Notes on Markov diffusions

Joe Neeman

November 3, 2016

Contents

Preface

These notes are a draft, and almost certainly full of mistakes. Please do not distribute them.

These are lecture notes for a course entitled *Geometry of Markov Diffusions*, given in the summer of 2016 at the University of Bonn. Most of the material contained in these notes comes from the book by Bakry, Gentil, and Ledoux [BGL14]. Compared to that book, these notes aim to be

- less general, less comprehensive, and therefore possibly easier for the uninitiated to digest; and
- read sequentially.

Prerequisites

These notes assume a basic knowledge of probability and functional analysis. For example, we will require martingales, the Hahn-Banach theorem, and a working knowledge of various types of convergence. These notes explicitly do *not* require any background in stochastic processes or stochastic calculus. Although there is a strong connection between diffusion semigroups and diffusion processes, we will not deal with it in these notes.

Outline

The main object of study in these notes is called a *compact Markov diffusion triple*, which roughly consists of a space, a probability measure on it, and some notion of the norm of the gradient of a function. Its construction, which requires some preliminaries, is the main task of Chapter 1. From these compact Markov diffusion triples we will construct *Markov semigroups*; if one were so inclined, one could also go on to construct *Markov processes*. The main point of these notes is that the analysis of these Markov semigroups can reveal non-trivial properties of the underlying probability measure.

In Chapters 2, 3, and 4, we introduce three families of functional inequalities: Poincaré inequalities, logarithmic Sobolev inequalities, and isoperimetric inequalities. We will show how these inequalities are related, how they follow from properties of our Markov triples, and some of their geometric consequences.

The final two chapters will expand our repertoire of Markov triples, by showing how to canonically put Markov triples on weighted Riemannian manifolds and infinite-dimensional Gaussian measures.

Acknowledgements

I would like to thank the following people (in alphabetical order) for pointing out errors and suggesting improvements: Claudio Bellani, Mathias Braun, and Mike Parucha.

Chapter 1

Markov diffusions

1.1 Some measure-theoretic preliminaries

A *kernel* on a measurable space $(E, \mathcal{F})$ is a function $p : E \times \mathcal{F} \to [0, 1]$ such that

- for every $x \in E$, $p(x, \cdot)$ is a probability measure on $(E, \mathcal{F})$; and
- for every $A \in \mathcal{F}$, $p(\cdot, A)$ is a measurable function $E \to [0, 1]$.

We say that a measurable space $(E, \mathcal{F})$ has the *measure decomposition property* if every measure μ on $(E \times E, \mathcal{F} \otimes \mathcal{F})$ can be decomposed as $\mu(dx, dy) = \mu_1(dx)p(x, dy)$ for some measure μ_1 on E and some kernel p. It is well-known that if E is a Polish space and $\mathcal{F}$ is its Borel σ-algebra then $(E, \mathcal{F})$ satisfies the measure decomposition property.

We say that a measurable space $(E, \mathcal{F})$ is *good* if it satisfies the measure decomposition property and $\mathcal{F}$ is countably generated. The measure space $(E, \mathcal{F}, m)$ is *good* if it satisfies the measure decomposition property and $\mathcal{F}$ is countably generated up to m-null sets.

We will use the bi-measure theorem [MRN53] to show that Markov operators have kernel representations:

Theorem 1.1.1. *Let $(E, \mathcal{F})$ be a good measurable space, and let $\nu : \mathcal{F} \times \mathcal{F} \to [0, \infty)$ be a map such that $\nu(\cdot, A)$ and $\nu(A, \cdot)$ are finite measures for any $A \in \mathcal{F}$. Then there exists a finite measure $\tilde{\nu}$ on $(E \times E, \mathcal{F} \otimes \mathcal{F})$ such that $\nu(A, B) = \tilde{\nu}(A \times B)$ for all $A, B \in \mathcal{F}$.*

From now on, all measurable spaces will be assumed to be good and all measures will be assumed to be σ-finite.

1.2 Markov operators and semigroups

Definition 1.2.1. *A* Markov operator *P on $(E, \mathcal{F})$ is a linear operator that sends bounded measurable functions to bounded measurable functions, and also satisfies*

(a) $P\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the constant function with value 1; and

(b) $Pf \geq 0$ whenever $f \geq 0$.

If instead of (a) we have $P\mathbb{1} \leq \mathbb{1}$ pointwise, then P is called a *sub-Markov* operator. Note that if P is sub-Markov and f takes values in $[0, 1]$ then $Pf \geq 0$ and $P(\mathbb{1} - f) \geq 0$, which implies that

$$0 \leq Pf \leq P\mathbb{1} \leq \mathbb{1},$$

and so $\|Pf\|_\infty \leq 1$. By linearity, $\|Pf\|_\infty \leq \|f\|_\infty$ for all bounded, measurable f.

Next, we observe that Markov operators satisfy a Jensen-type inequality:

Proposition 1.2.2. *If P is a Markov operator then for every convex function ϕ and every bounded, measurable f, $P\phi(f) \geq \phi(Pf)$.*

Proof. First, consider simple functions f. That is, let $A_1, \dots, A_k$ be a measurable partition of E and take $f = \sum_{i=1}^k a_i 1_{A_i}$ for some $a_i \in \mathbb{R}$. Then $P1_{A_1}, \dots, P1_{A_k}$ are non-negative and sum to $\mathbb{1}$. Applying Jensen's inequality pointwise,

$$P\phi(f) = \sum_{i=1}^k \phi(a_i) P1_{A_i} \geq \phi\left(\sum_{i=1}^k a_i P1_{A_i}\right) = \phi(Pf).$$

We may approximate any bounded, measurable f uniformly by simple functions. Since P is continuous in $L^\infty(\mu)$, a standard approximation argument implies that the claimed inequality holds for all bounded, measurable f. □

Definition 1.2.3. *The measure μ is* invariant *for P if*

$$\int Pf \, d\mu = \int f \, d\mu$$

for every $f \in L^1(\mu) \cap L^\infty$.

As a consequence of Proposition 1.2.2, whenever μ is invariant for P then P may be extended to a contraction on $L^p(\mu)$ for any $1 \leq p < \infty$. Indeed, applying Proposition 1.2.2 to the function $\phi(x) = |x|^p$ yields

$$\int |Pf|^p \, d\mu \leq \int P(|f|^p) \, d\mu = \int |f|^p \, d\mu$$

for every $f \in L^1(\mu) \cap L^\infty$. Hence $\|Pf\|_p \leq \|f\|_p$ for all $f \in L^\infty(\mu) \cap L^1(\mu)$. Since this set is dense in $L^p(\mu)$, we may extend P to a contraction on every $L^p(\mu)$. As a consequence, the requirement that $f \in L^\infty$ may be removed from Definition 1.2.3.

Definition 1.2.4. *The measure μ is* reversible *for P if*

$$\int f P g \, d\mu = \int g P f \, d\mu$$

for all $f, g \in L^2(\mu)$.

It is easy to see that reversibility implies invariance: if μ is a finite measure then it is enough to just plug in $g = \mathbb{1}$. In the σ-finite case, one can take a sequence g_n of non-negative functions in $L^2(\mu)$ increasing to $\mathbb{1}$.

Definition 1.2.5. *A* symmetric Markov semigroup *on $(E, \mathcal{F}, \mu)$ is a family $\{P_t : t \geq 0\}$ of operators satisfying*

(a) for every $t \geq 0$, P_t is a Markov operator;

(b) $P_0 f = f$ for all bounded measurable f;

(c) $P_{t+s} = P_t P_s$ for all $s, t \geq 0$;

(d) for every $t \geq 0$, μ is reversible for P_t; and

(e) for all $f \in L^2(\mu)$, $P_t f \to f$ in $L^2(\mu)$ as $t \to 0$.

Property (c) of Definition 1.2.5 is the reason for calling $\{P_t\}$ a semigroup. Property (e) is sometimes called a strong continuity property, because it is equivalent to requiring that $t \mapsto P_t$ is continuous in the strong operator topology for operators on $L^2(\mu)$. The choice of $L^2(\mu)$ as the space underlying this convergence will be convenient for us, but other choices are also possible.

1.3 The kernel representation

In the setting of good measurable spaces, every Markov operator is actually an integral operator with respect to a probability kernel.

Proposition 1.3.1. *Suppose P is a Markov operator that is continuous on $L^1(\mu)$. Then there exists a probability kernel p on $(E, \mathcal{F})$ such that for every $f \in L^\infty(\mu)$ and μ-almost every $x \in E$,*

$$(Pf)(x) = \int f(y)p(x, dy). \tag{1.1}$$

In particular, a Markov semigroup $\{P_t\}$ may be represented by a family of probability kernels $p_t : E \times \mathcal{F} \to [0, 1]$.

Proof. First, assume that μ is a probability measure and that P is a sub-Markov operator. Consider the map $\nu : \mathcal{F} \times \mathcal{F}$ defined by

$$\nu(A, B) = \int 1_A P 1_B \, d\mu.$$

We claim that $\nu(A, \cdot)$ and $\nu(\cdot, A)$ are both finite measures. Indeed, only the countable additivity property is non-trivial; to check this property, it suffices to show that if B_k is a decreasing sequence of sets with $\bigcap B_k = \emptyset$ then $\nu(B_k, A)$ and $\nu(A, B_k)$ both converge to zero. The first of these follows from the monotone convergence theorem:

$$\nu(B_k, A) = \int 1_{B_k} P 1_A \, d\mu,$$

where the integrand converges pointwise downwards to zero because $1_{B_k} \downarrow 0$ and $P1_A \ge 0$. On the other hand,

$$0 \le \nu(A, B_k) = \int 1_A P 1_{B_k} \, d\mu \le \int P 1_{B_k} \, d\mu.$$

The right-hand side converges to zero because $1_{B_k} \to 0$ in $L^1(\mu)$ and P is continuous on $L^1(\mu)$.

Now we may apply Theorem 1.1.1: there exists a measure $\tilde{\nu}$ on $E \times E$ such that $\tilde{\nu}(A \times B) = \int 1_A P 1_B \, d\mu$. By the measure decomposition property, we may decompose $\tilde{\nu}$ as $\tilde{\nu}(dx, dy) = \nu_1(dx)p(x, dy)$. That is,

$$\int_E 1_A P 1_B \, d\mu = \int_E 1_A(x) \int_E 1_B(y) p(x, dy) \, \nu_1(dx). \tag{1.2}$$

With $B = E$, we obtain

$$\nu_1(A) = \int_E 1_A P\mathbb{1}\, d\mu \leq \mu(A).$$

It follows then that ν_1 is absolutely continuous with respect to μ, with density $\frac{d\nu_1}{d\mu} = P\mathbb{1}$.

Since every bounded, measurable function may be approximated by a sum of indicator functions, (1.2) implies that for every bounded, measurable f and every $A \in \mathcal{F}$,

$$\begin{aligned}\int_E 1_A(x) Pf(x)\, \mu(dx) &= \int_E 1_A(x) \int_E f(y)p(x,dy)\, \nu_1(dx) \\ &= \int_E 1_A(x) \int_E f(y)p(x,dy) P\mathbb{1}(x)\, \mu(dx).\end{aligned}$$

Since A is arbitrary, it follows that

$$(Pf)(x) = P\mathbb{1}(x) \int_E f(y)p(x,dy)$$

for μ-almost every x.

Next, we consider the case that μ is σ-finite. Let A_n be a growing sequence of measurable sets of finite measure exhausting μ. Let μ_n be μ restricted to A_n, so that μ_n is a finite measure. If we define P_n by $P_n f = P(f 1_{A_n})$ then P_n is a sub-Markov operator that acts continuously on $L^1(\mu_n)$. Since μ_n is finite, we may apply the first part of the proof to find a kernel p_n satisfying

$$(P_n f)(x) = (P_n \mathbb{1})(x) \int f(y)p_n(x,dy)$$

for μ_n-almost every x. Finally, we may define the kernel p by setting $p(x,B) = \lim_{n\to\infty} p_n(x,B)$, where the limit exists because the definition of P_n ensures that $p_n(x,B)$ is non-decreasing in n. By the monotone convergence theorem, for every bounded, non-negative, measurable function f,

$$\lim_{n\to\infty} (P_n f)(x) = \lim_{n\to\infty} (P_n \mathbb{1})(x) \int f(y)p(x,dy).$$

Now, $P_n f = P(1_{A_n} f)$, and since $1_{A_n} f$ converges to f in $L^1(\mu)$ it follows that $P_n f \to Pf$ in $L^1(\mu)$; since the convergence is monotone, it must also hold for μ-almost every point $x \in E$. Applying the same reasoning to $P_n\mathbb{1}$, we conclude that

$$(Pf)(x) = (P\mathbb{1})(x) \int f(y)p(x,dy)$$

for μ-almost every x. In the case that P is a Markov operator, this is exactly (1.1). □

1.3.1 Density kernels

Sufficiently nice Markov operators (i.e., all the ones that we will be interested in) can even be represented by kernels that have densities with respect to some reference measure. From now on, we introduce the notation $\|P\|_{p,q}$ for the $L^p(\mu) \to L^q(\mu)$ norm of the operator P. (The measure μ is absent from the notation, but it should always be clear from the context.)

