomly is equivalent in probability to fixing a vector and choosing a subspace randomly.

Proposition 5.15. *Let $x \in \mathbb{R}^n \setminus 0$. For $L \in G_k(\mathbb{R}^n)$, let x_L be the orthogonal projection of x onto L. Then $\forall\, \epsilon \in (0,1)$,*

$$\mu_{n,k}\left(\left\{L \in G_k(\mathbb{R}^n) \; : \; \sqrt{\frac{n}{k}}||x_L|| \geq \frac{||x||}{1-\epsilon}\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}, \textit{ and}$$

$$\mu_{n,k}\left(\left\{L \in G_k(\mathbb{R}^n) \; : \; \sqrt{\frac{n}{k}}||x_L|| \leq (1-\epsilon)||x||\right\}\right) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}.$$

Proof. Let $x \in \mathbb{R}^n \setminus 0$. Normalizing if necessary, we can assume without loss of generality that $||x|| = 1$, i.e. that $x \in \mathbb{S}^{n-1}$. As O_n acts transitively on $G_k(\mathbb{R}^n)$, we can choose a $L_0 \in G_k(\mathbb{R}^n)$ such that the orbit of L_0 is $G_k(\mathbb{R}^n)$. Thus, applying a randomly chosen $U \in O_n$ to L_0 (with respect to ν_n) gives a random $L \in G_k(\mathbb{R}^n)$, so we can consider $\mu_{n,k}$ as the push-forward of ν_n under the map $U \to U(L_0)$. Thus,

$$\mu_{n,k}\left(\left\{L \in G_k(\mathbb{R}^n) \; : \; \sqrt{\frac{n}{k}}||x_L|| \geq \frac{1}{1-\epsilon}\right\}\right)$$

$$= \nu_n\left(\left\{U \in O_n \,:\, \sqrt{\frac{n}{k}}||x_{U(L_0)}|| \geq \frac{1}{1-\epsilon}\right\}\right),$$

and likewise,

$$\mu_{n,k}\left(\left\{L \in G_k(\mathbb{R}^n) \,:\, \sqrt{\frac{n}{k}}||x_L|| \leq (1-\epsilon)\right\}\right)$$
$$= \nu_n\left(\left\{U \in O_n \,:\, \sqrt{\frac{n}{k}}||x_{U(L_0)}|| \leq (1-\epsilon)\right\}\right).$$

Now we have to reckon with the length $||x_{U(L_0)}||$. But letting $y = U^{-1}x$, $||x_{U(L_0)}|| = ||y_{L_0}||$. O_n acts transitively on $\mathbb{S}^{n-1}$, so for a random $U \in O_n$, we get a random $y = U^{-1}x \in \mathbb{S}^{n-1}$. Thus, we can consider μ, the Borel probability measure on $\mathbb{S}^{n-1}$, as the push-forward of ν_n under the map $U \to U^{-1}x$. It follows that

$$\nu_n\left(\left\{U \in O_n \,:\, \sqrt{\frac{n}{k}}||x_{U(L_0)}|| \geq \frac{1}{1-\epsilon}\right\}\right)$$
$$= \mu\left(\left\{y \in \mathbb{S}^{n-1} \,:\, \sqrt{\frac{n}{k}}||y_{L_0}|| \geq \frac{1}{1-\epsilon}\right\}\right).$$

Likewise,

$$\nu_n\left(\left\{U \in O_n \,:\, \sqrt{\frac{n}{k}}||x_{U(L_0)}|| \leq (1-\epsilon)\right\}\right)$$
$$= \mu\left(\left\{y \in \mathbb{S}^{n-1} \,:\, \sqrt{\frac{n}{k}}||y_{L_0}|| \leq (1-\epsilon)\right\}\right).$$

The conclusion follows immediately from Corollary 5.10. □

5.4. **Proof of the J-L flattening lemma.** Now we are prepared to finally state and prove the Johnson-Lindenstrauss flattening lemma. In the below proof, ϵ will be a constant representing fidelity, or how much the projected vectors' distances will differ form the original vectors' distances, and Π will be a probability bound, indicating with what probability we would like our result to hold. In practice, these two constants are pre-specified, and we must choose a k and n such that our assumptions are satisfied.

Theorem 5.16 (Johnson-Lindenstrauss flattening lemma)**.** *Let $a_1, ..., a_N \in \mathbb{R}^n$ for some $N > 2$. Given a probability bound $\Pi > 0$ and a fidelity constant $\epsilon > 0$, choose an integer k such that*

$$\frac{N(N-1)}{2} \cdot \left(e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}\right) \leq \Pi.$$

Assuming that $k \leq n$, let $L \in G_k(\mathbb{R}^n)$ be chosen at random with respect to $\mu_{n,k}$. Let a_i' be the orthogonal projection of a_i onto L for $1 \leq i \leq N$. Then,

$$(1-\epsilon)||a_i - a_j|| \leq \sqrt{\frac{n}{k}}||a_i' - a_j'|| \leq \frac{||a_i - a_j||}{1-\epsilon}, \quad 1 \leq j < i \leq N$$

with probability at least $1 - \Pi$.