Proposition 1.3.2. *Suppose P is a Markov operator on $(E, \mathcal{F}, \mu)$ satisfying $\|P\|_{1,\infty} = M < \infty$. Then there exists a measurable $p : E \times E \to [0, \infty)$ such that $\|p\|_\infty = \|P\|_{1,\infty}$ and*

$$(Pf)(x) = \int f(y) p(x, y) \, \mu(dy)$$

for all bounded, measurable f and μ-almost every x.

Proof. We consider only the case that μ is a probability measure; as before, the σ-finite case is done by taking limits. The basic idea is to fix an x and to consider the linear functional $f \mapsto (Pf)(x)$, which is bounded on $L^1(\mu)$ by our assumption on P. Then it can be represented by an element of $L^\infty(\mu)$: there is some $p_x \in L^\infty(\mu)$ such that $\|p_x\|_\infty$ is bounded by the norm of our functional (which is in turn bounded by M) and such that

$$(Pf)(x) = \int f(y) p_x(y) \, \mu(dy)$$

for every f. There are at least two problems with this basic idea: the first is that Pf is only defined up to sets of measure zero, and so our linear functional does not make sense. The second is that we would need to somehow combine the $p_x(y)$ functions in such a way that they are measurable in x.

Both of the problems in the previous paragraph can be circumvented by dropping to a finite space: let $\mathcal{F}_n$ be a sequence of finite σ-algebras exhausting $\mathcal{F}$, satisfying $\mu(A) > 0$ for all n and all $A \in \mathcal{F}_n$ (such a sequence exists because of our assumption that $(E, \mathcal{F}, \mu)$ is a nice measurable space). Consider the Markov operator P_n on $(E, \mathcal{F}_n, \mu)$ defined by $P_n f = \mathbb{E}(Pf \mid \mathcal{F}_n)$. Then $\|P_n\|_{1,\infty} \le M$ and we may carry out the program of the previous paragraph to find a $(\mathcal{F}_n \otimes \mathcal{F}_n)$-measurable $p_n(x, y)$ satisfying $\|p_n\|_\infty \le M$ and

$$(P_n f)(x) = \int f(y) p_n(x, y) \, \mu(dy)$$

for every $f \in \mathcal{F}_n$ and every x. This time, there is no problem making sense of the linear functional $x \mapsto (Pf)(x)$, since a measurable function on a σ-algebra containing only sets of positive measure is defined pointwise.

Now we send $n \to \infty$. Since $P_n f = \mathbb{E}(Pf \mid \mathcal{F}_n)$ is a bounded martingale, it converges μ-almost surely to Pf. For any fixed x, $p_n(x, \cdot)$ is also a bounded martingale, because for any $A, B \in \mathcal{F}_n$

$$\int_{A\times B} p_{n+1}\, d\mu \otimes \mu = \int 1_A (P_{n+1} 1_B)\, d\mu = \int 1_A (P_n 1_B)\, d\mu = \int_{A\times B} p_n d\mu \otimes \mu$$

(where the middle equality follows from the definition of P_n). Hence p_n converges $(\mu \otimes \mu)$-almost everywhere to a limit p. Moreover, $\|p\|_\infty \le M$ and (by the dominated convergence theorem)

$$\int f(y) p_n(x,y)\, \mu(dy) \to \int f(y) p(x,y)\, \mu(dy)$$

for every $f \in L^\infty(\mu)$. Finally, the right hand side above must be equal to $Pf = \lim_{n\to\infty} P_n f$ μ-almost everywhere. □

1.4 Some examples

We give some examples of symmetric Markov semigroups. Note that in none of the examples below do we actually check the properties of Definition 1.2.5 – this may be taken as an exercise.

Example 1.4.1 (The heat semigroup on $\mathbb{R}^n$)**.** *Let μ be the Lebesgue measure on $\mathbb{R}^n$, and define*

$$p_t(x,y) = \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/4t}.$$

That is, for fixed x, $p_t(x, \cdot)$ is the density of a Gaussian random variable with mean x and covariance $2tI_n$ (where I_n denotes the $n \times n$ identity matrix). One can easily check that

$$(P_t f)(x) = \int_{\mathbb{R}^n} f(y) p_t(x,y)\, dy$$

defines a symmetric Markov semigroup on $(\mathbb{R}^n, \mathcal{B}, \mathrm{Leb})$*, where $\mathcal{B}$ is the Borel σ-algebra and* Leb *is the Lebesgue measure. This semigroup may also be described in terms of Brownian motion: if B_t is a standard Brownian motion on $\mathbb{R}^n$ then $P_t f(x) = \mathbb{E} f(x + \sqrt{2} B_t)$.*

Example 1.4.2 (The wrapped heat semigroup)**.** *Let p_t be the heat kernel on $\mathbb{R}$ from the previous example, and define the semigroup P_t^W on $([0,1], \mathcal{B})$ by*

$$(P_t^W f)(x) = \int_0^1 f(y) \sum_{k \in \mathbb{Z}} p_t(x, k+y)\, dy.$$

Then this defines a symmetric Markov semigroup with respect to the Lebesgue measure on $[0,1]$. In the same way that the heat semigroup on $\mathbb{R}^n$ was associated with a Brownian motion, this semigroup is associated with a "wrapped" Brownian motion: let $B_t^W = B_t - \lfloor B_t \rfloor$, where B_t is a Brownian motion on $\mathbb{R}$. Then $P_t^W f(x) = \mathbb{E} f(x + \sqrt{2} B_t^W)$. This is called a wrapped Brownian motion because it runs like a Brownian motion on $[0,1]$ until it hits one of the endpoints, at which point it wraps around to the other endpoint.

Example 1.4.3 (The reflected heat semigroup)**.** *With p_t again the heat kernel on $\mathbb{R}$, define the semigroup P_t^R on $([0,1], \mathcal{B})$ by*

$$(P_t^R f)(x) = \int_0^1 f(y) \sum_{k \in \mathbb{Z}} (p_t(x, 2k+y) + p_t(x, 2k-y))\, dy.$$

This also defines a symmetric Markov semigroup with respect to the Lebesgue measure on $[0,1]$. This one is associated, however, with a "reflected" Brownian motion

$$B_t^R = \begin{cases} B_t - \lfloor B_t \rfloor & \text{if } \lfloor B_t \rfloor \text{ is even} \\ 1 - B_t + \lfloor B_t \rfloor & \text{if } \lfloor B_t \rfloor \text{ is odd,} \end{cases}$$

which runs like a Brownian motion on $[0,1]$ until it hits the boundary, at which point it bounces off.

Example 1.4.4 (The Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$)**.** *Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$ (i.e., the measure whose density is $(2\pi)^{-n/2} e^{-|x|^2/2}$ with respect to the Lebesgue measure). Then*

$$(P_t f)(x) = \int_{\mathbb{R}^n} f(e^{-t} x + \sqrt{1 - e^{-2t}} y)\, d\gamma_n(y)$$

defines a symmetric Markov semigroup on $(\mathbb{R}^n, \mathcal{B}, \gamma_n)$. Note that representation above (unlike our previous examples) is not written in terms of a density kernel, but by a change of variables it can be put in that form.

1.5 Generators

The generator of a Markov semigroup is a (usually unbounded) operator that describes the semigroup's behavior for infinitesmally small t. (essentially, it is the derivative of the semigroup). As we will show later, the generator determines the semigroup completely.

Definition 1.5.1. *Let $\{P_t : t \geq 0\}$ be a symmetric Markov semigroup on $(E, \mathcal{F}, \mu)$. We define its* generator *L by*

$$Lf = \lim_{t \to 0} \frac{P_t f - f}{t},$$

with the domain $\mathcal{D}(L)$ being the set of $f \in L^2(\mu)$ such that the limit exists in $L^2(\mu)$. We may also define L on other domains: let $\mathcal{D}_p(L)$ be the set of $f \in L^p(\mu)$ such that the limit above exists in $L^p(\mu)$; then L is also defined as an operator from $\mathcal{D}_p(L)$ to $L^p(\mu)$. Note that there is no ambiguity in using the letter L for all of these operators, since if a sequence converges in both $L^p(\mu)$ and $L^q(\mu)$ then the limits coincide.

The semigroup property (Definition 1.2.5, part (c)) may be used to show that for any $f \in \mathcal{D}(L)$ and $t \geq 0$, $P_t f \in \mathcal{D}(L)$ and

$$LP_t f = P_t L f = \lim_{s \to 0} \frac{P_{t+s} f - P_t f}{s},$$

where the limit exists in $L^2(\mu)$; we leave this as an exercise.

The symmetry (Definition 1.2.5, part (d)) of $\{P_t : t \geq 0\}$ may also be seen in the generator:

Proposition 1.5.2. *For every $f, g \in \mathcal{D}(L)$,*

$$\int f L g \, d\mu = \int g L f \, d\mu.$$

For every $f \in \mathcal{D}_1(L)$,

$$\int L f \, d\mu = 0.$$

Proof. By the symmetry of P_t,

$$\frac{1}{t} \int f(P_t g - g) \, d\mu = \frac{1}{t} \int g(P_t f - f) \, d\mu.$$

The definition of L ensures that the left hand side converges to $\int fLg\,d\mu$ as $t \to 0$, while the right hand side converges to $\int gLf\,d\mu$. This proves the first claim.

For the second claim, recall that μ is invariant for P_t, and so $\int P_t f\,d\mu = \int f\,d\mu$ for all $f \in L^1(\mu)$. Hence,

$$\int \frac{P_t f - f}{t}\,d\mu = 0$$

for all t. Recall that $f \in \mathcal{D}_1(L)$ means that the integrand above converges to Lf in $L^1(\mu)$ as $t \to 0$. The second claim follows. $\square$

1.6 The carré du champ

There is another useful way, known as the *carré du champ* of characterizing the infinitesmal behavior of a Markov semigroup. Suppose that we are able to find an algebra $\mathcal{A}$ contained in $\mathcal{D}(L)$ (that is, $\mathcal{A}$ is a vector space of functions that is closed under multiplication). For example, if E is a subset of $\mathbb{R}^n$ then $\mathcal{A}$ might consist of smooth functions compactly supported in E.

Definition 1.6.1. *The* carré du champ *associated to L is the bilinear form Γ on $\mathcal{A} \times \mathcal{A}$ defined by*

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

We will abbreviate $\Gamma(f,f)$ by $\Gamma(f)$.

Note that we have defined Γ in terms of L, but one can also go (at least partly) the other way: if L is any operator satisfying the symmetry and invariance conditions in Proposition 1.5.2 and the formula $2\Gamma(f,g) = L(fg) - fLg - gLf$ then

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

for every $f, g \in \mathcal{A}$. If $\mathcal{A}$ is dense in $L^2(\mu)$ then this determines Lg for every $g \in \mathcal{A}$. Whether this determines L completely is a slightly more delicate point, which we will investigate later.

Proposition 1.6.2. *Suppose that Γ is the carré du champ associated to L, which is itself the generator of a symmetric Markov semigroup. Then $\Gamma(f) \geq 0$ pointwise for every $f \in \mathcal{A}$.*

Proof. By the definition of Γ, it suffices to show that if $f \in \mathcal{A}$ then $L(f^2) \geq 2fLf$ pointwise. Now, Proposition 1.2.2 implies that $P_t(f^2) \geq (P_t f)^2$ pointwise, and hence

$$L(f^2) = \lim_{t \to 0} \frac{P_t(f^2) - f^2}{t} \geq \lim_{t \to 0} \frac{(P_t f)^2 - f^2}{t} = \left. \frac{d(P_t f)^2}{dt} \right|_{t=0}.$$

On the other hand, the product rule yields

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

Since Γ is bilinear, the statement $\Gamma(f+g) \geq 0$ may be rearranged into a Cauchy-Schwarz inequality for Γ: for all $f, g \in \mathcal{A}$,

$$\Gamma(f, g) \leq \sqrt{\Gamma(f)\Gamma(g)}.$$

1.6.1 The examples again

We revisit our examples from Section 1.4 in order to compute their generators and carré du champs. In general, it can be tricky to figure out exactly what the domain of the generator is, but it usually suffices to specify the generator on a sufficiently large subset of its domain. We will say more about this later.

Example 1.6.3 (The heat semigroup on $\mathbb{R}^n$)**.** *If $f : \mathbb{R}^n \to \mathbb{R}$ is bounded and has two bounded, uniformly continuous derivatives then Taylor's theorem may be used to check that*

$$\frac{P_t f - f}{t} \stackrel{L^\infty}{\to} \Delta f,$$

where P_t is the heat semigroup of Example 1.4.1 and $\Delta f = \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i^2}$. Thus, $\mathcal{D}_\infty(L)$ contains all smooth, bounded functions, and $Lf = \Delta f$ for such functions. Since bounded functions of compact support belong to $L^2(\mathrm{Leb})$, one can easily check that $\mathcal{D}(L)$ contains all twice-differentiable functions of compact support (which we denote by $C_c^2(\mathbb{R}^n)$). Since $C_c^2(\mathbb{R}^n)$ is an algebra, we may define Γ on it. Some algebra reveals that

$$\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$$

for $f, g \in C_c^2(\mathbb{R}^n)$.

Example 1.6.4 (The wrapped heat semigroup)**.** *To find the generator of the wrapped heat semigroup, we first define the periodic extension $f^W : \mathbb{R} \to \mathbb{R}$ of a function $f : [0,1] \to \mathbb{R}$ by $f^W(x) = f(x - \lfloor x \rfloor)$. If f has continous second derivatives and satisfies $f^{(k)}(0) = f^{(k)}(1)$ for $k = 0, 1, 2$ (where $f^{(k)}$ means the kth derivative of f, and $f^{(0)} = f$) then f^W is C^2 and bounded on $\mathbb{R}$. Moreover, one can easily check by rearranging the definition of P_t^W that $P_t^W f(x) = P_t f^W(x)$ for all $x \in [0,1]$. Then we know from the previous example that*

$$\frac{P_t^W f - f}{t} = \frac{P_t f^W - f^W}{t} \overset{L^\infty}{\to} (f^W)'' = f''$$

on $[0,1]$ for all functions f as above. Since convergence in L^∞ implies convergence in $L^2([0,1])$, it follows that $\mathcal{D}(L)$ contains all smooth functions satisfying $f^{(k)}(0) = f^{(k)}(1)$ for all $k = 0, 1, 2$. This set of functions is an algebra, and on it we have $\Gamma(f, g) = f'g'$.

Example 1.6.5 (The reflected heat semigroup)**.** *The reflected heat semigroup can be analyzed in a similar way to the wrapped heat semigroup, but by taking a different extension: for $f : [0,1] \to \mathbb{R}$, define $f^R : \mathbb{R} \to \mathbb{R}$ by*

$$f^R(x) = \begin{cases} f(x - \lfloor x \rfloor) & \text{if } \lfloor x \rfloor \text{ is even} \\ f(1 - x + \lfloor x \rfloor) & \text{if } \lfloor x \rfloor \text{ is odd.} \end{cases}$$

In order to make f^R belong to $C^2(\mathbb{R})$, we need f to be in $C^2([0,1])$ and to satisfy $f'(0) = f'(1) = 0$. For such functions, we argue as in the previous example that $Lf = f''$ and $\Gamma(f, g) = f'g'$. Note that L and Γ are given by the same formulas in both the wrapped and reflected heat semigroups; the only difference is that their domains might be different (although we haven't shown this yet, since we haven't tried to actually characterize the domains, but only to find subsets of them).

Example 1.6.6 (The Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$)**.** *The generator may be computed in a similar way to the generator of the heat semigroup. Since we left out that computation we will include this one, but only in the case $n = 1$, and with the simplifying assumption that f has three bounded derivatives (instead of two uniformly continuous derivatives). By Taylor's* *theorem, if* $z = (e^{-t} - 1)x + \sqrt{1 - e^{-2t}}y$ *then*

$$\begin{aligned}(P_t f)(x) - f(x) &= \int_{\mathbb{R}} f(e^{-t}x - \sqrt{1 - e^{-2t}}y) - f(x)\, d\gamma_1(y) \\ &\le \int_{\mathbb{R}} f'(x) z \, d\gamma_1(y) + \frac{1}{2} \int_{\mathbb{R}} f''(x) z^2 \, d\gamma_1(y) \\ &\quad + \frac{\|f'''\|_\infty}{6} \int_{\mathbb{R}} z^3 \, d\gamma_1(y).\end{aligned}$$

We can compute

$$\begin{aligned}\int z \, d\gamma_1(y) &= (e^{-t} - 1)x \\ \int z^2 \, d\gamma_1(y) &= (e^{-t} - 1)^2 x^2 + 1 - e^{-2t} \\ \int z^3 \, d\gamma_1(y) &= O(t^{3/2}) \text{ as } t \to 0.\end{aligned}$$

Hence,

$$\frac{(P_t f)(x) - f(x)}{t} = f''(x) - x f'(x) + O(\sqrt{t})$$

as $t \to 0$*, where the* $O(\sqrt{t})$ *term is uniform in* x*. It follows that the left hand side converges in* L^∞ *(and therefore also in* $L^2(\gamma_1)$*, since* γ_1 *is a probability measure) to* $f''(x) - x f'(x)$*. That is, the generator of the Ornstein-Uhlenbeck semigroup on* $\mathbb{R}$ *acts as* $Lf(x) = f''(x) - x f'(x)$ *for all bounded* f *with three bounded derivatives. An analogous computation gives*

$$Lf(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle$$

in higher dimensions. However, the carré du champ operator remains as it did in the case of the heat semigroup: $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$*.*

1.7 Self-adjoint operators

Let $\mathcal{H}$ be a (separable) Hilbert space. If A and B are two operators defined on subsets of $\mathcal{H}$, then we write $A \subset B$ (and say that "B is an extension of A") if $\mathcal{D}(A) \subset \mathcal{D}(B)$ and $Af = Bf$ for all $f \in \mathcal{D}(A)$.

Definition 1.7.1. *An operator* $A : \mathcal{D}(A) \to \mathcal{H}$ *is* symmetric *if* $\langle Af, g \rangle = \langle f, Ag \rangle$ *for all* $f, g \in \mathcal{D}(A)$*.*

Definition 1.7.2. *If $\mathcal{D}(A)$ is dense in $\mathcal{H}$ then the* adjoint *of A is the operator A^* defined by*

$$\langle A^* f, g\rangle = \langle f, Ag\rangle$$

on $\mathcal{D}(A^*) = \{f \in \mathcal{H} : \text{the linear functional } g \mapsto \langle f, Ag\rangle \text{ is bounded}\}$. An operator A is self-adjoint *if $A^* = A$.*

In order to make sense of the definition of an adjoint, note that if $\mathcal{D}(A)$ is dense in $\mathcal{H}$ then for every bounded linear functional $\ell : \mathcal{D}(A) \to \mathbb{R}$ there is a unique $h \in \mathcal{H}$ such that $\ell(g) = \langle h, g\rangle$ for all $g \in \mathcal{D}(A)$. Hence, $\mathcal{D}(A^*)$ is defined to be exactly the set of f such that the equation

$$\langle h, g\rangle = \langle f, Ag\rangle \text{ for all } g \in \mathcal{D}(A)$$

uniquely defines h.

Note that a self-adjoint operator is always symmetric, but that a symmetric operator is not necessarily self-adjoint – the issue is that for a symmetric operator A, $\mathcal{D}(A)$ and $\mathcal{D}(A^*)$ might not coincide.

Exercise 1.7.1. *Show that if A is symmetric and densely defined then $A \subset A^*$ and A^* is closed. (In particular, every self-adjoint operator is closed.)*

Example 1.7.3. *Take $\mathcal{H} = L^2([0,1], \mathrm{Leb})$. Consider the operator A defined by $Af = f''$ on $\mathcal{D}(A) = \mathcal{C}_c^2([0,1])$ and the operator B defined by $Bf = f''$ on*

$$\mathcal{D}(B) = \{f \in \mathcal{C}^2([0,1]) : f(0) = f(1) = 0\}.$$

Note that $A \subset B$, that both are densely defined, and that both are symmetric operators because

$$\int_0^1 f(x)g''(x)\,dx = [fg']_0^1 - [f'g]_0^1 + \int_0^1 f''(x)g(x)\,dx$$

for all $f, g \in \mathcal{C}^2([0,1])$. Whenever both f and g belong to either $\mathcal{D}(A)$ or $\mathcal{D}(B)$, the right hand side reduces to $\int_0^1 f''g\,dx$.

Next, consider the adjoints of A and B: for $f \in \mathcal{C}^2$, consider the linear functional ℓ_f defined by

$$\ell_f(g) = \int_0^1 fg''\,dx.$$

Now, $f \in \mathcal{D}(A^*)$ if and only if ℓ_f is bounded on $\mathcal{D}(A)$. But for $g \in \mathcal{D}(A)$, we have

$$\ell_f(g) = \int_0^1 f''g\,dx \le \|f''\|_2 \|g\|_2,$$

and so we see that $\mathcal{D}(A^*)$ contains all of $\mathcal{C}^2([0,1])$. In particular, $\mathcal{D}(A^*)$ is substantially larger than $\mathcal{D}(A)$ and so A is certainly not self-adjoint.

On the other hand, $f \in \mathcal{D}(B^*)$ if and only if ℓ_f is bounded on $\mathcal{D}(B)$. For $g \in \mathcal{D}(B)$,

$$\ell_f(g) = [fg']_0^1 + \int_0^1 f'' g \, dx.$$

Now, if $f \in \mathcal{D}(B)$ then the first term on the right hand side disappears and so ℓ_f is bounded on $\mathcal{D}(B)$. On the other hand, if $f \notin \mathcal{D}(B)$ then we may find a sequence $g_n \in \mathcal{D}(B)$ such that $g_n \to 0$ but $g_n'(0) = g_n'(1) = 1$ for every n. For this sequence, $\ell_f(g_n) \to f(1) - f(0) \neq 0$ and so ℓ_f is not bounded on $\mathcal{D}(B)$. To summarize, $f \in \mathcal{C}^2([0,1])$ belongs to $\mathcal{D}(B^*)$ if and only if $f \in \mathcal{D}(B)$. In some sense, B is much closer to being self-adjoint than A is; the next exercise makes this more precise.

Exercise 1.7.2. *In the example above, show that $\bar{B} = B^*$ and that $\bar{B}$ is self-adjoint. (Hint: use Fourier series to characterize their domains.)*

1.7.1 Generators are self-adjoint

The entire reason for introducing self-adjoint operators in these notes is to apply their theory to generators of Markov semigroups. For the rest of this section, let $\{P_t : t \geq 0\}$ be a symmetric Markov semigroup on $(E, \mathcal{F}, \mu)$; let L be its generator.

Proposition 1.7.4.

(a) $\mathcal{D}(L)$ is dense in $L^2(\mu)$.

(b) L is closed.

(c) L is self-adjoint.

Proof. For $\epsilon > 0$, define the operator A_ϵ by $A_\epsilon f = \frac{1}{\epsilon} \int_0^\epsilon P_s f \, ds$. Note that A_ϵ computes with P_t, and that $A_\epsilon f \to f$ as $\epsilon \to 0$ for every $f \in L^2(\mu)$.

Define the operator L_t by $L_t f = \frac{1}{t}(P_t f - f)$; then $L_t f \to Lf$ as $t \to 0$ if and only if $f \in \mathcal{D}(L)$. Now, for any t and ϵ

$$\begin{aligned} L_t A_\epsilon f &= \frac{\int_t^{t+\epsilon} P_s f \, ds - \int_0^\epsilon P_s f \, ds}{\epsilon t} \\ &= \frac{\int_\epsilon^{t+\epsilon} P_s f \, ds - \int_0^t P_s f \, ds}{\epsilon t} \\ &= L_\epsilon A_t f. \end{aligned}$$

Taking $t \to 0$, $L_\epsilon A_t f \to L_\epsilon f$ (since L_ϵ is bounded for any fixed $\epsilon > 0$. We conclude that for any $\epsilon > 0$ and $f \in L^2(\mu)$ we have $A_\epsilon f \in \mathcal{D}(L)$ and $LA_\epsilon f = L_\epsilon f$. Since any $f \in L^2(\mu)$ may be approximated arbitrarily well by $A_\epsilon f \in \mathcal{D}(L)$, it follows that $\mathcal{D}(L)$ is dense in $L^2(\mu)$.

To check that L is closed, suppose that $f_n \in \mathcal{D}(L)$ with $f_n \to f$ and $Lf_n \to g$. Note that A_t commutes with L_t, and therefore also with L. Hence,

$$L_t f = \lim_{n\to\infty} L_t f_n = \lim_{n\to\infty} LA_t f_n = \lim_{n\to\infty} A_t L f_n = A_t g.$$

Taking $t \to 0$, it follows that $L_t f \to g$, which implies that $f \in \mathcal{D}(L)$ and $Lf = g$.