Proof. $\forall\, j < i \in [1, N]$, let $c_{ij} = a_i - a_j$ and let c_{ij}' be the orthogonal projection of c_{ij} onto L. There are $\binom{N}{2} = N(N-1)/2$ vectors c_{ij}, and for each pair (i, j), $||c_{ij}'|| = ||a_i' - a_j'||$. By Proposition 5.15, if we let

$$A_{ij} := \left\{L \in G_k(\mathbb{R}^n) \,:\, \sqrt{\frac{n}{k}}||c_{ij}'|| \geq \frac{||c_{ij}||}{1-\epsilon}\right\},$$

$$B_{ij} := \left\{ L \in G_k(\mathbb{R}^n) \,:\, \sqrt{\frac{n}{k}}||c'_{ij}|| \leq (1-\epsilon)||c_{ij}|| \right\},$$

then

$$\mu_{n,k}(A_{ij}) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4},$$

and

$$\mu_{n,k}(B_{ij}) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}.$$

Likewise,

$$\mu_{n,k}(A_{ij} \cap B_{ij}) \leq \min\Big(\mu_{n,k}(A_{ij}), \mu_{n,k}(B_{ij})\Big) \leq e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}.$$

Thus the probability of both A_{ij} and B_{ij} for all $j < i$ is bounded above by

$$\mu_{n,k}\left(\bigcup_{j<i}(A_{ij} \cap B_{ij})\right) \leq \frac{N(N-1)}{2}e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4} \leq \Pi.$$

So, the probability that the event $A_{ij}^c \cup B_{ij}^c =$

$$\left\{(1-\epsilon)||a_i - a_j|| \leq \sqrt{\frac{n}{k}}||a'_i - a'_j|| \leq \frac{||a_i - a_j||}{1-\epsilon}, \quad 1 \leq i, j \leq N\right\}$$

occurs for all pairs $j < i$ is

$$1 - \mu_{n,k}\left(\bigcup_{j<i}(A_{ij} \cap B_{ij})\right) \geq 1 - \Pi,$$

concluding the proof. □

Observe the key role that concentration of measure plays in both the propositions leading up to the Johnson-Lindenstrauss lemma, and the proof of the result itself.

Remark 5.17. In the statement of the Johnson-Lindenstrauss lemma, we can in fact choose our dimension k independently of the ambient dimension n by noting $n \geq N$ and instead dealing only with N. Namely, for $\epsilon, \Pi > 0$ fixed,

$$k > \frac{4\log(N^2/\Pi)}{\epsilon^2}$$

satisfies the condition

$$\frac{N(N-1)}{2} \cdot \left(e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}\right) \leq \Pi.$$

To see this, observe that $0 < k \leq n$, so

$$e^{-n\epsilon^2/4} \leq e^{-k\epsilon^2/4},$$

and thus

$$k > \frac{4\log(N^2/\Pi)}{\epsilon^2} \implies -k\epsilon^2/4 < -\log(N^2/\Pi) \implies e^{-k\epsilon^2/4} < \frac{\Pi}{N^2}.$$

In this case, it follows that

$$\frac{N(N-1)}{2} \cdot \left(e^{-k\epsilon^2/4} + e^{-n\epsilon^2/4}\right) \leq N(N-1)e^{-k\epsilon^2/4} < \frac{\Pi(N^2-N)}{N^2} \leq \Pi,$$

as desired. Thus, in practice, we can choose the dimension of k to be of order $\log(N)$, which is much lower than N and provides a serious computational reduction. □

REFERENCES

[1] R.B. Ash: *Probability and Measure Theory*, (with contributions from C. Doléans-Dade), 2nd Edition, Academic Press, 2000.

[2] D. Bakry: *Un critère de non explosion pour certaines diffusions sur une variété riemannienne complète* C.R. Acad. Sc. Paris, t.303, Série I, No. 1, 1986, 23-26.

[3] D. Bakry, I. Gentil, M. Ledoux: *Analysis and Geometry of Markov Diffusion Operators*, Springer Verlag,2014.

[4] F. Baudoin: *Diffusion Processes and Stochastic Calculus.* European Mathematical Society, 2014.

[5] E. Çinlar: *Probability and Stochastics*, Graduate Texts in Math., vol. 261 Springer Verlag, 2011.

[6] M. Ledoux: *The Concentration of Measure Phenomenon. AMS Surveys and Monographs. 89, 2001.*

[7] M. Ledoux: *Isoperimetry and Gaussian analysis.* In: Bernard P. (eds) Lectures on Probability Theory and Statistics. Lecture Notes in Mathematics, vol 1648. Springer, Berlin, Heidelberg, 1996.

[8] L. Lorenzo, M. Bertoldi: *Analytical Methods for Markov Semigroups*, Chapman & Hall, 2007.

[9] C. B. Morrey Jr: *Multiple Integrals in the Calculus of Variations*, Springer Verlag, 1966.

[10] L.I. Nicolaescu: *Introduction to Real Analysis*, World Scientific, 2019.

[11] A. Barvinok: *Math 710 lecture notes.*

[12] T. Tao: *Compressed sensing and single pixel cameras* Blog post, April 2007.

[13] A. Pazy: *Semigroups of Linear Operators and Applications to Partial Differential Equations* Springer-Verlag New-York Inc., 1983.

[14] J Neeman: *Notes on Markov Diffusions* Lecture notes, November 2016.

[15] D. Bakry, F. Barthe, P. Cattiaux, A. Guillin: *A Simple Proof of the Poincaré Inequality for a Large Class of Probability Measures Including the Log-Concave Case.* Electronic Communications in Probability. 13. 2008.