Finally, we check that L is self-adjoint: the only thing to check is that if $f \in \mathcal{D}(L^*)$ then $f \in \mathcal{D}(L)$. Suppose, then, that $f \in \mathcal{D}(L^*)$. Then there is a constant C such that for all $g \in \mathcal{D}(L)$,

$$C\|g\|_2 \geq \int_E f Lg\, d\mu = \lim_{t\to 0} \int_E f L_t g\, d\mu = \lim_{t\to 0} \int_E g L_t f\, d\mu.$$

In other words, there is a bounded linear functional ℓ on $\mathcal{D}(L)$ such that

$$\int_E g L_t f\, d\mu \to \ell(g).$$

Since $\mathcal{D}(L)$ is dense, it follows that there is some $h \in L^2(\mu)$ such that $\ell(g) = \int_E hg\, d\mu$ for all $g \in \mathcal{D}(L)$ and $L_t f \rightharpoonup h$. This is almost what we want, except that we need to upgrade the weak convergence to strong convergence.

Recalling that $LA_t = L_t f \rightharpoonup h$ and that $A_t f \to f$ as $t \to 0$, we see that the pair $(A_t f, LA_t f) \rightharpoonup (f, h)$ in $L^2(\mu) \times L^2(\mu)$. Since the graph of L is convex and closed, it is also weakly closed. Hence (f, h) belongs to the graph of L, and so $f \in \mathcal{D}(L)$. □

Definition 1.7.5. *A self-adjoint operator A on $\mathcal{H}$ is* non-negative *if $\langle Af, f\rangle \geq 0$ for all $f \in \mathcal{D}(A)$. A is* non-positive *if $-A$ is non-negative.*

Proposition 1.7.6. *If L is the generator of a symmetric Markov semigroup then L is non-positive.*

Proof. For $f \in \mathcal{D}(L)$,

$$\int fLf\, d\mu = \lim_{t\to 0} \frac{1}{t} \int f P_t f - f^2\, d\mu = \lim_{t\to 0} \frac{\|P_{t/2} f\|_2^2 - \|f\|_2^2}{t}.$$

For any $t > 0$, the numerator is non-positive because P_t is a contraction on $L^2(\mu)$. □

To summarize, every generator of a symmetric Markov semigroup is a non-positive self-adjoint operator. It turns out that the converse is partly true: every non-positive self-adjoint operator on $L^2(\mu)$ is the generator of some strongly continuous contraction semigroup on $L^2(\mu)$ (we will say what that means later).

1.7.2 The Friedrichs extension

The problem with self-adjoint operators is that it is sometimes difficult to actually specify them. Most of the time (as in Example 1.7.3), we want to define our operators only on a set of sufficiently smooth functions; this will never be a large enough domain for the operator to be self-adjoint. Fortunately, there is a general recipe for finding a self-adjoint extension of a densely-defined, non-negative symmetric operator. Such an extension is not always unique.

Suppose that A is a densely defined, non-negative symmetric operator on a Hilbert space $\mathcal{H}$. Now define the positive, symmetric bilinear form $\langle \cdot, \cdot \rangle_A$ on $\mathcal{D}(A)$ by

$$\langle f, g \rangle_A = \langle f, g \rangle + \langle Af, g \rangle = \langle f, g \rangle + \langle f, Ag \rangle.$$

Now let $\mathcal{H}_1 \subseteq \mathcal{H}$ be the elements of the form $\lim_{n \to \infty} f_n$, where the limit is with respect to $\|\cdot\|$ and $f_n \in \mathcal{D}(A)$ is a sequence that is Cauchy with respect to $\|\cdot\|_A$. Since A is non-negative, every sequence that is Cauchy with respect to $\|\cdot\|_A$ is also Cauchy with respect to $\|\cdot\|$, and so it always has a limit in $\mathcal{H}$. One can then extend $\|\cdot\|_A$ to $\mathcal{H}_1$, by defining $\|f\|_A = \lim_{n \to \infty} \|f_n\|_A$ whenever f_n is a sequence as above.

Exercise 1.7.3.

(a) *Check that the definition of $\|f\|_A$ for $f \in \mathcal{H}_1$ above does not depend on the choice of f_n.*

(b) *Show that $\mathcal{H}_1$ is complete with respect to $\|\cdot\|_A$.*

Now we define a new operator $\tilde{A}$ on $\mathcal{H}$ by $\langle \tilde{A} f, g \rangle = \langle f, g \rangle_A - \langle f, g \rangle$ on the domain

$$\mathcal{D}(\tilde{A}) = \{f \in \mathcal{H}_1 : \exists\, C(f) \text{ with } \langle f, g \rangle_A \le C(f) \|g\| \text{ for all } g \in \mathcal{H}_1\}.$$

Note that our definition of $\tilde{A}$ makes sense because for every $f \in \mathcal{D}(\tilde{A})$, the functional $\langle f, \cdot \rangle_A - \langle f, \cdot \rangle$ extends uniquely to a bounded linear functional on $\mathcal{H}$.

Exercise 1.7.4. *Show that $\tilde{A}$ is a self-adjoint extension of A. It is called the* Friedrichs extension.

1.8 Spectral theory of self-adjoint operators

Let $\mathcal{H}$ be a (separable) Hilbert space.

Definition 1.8.1. *A family $\{\mathcal{H}_\lambda : \lambda \geq 0\}$ of closed linear subspaces is a* spectral decomposition *of $\mathcal{H}$ if*

(a) $\mathcal{H}_\lambda \subseteq \mathcal{H}_{\lambda'}$ whenever $\lambda \leq \lambda'$;

(b) $\mathcal{H}_\lambda = \bigcup_{\lambda' > \lambda} \mathcal{H}_{\lambda'}$; and

(c) $\bigcup_{\lambda \geq 0} \mathcal{H}_\lambda = \mathcal{H}$.

Take a spectral decomposition $\{\mathcal{H}_\lambda\}$ and consider the orthogonal projection operators $E_\lambda : \mathcal{H} \to \mathcal{H}_\lambda$. For any $f \in \mathcal{H}$, the function $\lambda \mapsto \langle E_\lambda f, f\rangle = \|E_\lambda f\|^2$ is non-decreasing, right-continuous, and bounded by $\|f\|^2$. For any $f, g \in \mathcal{H}$,

$$\langle E_\lambda f, g\rangle = \frac{1}{2}\left(\langle E_\lambda (f+g), f+g\rangle - \langle E_\lambda f, f\rangle - \langle E_\lambda g, g\rangle\right),$$

and hence the function $\lambda \mapsto \langle E_\lambda f, g\rangle$ is right-continuous and has bounded variation. For a bounded, measurable function $\psi : \mathbb{R}_+ \to \mathbb{R}$, define the operator $T_\psi : \mathcal{H} \to \mathcal{H}$ by

$$\langle T_\psi f, g\rangle = \psi(0)\langle E_0 f, g\rangle + \int_0^\infty \psi(\lambda)\, d\langle E_\lambda f, g\rangle. \tag{1.3}$$

This defines a bounded operator, because if $\|f\| = \|g\| = 1$ then the total variation of the function $\lambda \mapsto \langle E_\lambda f, g\rangle$ is at most 3. If ψ is unbounded, we may define T_ψ on the domain

$$\mathcal{D}(T_\psi) = \left\{ f \in \mathcal{H} : \int_0^\infty \psi^2(\lambda) d\langle E_\lambda f, f\rangle < \infty \right\}$$

by the same formula (1.3).

Exercise 1.8.1. *Our definition of T_ψ in* (1.3) *satisfies the following properties:*

(a) If $\phi, \psi : \mathbb{R}_+ \to \mathbb{R}$ are two bounded, measurable functions then $T_{\phi\psi} = T_\phi T_\psi$.

(b) If $\phi : \mathbb{R}_+ \to \mathbb{R}$ is bounded and measurable then T_ϕ is symmetric and as norm at most $\|\phi\|_\infty$.

(c) *If ψ_n is a sequence of bounded, measurable functions converging either monotonically or uniformly on compact sets to a measurable function ψ then $T_{\psi_n} f \to T_\psi f$ for all $f \in \mathcal{D}(T_\psi)$.*

(d) *For any bounded, measurable ψ and any $f, g \in \mathcal{D}(T_\psi)$,*

$$d\langle E_\lambda T_\psi f, g\rangle = \psi(\lambda) d\langle E_\lambda f, g\rangle.$$

(e) *For any measurable ψ, T_ψ is self-adjoint.*

(f) *For any non-negative, measurable ψ, T_ψ is non-negative.*

Theorem 1.8.2 (Spectral theorem for self-adjoint operators)**.** *If $A : \mathcal{D}(A) \to \mathcal{H}$ is non-negative and self-adjoint then there exists a spectral decomposition $\{\mathcal{H}_\lambda\}$ such that when ψ is the identity function $x \mapsto x$ then $T_\psi = A$.*

We call the spectral decomposition in Theorem 1.8.2 the *spectral decomposition of A*. In this case, we write $\psi(A)$ instead of T_ψ.

Exercise 1.8.2. *If ψ is a polynomial then $\psi(A)$ as defined above coincides with the "usual" definition of $\psi(A)$ (e.g. $A^2 f = A(Af)$, with $\mathcal{D}(A^2) = \{f \in \mathcal{D}(A) : Af \in \mathcal{D}(A)\}$).*

The spectral decomposition of an operator is closely related to its set of eigenvalues and eigenfunctions. In particular, we will use the following characterization of the kernel of A:

Proposition 1.8.3. *If $\{\mathcal{H}_\lambda\}$ is the spectral decomposition of the non-negative, self-adjoint operator A then* $\ker A = \mathcal{H}_0$.

Proof. First, suppose that $f \in \ker A$. Then $\langle Af, f\rangle = 0$ and so by Theorem 1.8.2 and (1.3),

$$0 = \langle Af, f\rangle = \int_0^\infty \lambda d\langle E_\lambda f, f\rangle.$$

It follows that the measure $d\langle E_\lambda f, f\rangle$ is identically zero and so $\langle E_\lambda f, f\rangle$ is constant for $\lambda \in (0, \infty)$. Right-continuity (Definition 1.8.1, part (b)) of the spectral decomposition implies that $\langle E_\lambda f, f\rangle \to \|E_0 f\|^2$ as $\lambda \to 0$; exhaustiveness (part (c)) implies that $\langle E_\lambda f, f\rangle \to \|f\|^2$ as $\lambda \to \infty$. We conclude that $\|f\| = \|E_0 f\|$, and so $f \in \mathcal{H}_0$.

On the other hand, if $f \in \mathcal{H}_0$ then $E_\lambda f$ is constant for $\lambda \in (0, \infty)$ and so the measure $d\langle E_\lambda f, g\rangle$ is zero on $(0, \infty)$ for any g. Theorem 1.8.2 and (1.3) imply that

$$\langle Af, g\rangle = \int_0^\infty \lambda d\langle E_\lambda f, g\rangle = 0$$

for any $g \in \mathcal{D}(A)$. Since $\mathcal{D}(A)$ is dense, it follows that $Af = 0$. □

1.8.1 The semigroup generated by a self-adjoint operator

Now suppose that L is a non-positive, self-adjoint operator. Then $-L$ is non-negative. Since the function $\lambda \mapsto e^{-t\lambda}$ is bounded and non-negative on $\mathbb{R}^+$, we may apply the spectral theorem to define $P_t = e^{tL}$.

Proposition 1.8.4. *The family $\{P_t : t \geq 0\}$ satisfies the following properties*

(a) $P_{t+s} = P_t P_s$ *for all* $s, t \geq 0$*;*

(b) for every $t \geq 0$ *and* $f \in L^2(\mu)$*,* $\|P_t f\|_2 \leq \|f\|_2$*;*

(c) for every $t \geq 0$*,* μ *is reversible for* P_t*;*

(d) for all $f \in L^2(\mu)$*,* $P_t f \to f$ *in* $L^2(\mu)$ *as* $t \to 0$*;*

(e) for every $f \in \mathcal{D}(L)$*,* $Lf = \lim_{t\to 0} \frac{P_t f - f}{t}$*; and*

(f) for any $f \in L^2(\mu)$*, every* $t > 0$*, and every* $k \in \mathbb{N}$*,* $P_t f \in \mathcal{D}(L^k)$*.*

Note that Proposition 1.8.4 doesn't assert that P_t is a Markov operator for any t. Using Hille-Yosida theory, it is possible to fully characterize the generators L that give rise to Markov semigroups. We will not present this characterization, however, since we are interested in a special class of generators (diffusion generators) that will turn out to always produce Markov semigroups.

Proof. The semigroup property $P_{t+s} = P_t P_s$ follows from part (a) of Exercise 1.8.1. The bound $\|P_t f\|_2 \leq \|f\|_2$ follows from part (b) of Exercise 1.8.1. The reversibility of μ also follows from part (b) of Exercise 1.8.1 (recalling that the inner product here is $\langle f, g\rangle = \int_E fg\, d\mu$). The convergence $P_t f \to f$ follows from part (c) of Exercise 1.8.1, since the function $\lambda \mapsto e^{-t\lambda}$ converges pointwise and monotonically to the function with constant value 1. Since $\lambda \mapsto \frac{e^{-t\lambda}-1}{t}$ converges monotonically to the function $\lambda \mapsto -\lambda$ as $t \to 0$, the same part of Exercise 1.8.1, shows that $\frac{1}{t}(P_t f - f) \to Lf$ if $f \in \mathcal{D}(L)$.

Finally, note that $g \in \mathcal{D}(L^k)$ whenever $\int_0^\infty \lambda^{2k}\, d\langle E_\lambda g, g\rangle < \infty$. Now, part (d) implies that

$$d\langle E_\lambda P_t f, P_t f\rangle = d\langle E_\lambda P_{2t} f, f\rangle = e^{-2\lambda t} d\langle E_\lambda f, f\rangle$$

for any $f \in L^2(\mu)$. Hence,

$$\int_0^\infty \lambda^{2k}\, d\langle E_\lambda P_t f, P_t f\rangle = \int_0^\infty e^{-2t\lambda}\lambda^{2k}\, d\langle E_\lambda f, f\rangle < \infty. \qquad \square$$

1.9 Compact Markov diffusions

In all the examples we've seen so far, L was a second-order differential operator and Γ was a first-order differential operator. From now on, we will be interested only in semigroups of this form. We will encode this property (which we call the *diffusion* property) as an algebraic assumption on Γ.

Definition 1.9.1 (Diffusion carré du champ)**.** *Suppose that $\mathcal{A}$ is an algebra of functions $E \to \mathbb{R}$ that is closed under composition with smooth functions: for every $f_1, \dots, f_k \in \mathcal{A}$ and any C^∞ function $\Psi : \mathbb{R}^k \to \mathbb{R}$, the function $\Psi(f_1, \dots, f_k)$ also belongs to $\mathcal{A}$. We say that a bilinear form $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is a* diffusion carré du champ *if for every Ψ as above,*

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

Definition 1.9.1 may be seen as a sort of chain rule for Γ, which may be seen to imply a sort of second-order chain rule for L: recall that for a symmetric Markov semigroup, its generator L and carré du champ form Γ are related by the formula

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

If Γ is a diffusion carré du champ then by some manipulations, we observe that

$$L\Psi(f_1, \dots, f_k) = \sum_{i=1}^{k} \partial_i \Psi(f_1, \dots, f_k) L f_i + \sum_{i,j=1}^{k} \partial_i \partial_j \Psi(f_1, \dots, f_k) \Gamma(f_i, f_j). \tag{1.4}$$

For the special case $k = 1$, we have

$$\begin{aligned} \Gamma(\psi(f), g) &= \psi'(f)\Gamma(f, g) \\ L\psi(f) &= \psi'(f) L f + \psi''(f)\Gamma(f). \end{aligned}$$

In particular, if $f \in \mathcal{A}$ is a constant function then $Lf = 0$ and $\Gamma(f, g) = 0$ for all $g \in \mathcal{A}$.

Definition 1.9.2. *An operator L satisfying* (1.4) *is called a* diffusion generator*. A symmetric Markov semigroup whose generator is a diffusion generator is called a* diffusion semigroup*.*

Example 1.9.3. *Since we have already seen several examples of diffusion semigroups, let us consider an example which is not a diffusion semigroup. Let μ be a probability measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} x^2\, d\mu(x) < \infty$. Define*

$$(P_t f)(x) = \sum_{k\geq 0} \frac{t^k}{k!} \int_{\mathbb{R}} f(x+y)\, d\mu^{*k}(y),$$

where μ^{*k} denotes the k-fold convolution of μ with itself, and μ^{*0} is the unit mass at zero. One can check that this defines a symmetric Markov semigroup with respect to the Lebesgue measure (it is usually known as a compound Poisson *semigroup). Then*

$$\frac{P_t f(x) - f(x)}{t} = \sum_{k\geq 1} \frac{t^{k-1}}{k!} \int_{\mathbb{R}} f(x+y)\, d\mu^{*k}(y).$$

If $f \in L^2(\mathbb{R}, \mathrm{Leb})$ then the function $x \mapsto \int f(x+y)\, d\mu^{*k}(y)$ also belongs to $L^2(\mathbb{R}, \mathrm{Leb})$, and its norm is at most $\|f\|_2$. It follows that for every $f \in L^2(\mathbb{R}, \mathrm{Leb})$,

$$\left\| \frac{P_t f - f}{t} - \int_{\mathbb{R}} f(\cdot + y)\, d\mu(y) \right\|_2 \leq \sum_{k\geq 2} \frac{t^{k-1}}{k!} \|f\|_2,$$

and the right hand side converges to zero as $t \to 0$. Hence, $\mathcal{D}(L) = L^2(\mathbb{R}, \mathrm{Leb})$ and

$$(Lf)(x) = \int_{\mathbb{R}} f(x+y)\, d\mu(y).$$

This is certainly not a second-order differential operator, and it also fails to be a diffusion operator. One way to see this is that if L is a diffusion operator and ψ is constant on a neighborhood of $f(x)$ then $L\psi(f)(x) = 0$; this property clearly does not hold for the operator L above.

1.9.1 Dirichlet forms

Definition 1.9.4. *A bilinear map $\mathcal{E} : \mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E}) \to \mathbb{R}$ is a* Dirichlet form *if*

(a) $\mathcal{D}(\mathcal{E})$ is a dense subset of $L^2(\mu)$ for some measure μ;

(b) $\mathcal{E}(f, g) = \mathcal{E}(g, f)$ for all $f, g \in \mathcal{D}(\mathcal{E})$;

(c) $\mathcal{E}(f) := \mathcal{E}(f, f) \geq 0$ for all $f \in \mathcal{D}(\mathcal{E})$;

(d) *$\mathcal{D}(\mathcal{E})$ is complete with respect to the inner product $\langle f, g\rangle_{\mathcal{E}} = \int fg \, d\mu + \mathcal{E}(f,g)$; and*

(e) *for every $f \in \mathcal{D}(\mathcal{E})$, $(0 \vee f \wedge 1) \in \mathcal{E}$ and $\mathcal{E}(0 \vee f \wedge 1) \leq \mathcal{E}(f)$.*

Note that if we have a symmetric, non-negative bilinear form $\mathcal{E}$ defined on some dense $D \subset L^2(\mu)$ then we can always try to extend it to the closure of D under the norm $\|\cdot\|_{L^2(\mu)} + \mathcal{E}(\cdot)$. This may be done unambiguously if and only if $\mathcal{E}$ is *closable* on D, in the sense that if $f_n \to 0$ in $L^2(\mu)$ and f_n is Cauchy with respect to $\langle\cdot,\cdot\rangle_{\mathcal{E}}$ then $\mathcal{E}(f_n) \to 0$.

Now suppose that Γ is a diffusion carré du champ defined on an algebra $\mathcal{A}$ that is dense in $L^2(\mu)$. Suppose also that Γ is non-negative, in the sense that $\Gamma(f,f) \geq 0$ pointwise for all $f \in \mathcal{A}$. We may then define a symmetric, non-negative bilinear form $\mathcal{E}$ on $\mathcal{A}$ by

$$\mathcal{E}(f,g) = \int \Gamma(f,g) \, d\mu.$$

If we suppose for now that this is closable, we may take $\mathcal{D}(\mathcal{E})$ to be the completion of $\mathcal{A}$ under $\langle\cdot,\cdot\rangle_{\mathcal{E}}$ and extend $\mathcal{E}$ to $\mathcal{D}(\mathcal{E}) \times \mathcal{D}(\mathcal{E})$ as above.

We claim that this $\mathcal{E}$ is a Dirichlet form; to do this, it suffices to show property (e). For $\epsilon > 0$, let ψ_ϵ be a smooth approximations to the function $\psi_0(x) = 0 \vee x \wedge 1$; specifically, suppose that $0 \leq \psi_\epsilon \leq \psi_0$ and that ψ_ϵ' converges upwards to $1_{(0,1)}$. By the diffusion property of Γ, $\psi_\epsilon(f) \in \mathcal{A}$ for every $\epsilon > 0$ and

$$\Gamma(\psi_\epsilon(f)) = \psi_\epsilon'(f)^2 \Gamma(f).$$

In particular, $\psi_\epsilon(f)^2$ converges upwards to $\psi_0(f)^2$ and $\Gamma(\psi_\epsilon(f))$ converges upwards to $1_{\{f \in (0,1)\}} \Gamma(f)$. It follows that $\psi_\epsilon(f)$ is Cauchy with respect to $\langle\cdot,\cdot\rangle_{\mathcal{E}}$, and so its limit in $L^2(\mu)$ (namely, $\psi_0(f)$) belongs to $\mathcal{D}(\mathcal{E})$. Finally, $\psi_\epsilon' \leq 1$ everywhere; hence

$$\mathcal{E}(\psi_\epsilon(f)) = \int \psi_\epsilon'(f)^2 \Gamma(f) \, d\mu \leq \int \Gamma(f) \, d\mu = \mathcal{E}(f).$$

Taking $\epsilon \to 0$ shows that the $\mathcal{E}(\psi_0(f)) \leq \mathcal{E}(f)$.

As a final remark on Dirichlet forms, note that a Dirichlet form may always be used to define a non-positive, self-adjoint operator L by setting

$$\int gLf \, d\mu = -\mathcal{E}(f,g)$$

on the domain

$$\mathcal{D}(L) = \{f \in \mathcal{D}(\mathcal{E}) : \exists\, C \text{ such that } \mathcal{E}(f,g) \leq C\|g\|_2 \text{ for all } g \in \mathcal{D}(\mathcal{E})\}.$$

Indeed, such an operator is clearly non-positive, and it is self-adjoint for the same reason that the Friedrichs extension is self-adjoint. Note that $\mathcal{D}(\mathcal{E})$ plays the same role here as $\mathcal{H}_1$ played in defining the Friedrichs extension.

1.9.2 Compact Markov diffusion triple

The following definition gives the main setting for the rest of this course. The basic idea is to start with a space, a measure, and a carré du champ operator and build everything up from those. For simplicity, we will restrict ourselves initially to "compact" Markov triples. This does not mean that the space need necessarily be compact (indeed, we will never even put a topology on it). The book [BGL14] on which these notes are based go into much more detail on various less restrictive settings.

Definition 1.9.5. *Let $(E, \mathcal{F}, \mu)$ be a good measure space, where μ is a probability measure. For $\mathcal{A} \subset L^2(\mu)$, let $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ be a symmetric, bilinear form. We say that (E, μ, Γ) is a* compact Markov diffusion triple *if the following conditions are all satisfied:*

(a) $\mathcal{A}$ *is dense in* $L^2(\mu)$.

(b) $\mathcal{A}$ *is closed under composition with smooth functions* $\Psi : \mathbb{R}^k \to \mathbb{R}$ *(in particular, it is an algebra and it contains the constant functions).*

(c) $\Gamma(f) = \Gamma(f, f) \geq 0$ *pointwise for all* $f \in \mathcal{A}$.

(d) Γ *is a diffusion carré du champ.*

(e) If $\Gamma(f) = 0$ *then* f *is constant.*

Now define $\mathcal{E}(f, g) = \int \Gamma(f, g)\, d\mu$ *for* $f, g \in \mathcal{A}$.

(f) For every $f \in \mathcal{A}$*, there exists a constant* C *such that* $\mathcal{E}(f, g) \leq C\|g\|_{L^2(\mu)}$ *for all* $g \in \mathcal{A}$.

It follows that $\mathcal{E}$ *is closable, and so we may extend it to a Dirichlet form. Let* L *be the self-adjoint operator defined by* $\int_E f Lg\, d\mu = -\mathcal{E}(f, g)$ *on the domain where this makes sense. Note that (f) implies* $\mathcal{A} \subset \mathcal{D}(L)$*. Let* $P_t = e^{tL}$.

(g) $L\mathcal{A} \subset \mathcal{A}$.

(h) $P_t\mathcal{A} \subset \mathcal{A}$.

Most of the requirements in the definition above should be familiar by now. Part (e) is new; it is called the *ergodicity* condition. In the case that $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$ on some $E \subset \mathbb{R}^n$, it is equivalent to assuming that E is connected. Part (h) is also new, but in the concrete cases we consider it will mostly be easy to show.

As we have already mentioned, the book [BGL14] considers several more general settings. Essentially, it turns out to be desirable to weaken assumption (b) and remove assumption (h). For some motivation, suppose that $E = \mathbb{R}^n$. Under assumption (b), $\mathcal{A}$ contains constant functions. Hence, the assumption that $\mathcal{A} \subset L^2(\mu)$ can only hold if μ is finite; in particular, the heat semigroup is already excluded. If we restricted the smooth functions Ψ to those satisfying $\Psi(0) = 0$, we could try to take $\mathcal{A} = C_c^\infty(\mathbb{R}^n)$. However, the heat semigroup maps compactly supported functions to fully supported functions, so (h) will fail.

To deal with this issue, [BGL14] introduces two algebras: they begin by defining everything on an "inner" algebra $\mathcal{A}_0$ (think of $C_c^\infty(E)$) and then later extending everything to an "outer" algebra $\mathcal{A}$ (think of $C^\infty(E)$). Only the inner algebra $\mathcal{A}_0$ is assumed to belong to $L^2(\mu)$, and it is only assumed to be closed under composition with smooth functions vanishing at zero. Part (h) is assumed for $\mathcal{A}$, but not for $\mathcal{A}_0$.

Although the setting with two algebras is rather more general than the one that we present, the main ideas may already be developed in the setting of Definition 1.9.5. Essentially all of the interesting results that we prove will extend to the more general setting, but the proofs usually have some extra technicalities (for example, the assumption that $\mathcal{A} \subset L^2(\mu)$ will simplify our life substantially. Therefore, we will restrict this presentation to compact diffusion Markov triples (which we will usually abbreviate to *compact Markov triple*).

1.10 Properties of diffusion semigroups

1.10.1 The Markov property

We originally constructed semigroups P_t from generators L using spectral theory, from which it wasn't particularly clear how to check if the resulting semigroup is a Markov semigroup. One nice consequence of Definition 1.9.5 is that the resulting semigroup is always a Markov semigroup:

Proposition 1.10.1. *Let (E, μ, Γ) be a compact Markov diffusion triple and let P_t be its semigroup. Then P_t is a Markov semigroup.*

Proof. First, we show that $P_t \mathbb{1} = \mathbb{1}$. For this, note that because of Definition 1.9.5 (d), $L\mathbb{1} = 0$. Letting $\{\mathcal{H}_\lambda : \lambda \geq 0\}$ be the spectral decomposition of $-L$, it follows (from Proposition 1.8.3) that $\mathbb{1} \in \mathcal{H}_0$. Then $\langle E_\lambda \mathbb{1}, f\rangle = \langle \mathbb{1}, f\rangle = \int f\, d\mu$ for all $\lambda \geq 0$ and all $f \in L^2(\mu)$; hence,

$$\langle P_t \mathbb{1}, f\rangle = \langle E_0 \mathbb{1}, f\rangle + \int_0^\infty e^{-\lambda t}\, d\langle E_\lambda \mathbb{1}, f\rangle = \langle \mathbb{1}, f\rangle.$$

It follows that $P_t \mathbb{1} = \mathbb{1}$.

In order to show that P_t preserves non-negativity, we will show the equivalent statement that $P_t f \leq 1$ whenever $f \leq 1$. First, introduce the resolvent operator $R_\lambda = (\lambda \operatorname{Id} - L)^{-1}$, which may be defined using spectral theory as $\psi_\lambda(-L)$ where $\psi_\lambda(x) = 1/(\lambda + x)$. Since

$$\sum_{k\geq 0} \frac{t^k}{k!} \left(\frac{\lambda^2}{\lambda + x}\right)^k = \sum_{k \geq 0} \frac{t^k}{k!} \left(\lambda - \frac{x}{1 + x/\lambda}\right)^k = \exp\left(t\lambda - \frac{tx}{1 + x/\lambda}\right),$$

it follows that

$$\lim_{\lambda \to \infty} e^{-t\lambda} \sum_{k \geq 0} \frac{t^k}{k!} (\lambda^2 R_\lambda)^k = P_t.$$

Hence, it suffices to show that $\lambda R_\lambda f \leq 1$ whenever $f \leq 1$. Let $g = \lambda R_\lambda f$, so that $\lambda g = Lg + \lambda f$. Note that such a g is automatically in $\mathcal{D}(L)$ (and hence also in $\mathcal{D}(\mathcal{E})$). Now let $\psi : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying $\psi(0) = 0$ and $\|\psi'\|_\infty \leq 1$. Then

$$\lambda \int g(g - \psi(g))\, d\mu = \int Lg(g - \psi(g))\, d\mu + \lambda \int f(g - \psi(g))\, d\mu.$$

Now,

$$\int Lg(g - \psi(g))\, d\mu = -\int \Gamma(g, g - \psi(g))\, d\mu = -\int (1 - \psi'(g))\Gamma(g, g)\, d\mu \leq 0.$$

Hence,

$$\int (f - g)(g - \psi(g))\, d\mu \geq 0.$$

Now take ψ to approach the function $x \mapsto x \wedge 1$; we obtain

$$\int (1 - g)(g - g \wedge 1) \geq \int (f - g)(g - g \wedge 1)\, d\mu \geq 0,$$

and so it must be that $g \leq 1$ μ-almost surely. $\square$

1.10.2 Ergodicity

Those who are familiar with the notion of ergodicity in other settings may have been surprised that we used the same word to describe part (e) of Definition 1.9.5. However, this assumption does turn out to be closely related to a more standard usage of the term. Indeed, Definition 1.9.5 (e) is the main property driving the convergence in the following proposition:

Proposition 1.10.2. *For a compact Markov triple* (E, μ, Γ) *with semigroup* P_t *and any* $f \in L^2(\mu)$, $P_t f \to \int_E f \, d\mu$ *in* $L^2(\mu)$ *as* $t \to \infty$.

Proof. Let L be the generator of P_t, and recall that $Lg = 0$ for all constant functions g. Definition 1.9.5 (e) implies that only constant functions belong to $\ker L$. That is, if $\{\mathcal{H}_\lambda\}$ is the spectral decomposition of $-L$ then (by Proposition 1.8.3) $\mathcal{H}_0$ is the space of constant functions. In particular, $E_0 f$ is the constant function $x \mapsto \int_E f \, d\mu$.

By the right-continuity of the spectral decomposition, $E_\lambda f \to E_0 f$ as $\lambda \to 0$. Fix f and $\epsilon > 0$, and choose $\lambda_\epsilon > 0$ small enough so that $\|E_{\lambda_\epsilon} f - E_0 f\| \le \epsilon$. According to the definition of P_t,

$$\langle P_{2t} f, f \rangle = \langle E_0 f, f \rangle + \int_0^\infty e^{-2t\lambda} d\langle E_\lambda f, f \rangle.$$

Since $d\langle E_\lambda f, f \rangle$ is a positive measure, we may insert the bound $e^{-2t\lambda} \le 1$ for $\lambda \le \lambda_\epsilon$ and $e^{-2t\lambda} \le e^{-2t\lambda_\epsilon}$ otherwise. Then

$$\langle P_{2t} f, f \rangle \le \langle E_{\lambda_\epsilon} f, f \rangle + e^{-2t\lambda_\epsilon}(\|f\|^2 - \langle E_{\lambda_\epsilon} f, f \rangle) \le \|E_0 f\|^2 + \epsilon + e^{-2t\lambda_\epsilon}\|f\|^2.$$

On the other hand, the reversibility of μ with respect to P_t implies that $\langle P_{2t} f, f \rangle = \|P_t f\|^2$, and so

$$\limsup_{t \to \infty} \|P_t f\|^2 \le \|E_0 f\|^2 + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have $\limsup \|P_t f\|^2 \le \|E_0 f\|^2$. On the other hand, the definition of $P_t f$ also implies that

$$\langle P_t f, E_0 f \rangle = \langle E_0 f, E_0 f \rangle + \int_0^\infty e^{-t\lambda} d\langle E_\lambda f, E_0 f \rangle,$$

where the integral is zero because $\langle E_\lambda f, E_0 f \rangle$ is constant for $\lambda > 0$. Hence, $\langle P_t f, E_0 f \rangle = \|E_0 f\|^2$ for all t. It follows that $\|P_t f\| \ge \|E_0 f\|$ for all t; hence, $\lim_{t \to \infty} \|P_t f\| = \|E_0 f\|$ and $P_t f \to E_0 f$. $\square$

1.10.3 Examples

Example 1.10.3 (The heat semigroup on $\mathbb{R}^n$)**.** *The triple* $(\mathbb{R}^n, \mathrm{Leb}, \Gamma)$ *corresponding to the heat semigroup is not a compact Markov triple because* Leb *is not finite.*

Example 1.10.4 (The wrapped heat semigroup)**.** *Take $E = [0,1]$ and let μ be the Lebesgue measure. Let $\mathcal{A}$ be the set of $f \in C^\infty[0,1]$ satisfying $f(0) = f(1)$, and let $\Gamma(f,g) = f'g'$ for $f, g \in \mathcal{A}$. Then (E, μ, Γ) is a compact Markov triple. All of the properties are immediate, except possibly for (h), which follows from our earlier explicit form of P_t.*

In this example, we can write the $\mathcal{E}$ and L (on their whole domains) explicitly in terms of Fourier series: let

$$f(x) = f_0 + \sum_{k \geq 1} \hat{f}_k \sin(2\pi k x) + \check{f}_k \cos(2\pi k x),$$

and similarly for g, where the sum converges in L^2. In order to compute the Fourier expansion of f', note that

$$\int_0^1 f'(x) \cos(2\pi k x)\, dx = 2\pi k \int_0^1 f(x) \sin(2\pi k x)\, dx,$$

and similarly for $\int_0^1 f'(x) \sin(2\pi k x)\, dx$. Then we have

$$f'(x) = 2\pi \sum_{k \geq 1} k \hat{f}_k \cos(2\pi k x) - k \check{f}_k \sin(2\pi k x).$$

Since the terms $\cos(2\pi k x)$ and $\sin(2\pi k x)$ are all orthogonal and have L^2 norm $1/\sqrt{2}$,

$$\mathcal{E}(f,g) = 2\pi^2 \sum_{k \geq 1} k^2 \hat{f}_k \hat{g}_k + k^2 \check{f}_k \check{g}_k.$$

It's clear from this expression that the domain of $\mathcal{E}$ is the set of $f \in L^2(\mu)$ such that

$$\sum_{k \geq 1} k^2 (\hat{f}_k^2 + \check{f}_k^2) < \infty.$$

Recall that $\mathcal{D}(L)$ is the set of $f \in \mathcal{D}(\mathcal{E})$ such that there is a constant C for which $\mathcal{E}(f,g) \leq C\|g\|_2$ for all $g \in \mathcal{D}(\mathcal{E})$. From the expression for $\mathcal{E}$ above and the Cauchy-Schwarz inequality, $\mathcal{D}(L)$ is the set of f such that

$$\sum_{k \geq 1} k^4 (\hat{f}_k^2 + \check{f}_k^2) < \infty.$$

On this set,

$$Lf(x) = \sum_{k \geq 1} k^2 (\hat{f}_k \cos(2\pi kx) + \check{f}_k \sin(2\pi kx)).$$

Example 1.10.5 (The reflected heat semigroup)**.** *The reflected heat semigroup is similar to the wrapped one, except that we let $\mathcal{A}$ be the set of $f \in C^\infty[0,1]$ satisfying $f'(0) = f'(1) = 0$. This leads to some differences in $\mathcal{E}$ and L; for example, the difference in boundary conditions leads to a different relationship between the Fourier expansions of f and f':*

$$f'(x) = \sum_{k \geq 1} (f(1) - f(0) + 2\pi k \hat{f}_k) \cos(2\pi kx) - 2\pi k \check{f}_k \sin(2\pi kx).$$

Hence, the domain of $\mathcal{E}$ becomes the set of $f \in L^2(\mu)$ such that there exists some $a \in \mathbb{R}$ with

$$\sum_{k \geq 1} (a + k\hat{f}_k)^2 + k^2 \check{f}_k^2 < \infty.$$

We leave the rest of this example as an exercise.

Example 1.10.6 (The Ornstein-Uhlenbeck semigroup on $\mathbb{R}$)**.** *The Ornstein-Uhlenbeck semigroup on $\mathbb{R}$ is associated with the triple $(\mathbb{R}, \gamma_1, \Gamma)$, where γ_1 is the standard Gaussian measure on $\mathbb{R}$ and $\Gamma(f,g) = f'g'$. We take $\mathcal{A}$ to be the set of smooth functions whose derivatives of all orders decay faster than any polynomial as $x \to \infty$. One can check that $\mathcal{A}$ is preserved under P_t by using the explicit formula for P_t that we gave earlier. Recall that $Lf(x) = f''(x) - xf'(x)$.*

Like the wrapped heat semigroup, the Ornstein-Uhlenbeck semigroup may be studied using explicit spectral expansions. For this, we introduce the Hermite polynomials, defined by

$$h_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2},$$

or equivalently in terms of the generating function

$$e^{xt - t^2/2} = \sum_{k=0}^{\infty} h_k(x) \frac{t^k}{k!}.$$

Note that $h_0(x) = 1$, $h_1(x) = x$, $h_2(x) = x^2 - 1$, and that in general h_k is a polynomial of degree k. One can check that $\{h_k, k \geq 0\}$ is an orthogonal basis of $L^2(\mathbb{R}, \gamma_1)$, and that $\|h_k\|_2^2 = k!$. The h_k also satisfy the recursion

$$Lh_k(x) = h_k''(x) - xh_k'(x) = -kh_k(x).$$

It follows that if $f = \sum_{k\geq 0} \hat{f}_k h_k/\sqrt{k!}$ then

$$
\begin{aligned}
P_t f &= \sum_{k\geq 0} e^{-tk} \hat{f}_k \frac{h_k}{\sqrt{k!}} \\
Lf &= -\sum_{k\geq 0} k \hat{f}_k \frac{h_k}{\sqrt{k!}} \\
\mathcal{E}(f) &= \sum_{k\geq 0} k \hat{f}_k^2,
\end{aligned}
$$

where $f \in \mathcal{D}(L)$ whenever $\sum k^2 \hat{f}_k < \infty$ and $f \in \mathcal{D}(\mathcal{E})$ whenever $\sum k \hat{f}_k < \infty$.

1.11 Γ_2, curvature, and dimension

Recall that Γ may be written in terms of L as

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

We define a bilinear form Γ_2 by a very similar formula, but where multiplication is replaced by Γ.

Definition 1.11.1. *For $f, g \in \mathcal{A}$, define*

$$2\Gamma_2(f, g) = L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(g, Lf).$$

By similar formulas, one can define a whole sequence Γ_k of bilinear forms. It isn't clear how useful this is, although Γ_3 has recently received some attention [LNP15]. If Γ is a diffusion carré du champ then one can derive a chain rule for Γ_2. We will explore the general case later, but the single-variable version is

$$\Gamma_2(\psi(f)) = \psi'(f)^2\Gamma_2(f) + \psi'(f)\psi''(f)\Gamma(f, \Gamma(f)) + \psi''(f)^2\Gamma(f)^2$$

for every smooth $\psi : \mathbb{R} \to \mathbb{R}$.

Definition 1.11.2. *For $\rho \in \mathbb{R}$ and $n \in [1, \infty]$, a compact Markov triple (E, μ, Γ) is said to satisfy the* curvature-dimension condition $\mathrm{CD}(\rho, n)$ *if*

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

μ-almost everywhere for every $f \in \mathcal{A}$.

The motivation for Definition 1.11.2 is probably unclear so far. As we will show later, if (E, μ, Γ) is the compact Markov triple corresponding to the heat semigroup on a compact n-dimensional Riemannian manifold with Ricci curvature bounded below by ρ then (E, μ, Γ) satisfies $\mathrm{CD}(\rho, n)$; this is the motivation for the terminology.

Note that $\mathrm{CD}(\rho, n)$ trivially implies $\mathrm{CD}(\rho', n')$ for every $\rho' \le \rho$ and $n' \ge n$. However, there is not necessarily a best possible ρ or a best possible n for a given Markov triple – it could be that one can increase ρ at the cost of decreasing n, or vice versa.

1.11.1 Examples

Example 1.11.3 (The heat semigroup on $\mathbb{R}^n$)**.** *Technically, we defined* $\mathrm{CD}(\rho, n)$ *only for compact Markov triples, and the heat semigroup on* $\mathbb{R}^n$ *is not a compact Markov triple. However, there is no obstacle in extending Definition 1.11.2 to cover this case. Since* $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$ *and* $Lf = \Delta f$*, we have*

$$2\Gamma_2(f, g) = \Delta \langle \nabla f, \nabla g \rangle - \langle \nabla f, \nabla \Delta g \rangle - \langle \nabla g, \nabla \Delta f \rangle = 2\langle \nabla^2 f, \nabla^2 g \rangle,$$

where $\nabla^2 f$ *denotes the matrix* $(\partial_i \partial_j f)_{i,j=1}^n$ *and* $\langle A, B \rangle = \mathrm{tr}(A^T B)$ *for* $n \times n$ *matrices* A *and* B*. Hence, the* $\mathrm{CD}(\rho, k)$ *condition becomes*

$$\|\nabla^2 f\|_2^2 \ge \rho \|\nabla f\|_2^2 + \frac{1}{k}(\Delta f)^2. \tag{1.5}$$

For any $\rho > 0$*, this condition does not hold for all* f *in* C_c^∞ *(for example, by taking* f *to be linear on some neighborhood, so* ∇f *is non-zero but* $\nabla^2 f$ *and* Lf *are zero). As for* k*, we have the inequality* $\|A\|_2^2 \ge \frac{1}{n} \mathrm{tr}(A)^2$ *for any* $n \times n$ *matrix* A*. Since* $Lf = \mathrm{tr}\, \nabla^2 f$*, it follows that* (1.5) *is satisfied for* $\rho = 0$ *and* $k = n$*; hence the heat semigroup on* $\mathbb{R}^n$ *satisfies* $\mathrm{CD}(0, n)$*. Note also that* n *is the smallest value of* k *for which* (1.5) *holds, by considering a function* f *satisfying* $f(x) = x^2$ *on some neighborhood.*

Example 1.11.4 (The Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$)**.** *For the Ornstein-Uhlenbeck semigroup,* $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$ *and* $Lf = \Delta f - \langle x, \nabla f \rangle$*. Compared to the computation for the heat semigroup on* $\mathbb{R}^n$*, note that the only difference will involve the term* $\langle x, \nabla f \rangle$ *in* L*. Hence,*

$$\begin{aligned} 2\Gamma_2(f, g) &= 2\langle \nabla^2 f, \nabla^2 g \rangle - \langle x, \nabla \langle \nabla f, \nabla g \rangle \rangle + \langle \nabla f, \nabla \langle x, \nabla g \rangle \rangle + \langle \nabla g, \nabla \langle x, \nabla f \rangle \rangle \\ &= 2\langle \nabla^2 f, \nabla^2 g \rangle + 2\langle \nabla f, \nabla g \rangle. \end{aligned}$$

In particular, $\Gamma_2(f) = \|\nabla^2 f\|_2^2 + \Gamma(f)$*, and so it is immediately obvious that this satisfies* $\mathrm{CD}(1,\infty)$*. By choosing* $f(x) = x_1$ *(or more precisely, choosing a sequence of functions in* $\mathcal{A}$ *that equal* x_1 *on larger and larger compact sets), we have* $\Gamma_2(f) = 1$*,* $\Gamma(f) = 1$*, and* $Lf = -x_1$*; since this is unbounded, the Ornstein-Uhlenbeck semigroup does not satisfy* $\mathrm{CD}(\rho, n)$ *for any finite* n*:* $\mathrm{CD}(1,\infty)$ *is the best possible curvature-dimension condition.*

Example 1.11.5 (The ultraspheric semigroup on $(-1,1)$)**.** *For* $n \geq 1$*, consider the operator* L_n *that acts on* $C^\infty[-1,1]$ *as*

$$(L_n f)(x) = (1-x^2)f''(x) - nxf'(x).$$

The corresponding carré du champ operator is $\Gamma(f,g) = (1-x^2)f'g'$*. One can easily see using integration by parts that*

$$\int_{-1}^{1} (1-x^2)^{n/2-1} L_n f \, dx = 0$$

for every smooth f*. Hence, this defines a compact Markov triple with invariant measure* $\mu = C_n(1-x^2)^{n/2-1}$*, where* C_n *is a normalizing constant. As we will see later, this triple can be obtained by starting with the heat semigroup on the sphere* S^n *and projecting it on some axis. For now, we just consider this an an example of the curvature-dimension condition.*

After some slightly tedious computations, one can check that

$$\Gamma_2(f) = (n-1)\Gamma(f) + \frac{1}{n}(L_n f)^2 + \left(1 - \frac{1}{n}\right)(1-x^2)^2(f'')^2.$$

Hence, this triple satisfies $\mathrm{CD}(n-1, n)$ *(and this is the best possible in either parameter).*

Chapter 2

Poincaré inequalities

Curvature-dimension conditions turn out to imply many important inequalities. We will begin with the observation that a $\mathrm{CD}(\rho, \infty)$ condition is equivalent to exponential decay of various quantities along the semigroup P_t.

Proposition 2.0.1. *Let P_t be the semigroup of a compact Markov triple. The following are equivalent:*

(a) the curvature-dimension condition $\mathrm{CD}(\rho, \infty)$ holds for some $\rho \in \mathbb{R}$;

(b) for every $f \in \mathcal{A}$ and every $t \geq 0$,

$$\Gamma(P_t f) \leq e^{-2\rho t} P_t \Gamma(f);$$

(c) for every function $f \in \mathcal{A}$ and every $t \geq 0$,

$$P_t(f^2) - (P_t f)^2 \leq \frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma(f);$$

(d) for every function $f \in \mathcal{A}$ and every $t \geq 0$,

$$P_t(f^2) - (P_t f)^2 \geq \frac{e^{2\rho t} - 1}{\rho} \Gamma(P_t f).$$

In the last two conditions, if $\rho = 0$ then we take $(1-e^{-2\rho t})/\rho = (e^{2\rho t}-1)/\rho = 2t$.

Proof. To see that (a) implies (b), choose $f \in \mathcal{A}$ and define

$$\Lambda(s) = e^{-2\rho s} P_s \Gamma(P_{t-s} f).$$

Applying the chain rule and the definition of Γ_2,

$$\Lambda'(s) = 2e^{-2\rho s} P_s(\Gamma_2(P_{t-s} f) - \rho \Gamma(P_{t-s} f)).$$

Under the $\mathrm{CD}(\rho, \infty)$ assumption, $\Lambda'(s)$ is non-negative pointwise and so $\Lambda(t) \geq \Lambda(0)$, which is equivalent to claim (b).

To prove that (b) implies (c), choose $f \in \mathcal{A}$ and define

$$\Lambda(s) = P_s(P_{t-s} f)^2.$$

By the chain rule and the definition of Γ, $\Lambda'(s) = 2P_s\Gamma(P_{t-s}f)$. Hence,

$$\begin{aligned}
\Lambda(t) - \Lambda(0) &= 2 \int_0^t P_s \Gamma(P_{t-s} f)\, ds \\
&\leq 2 \int_0^t e^{-2\rho(t-s)} P_t \Gamma(f)\, ds \\
&= \frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma(f),
\end{aligned}$$

where the inequality follows from (b). This proves (c).

The proof that (b) implies (d) is very similar, except that instead of the inequality $P_s\Gamma(P_{t-s}f) \leq e^{-2\rho(t-s)} P_t \Gamma(f)$ that we used above, we use the inequality $P_s\Gamma(P_{t-s}f) \geq e^{2\rho s}\Gamma(P_t f)$. This yields

$$\Lambda(t) - \Lambda(0) \geq 2 \int_0^t e^{2\rho s} \Gamma(P_t f)\, ds = \frac{e^{2\rho t} - 1}{\rho} \Gamma(P_t f).$$

Now let us show that (c) implies (a). Note that the definition of L allows us to write

$$P_t h = h + tLh + \frac{t^2}{2} L^2 h + o(t^2) \tag{2.1}$$

as $t \to 0$ for any $h \in \mathcal{A}$. Recall that the definition of L involves $L^2(\mu)$ convergence, which implies pointwise convergence μ-almost everywhere. Hence, the equation above makes sense pointwise μ-almost everywhere. Applying (2.1) for $h = f$ and $h = f^2$,

$$\begin{aligned}
P_t(f^2) - (P_t f)^2 &= tL(f^2) + \frac{t^2}{2} L^2(f^2) - 2tfLf - t^2(Lf)^2 - t^2 f L^2 f + o(t^2) \\
&= 2t\Gamma(f) + \frac{t^2}{2} L^2(f^2) - t^2(Lf)^2 - t^2 f L^2 f + o(t^2).
\end{aligned}$$

On the other hand,

$$\frac{1-e^{-2\rho t}}{\rho} P_t\Gamma(f) = 2t\Gamma(f) - 2\rho t^2\Gamma(f) + 2t^2 L\Gamma(f) + o(t^2).$$

Assuming (c) and taking $t \to 0$, we have

$$\frac{1}{2}L^2(f^2) - (Lf)^2 - fL^2 f \leq -2\rho\Gamma(f) + 2L\Gamma(f).$$

Now, note that the right hand side above can also be written as $2\Gamma(f, Lf) + L\Gamma(f)$. After rearranging, we conclude that

$$L\Gamma(f) - 2\Gamma(f, Lf) \geq 2\rho\Gamma(f),$$

which is exactly the $\mathrm{CD}(\rho, \infty)$ condition.

The proof that (d) implies (a) is analogous, and we omit it. □

We will study part (c) of Proposition 2.0.1 in much more detail. To begin with, suppose that $\rho > 0$ and take the limit as $t \to \infty$. By ergodicity (Proposition 1.10.2), $P_t h \to \int h \, d\mu$ as $t \to \infty$, and so

$$\int f^2 \, d\mu - \left(\int f, d\mu\right)^2 \leq \frac{1}{\rho} \int \Gamma(f) \, d\mu.$$

We call such an inequality a Poincaré inequality.

Definition 2.0.2. *Let μ be a probability measure and $\mathcal{E}$ be a Dirichlet form on $L^2(\mu)$. We say that μ and $\mathcal{E}$ satisfy a* Poincaré inequality *with constant C if*

$$\int f^2 \, d\mu - \left(\int f \, d\mu\right)^2 \leq C\mathcal{E}(f) \tag{2.2}$$

for every $f \in \mathcal{D}(\mathcal{E})$. The Poincaré constant *of μ and $\mathcal{E}$ is the smallest C such that* (2.2) *holds for every $f \in \mathcal{D}(\mathcal{E})$.*

Usually, we will apply Definition 2.0.2 in the context of a compact Markov triple. In this case, since $\mathcal{A}$ is dense in $\mathcal{D}(\mathcal{E})$ with respect to $\langle \cdot, \cdot \rangle_{\mathcal{E}}$, it suffices to check (2.2) for all $f \in \mathcal{A}$. We will also sometimes say that μ and Γ satisfy a Poincaré inequality, in which case it is implied that $\mathcal{E}(f) = \int \Gamma(f) \, d\mu$.

As we have argued above, if a compact Markov triple satisfies $\mathrm{CD}(\rho, \infty)$ then it satisfies a Poincaré inequality with constant $\frac{1}{\rho}$. In fact, for finite t Proposition 2.0.1 gives something more precise, which we will call a *local Poincaré inequality* for the semigroup P_t.

Corollary 2.0.3. *The compact Markov triple* (E, μ, Γ) *satisfies* $\mathrm{CD}(\rho, \infty)$ *if and only if for every* $t \geq 0$ *and* μ*-almost every* $x \in E$*, the measure* $p_t(x, \cdot)$ *satisfies a Poincaré inequality with constant* $\frac{1-e^{-2\rho t}}{\rho}$*, where* p_t *is the kernel representation of the semigroup* P_t*.*

2.1 Poincaré inequalities under $\mathrm{CD}(\rho, n)$

Recall that if $n < \infty$ then the condition $\mathrm{CD}(\rho, n)$ is more restrictive than the condition $\mathrm{CD}(\rho, \infty)$. It turns out that under $\mathrm{CD}(\rho, n)$ for $\rho > 0$ and $n \in (1, \infty)$, the finiteness of n lets us get a better Poincaré inequality that we did for $\mathrm{CD}(\rho, \infty)$. When specialized to the case of the heat semigroup on a compact Riemannian manifold, this is known as Lichnerowicz's theorem.

Theorem 2.1.1. *Let* (E, μ, Γ) *be a compact Markov triple. If the curvature-dimension condition* $\mathrm{CD}(\rho, n)$ *holds for some* $\rho > 0$*,* $n > 1$ *then* μ *satisfies a Poincaré inequality with constant* $C = \frac{n-1}{\rho n}$*.*

Lemma 2.1.2. *Suppose that the compact Markov triple* (E, μ, Γ) *satisfies* $\mathrm{CD}(\rho, \infty)$ *for some* $\rho > 0$*. Then* μ *satisfies a Poincaré inequality with constant* $C > 0$ *if and only if*

$$\int_E \Gamma(f)\, d\mu \leq C \int_E (Lf)^2\, d\mu \tag{2.3}$$

for all $f \in \mathcal{D}(L)$*.*

Proof. Fix $f \in \mathcal{A}$ and consider the function $\Lambda(t) = \int_E (P_t f)^2\, d\mu$. Then

$$\Lambda'(t) = -2 \int_E \Gamma(P_t f)\, d\mu$$

and

$$\Lambda''(t) = -4 \int_E \Gamma(P_t f, L P_t f) = 4 \int_E (L P_t f)^2\, d\mu.$$

Under the assumption (2.3), we have $\Lambda''(t) \geq -\frac{2}{C} \Lambda'(t)$ for all $t \geq 0$. Hence,

$$\begin{aligned} \int f^2\, d\mu - \left(\int f\, d\mu \right)^2 &= -\int_0^\infty \Lambda'(t)\, dt \\ &\leq \frac{C}{2} \int_0^\infty \Lambda''(t)\, dt \\ &= -\frac{C}{2} \Lambda'(0) \\ &= C \int_E \Gamma(f)\, d\mu. \end{aligned}$$

Here, we are implicitly using the $\mathrm{CD}(\rho, \infty)$ condition in order to conclude that $\Gamma(P_t f)$ is decaying exponentially in t, so all of the integrals converge. This proves one of the claimed directions under the assumption that $f \in \mathcal{A}$; for $f \in \mathcal{D}(L)$, we use a standard approximation argument.

For the converse, suppose that μ satisfies a Poincaré inequality with constant C, and fix some $f \in \mathcal{D}(L)$ with mean zero. By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\int_E \Gamma(f)\, d\mu &= \int_E f(-Lf)\, d\mu \\
&\leq \sqrt{\int_E f^2\, d\mu \int_E (Lf)^2\, d\mu} \\
&\leq \sqrt{C \int_E \Gamma(f)\, d\mu \int_E (Lf)^2\, d\mu},
\end{aligned}$$

which proves (2.3). □

Proof of Theorem 2.1.1. Integrating the $\mathrm{CD}(\rho, n)$ condition yields

$$\int_E \Gamma_2(f)\, d\mu \geq \rho \int_E \Gamma(f)\, d\mu + \frac{1}{n} \int_E (Lf)^2\, d\mu.$$

Since $\int_E Lh\, d\mu = 0$ for every $h \in \mathcal{A}$, we have

$$\begin{aligned}
\int_E \Gamma_2(f) &= \frac{1}{2} \int_E L\Gamma(f)\, d\mu - \int_E \Gamma(f, Lf)\, d\mu \\
&= \frac{1}{2} \int_E L\Gamma(f)\, d\mu - \frac{1}{2} \int_E L(fLf)\, d\mu + \frac{1}{2} \int_E (Lf)^2 + fL^2 f\, d\mu \\
&= \int_E (Lf)^2\, d\mu.
\end{aligned}$$

Applying this to the integrated $\mathrm{CD}(\rho, n)$ condition, we have

$$\frac{n-1}{n} \int_E (Lf)^2\, d\mu \geq \rho \int_E \Gamma(f)\, d\mu.$$

Finally, apply Lemma 2.1.2 to this inequality. □

Note that (unlike the $\mathrm{CD}(\rho, \infty)$ case), we do not obtain a characterization of $\mathrm{CD}(\rho, n)$ in terms of any Poincaré inequality. To see that there is no such characterization, recall that the Ornstein-Uhlenbeck semigroup satisfies $\mathrm{CD}(1, \infty)$ (and hence also a family of local Poincaré inequalities), but not $\mathrm{CD}(\rho, n)$ for any $n < \infty$.

2.2 Spectral gap

Definition 2.2.1. *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$. Its* spectrum $\sigma(A)$ *is the set of $\lambda \in \mathbb{R}$ such that $A - \lambda \,\mathrm{Id}$ does not have a bounded inverse.*

Usually, one defines a spectrum over the complex numbers for operators acting on complex Hilbert spaces; it is then possible to show that self-adjoint operators have an entirely real spectrum. To save time, we will skip this step and only define the real spectrum for self-adjoint operators.

Exercise 2.2.1. *If A is self-adjoint and non-negative then $\sigma(A) \subset \mathbb{R}_+$.*

One easily sees that every eigenvalue of A belongs to its spectrum: if $Af = \lambda f$ for $f \neq 0$ then $(A - \lambda \,\mathrm{Id})f = 0$ and so $A - \lambda \,\mathrm{Id}$ certainly does not have a bounded inverse. The converse is not true – not every $\lambda \in \sigma(A)$ is necessarily an eigenvalue – but an approximate converse is true.

Lemma 2.2.2. *If A is self-adjoint then $\lambda \in \sigma(A)$ if and only if there is some sequence $f_n \in \mathcal{D}(A)$ such that $\|f_n\| = 1$ for all n but $\|(A - \lambda \,\mathrm{Id})f_n\| \to 0$.*

In the case that λ is an eigenvalue of A one can just take the sequence $f_n = f$, where f is a normalized eigenfunction with eigenvalue λ.

Proof. One direction is clear: if there is a sequence f_n as in the lemma then any inverse of $A - \lambda \,\mathrm{Id}$ could not possibly be bounded; hence $\lambda \in \sigma(A)$.

To show the other direction, fix $\lambda \in \sigma(A)$; let $\mathcal{R}$ denote the range of $A - \lambda \,\mathrm{Id}$. One way that $A - \lambda \,\mathrm{Id}$ could fail to have an inverse is that it could fail to be injective. In that case, $A - \lambda \,\mathrm{Id}$ has a non-trivial null space, and so λ is an eigenvalue of A. As we already discussed, this implies the existence of a sequence f_n as above.

It remains to consider the case that $A - \lambda \,\mathrm{Id}$ is injective. In this case, we claim that $\mathcal{R}$ is dense in $\mathcal{H}$. Indeed, suppose that $h \in \mathcal{R}^\perp$. Then for every $f \in \mathcal{D}(A)$, $\langle h, (A - \lambda \,\mathrm{Id})f\rangle = 0$. It follows that $h \in \mathcal{D}(A^*) = \mathcal{D}(A)$ and that

$$0 = \langle h, (A - \lambda \,\mathrm{Id})f\rangle = \langle (A - \lambda \,\mathrm{Id})h, f\rangle$$

for every $f \in \mathcal{D}(A)$; hence, $(A - \lambda \,\mathrm{Id})h = 0$. Since $A - \lambda \,\mathrm{Id}$ was assumed injective, it follows that $h = 0$ and $\mathcal{R}^\perp = \{0\}$. In particular, $\mathcal{R}$ is dense in $\mathcal{H}$.

Since $A - \lambda \,\mathrm{Id}$ is injective, we can certainly define an inverse $(A - \lambda \,\mathrm{Id})^{-1} : \mathcal{R} \to \mathcal{H}$. Suppose first that this inverse is bounded on $\mathcal{R}$. For any $g \in \mathcal{H}$, the density of $\mathcal{R}$ implies we may choose a sequence $g_n \in \mathcal{R}$ such that $g_n \to g$.

Setting $f_n = (A - \lambda\,\mathrm{Id})^{-1} g_n$, the boundedness of $(A - \lambda\,\mathrm{Id})^{-1}$ implies the sequence f_n converges, to f say. Since A is a closed operator and both f_n and $(A - \lambda\,\mathrm{Id}) f_n$ converge, it follows that $f = \lim f_n \in \mathcal{D}(A)$ and

$$(A - \lambda\,\mathrm{Id}) f = \lim_{n\to\infty} (A - \lambda\,\mathrm{Id}) f_n = \lim_{n\to\infty} g_n = g.$$

Hence, $g \in \mathcal{R}$ and so $\mathcal{R} = \mathcal{H}$. In particular, $(A - \lambda\,\mathrm{Id})^{-1}$ is a bounded operator on all of $\mathcal{H}$, contradicting our assumption that $\lambda \in \sigma(A)$.

To conclude, $\lambda \in \sigma(A)$ implies that the operator $(A - \lambda\,\mathrm{Id})^{-1} : \mathcal{R} \to \mathcal{H}$ is unbounded. Hence, there exist $g_n \in \mathcal{R}$, $\|g_n\| = 1$ such that $\|(A - \lambda\,\mathrm{Id})^{-1} g_n\| \to \infty$. Setting f_n to be $(A - \lambda\,\mathrm{Id})^{-1} g_n$ (renormalized to have norm 1) gives the desired sequence. □

Proposition 2.2.3. *If the compact Markov triple (E, μ, Γ) satisfies a Poincaré inequality with constant C then $\sigma(L)$ belongs to $(-\infty, -\frac{1}{C}] \cup \{0\}$.*

The conclusion of Proposition 2.2.3 is known as a *spectral gap*, because it says that there is a gap between the largest and second-largest elements of the spectrum. Note that $0 \in \sigma(L)$ because $L\mathbb{1} = 0$.

Proof. Take $\lambda \in \sigma(L)$, $\lambda \neq 0$. By Lemma 2.2.2, we may choose a sequence $f_n \in \mathcal{D}(L)$ such that $\|f_n\|_2 = 1$ and $\|Lf_n - \lambda f_n\|_2 \to 0$. Since $\int_E Lf_n \, d\mu = 0$ for every n, it follows in particular that $\int_E f_n \, d\mu \to 0$ as $n \to \infty$. Now, the Poincaré inequality yields

$$\int_E f_n^2 \, d\mu - \left(\int_E f_n \, d\mu\right)^2 \leq C \int_E \Gamma(f_n) \, d\mu = -C \int_E f_n L f_n \, d\mu.$$

As $n \to \infty$, the left hand side converges to 1. On the right hand side, we may replace Lf_n in the limit by λf_n; hence, the right hand side converges to $-C\lambda$. We conclude that $\lambda \leq -\frac{1}{C}$. □

2.3 Decay of variance

For a probability measure μ and a function $f \in L^2(\mu)$, define

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

Note that $\mathrm{Var}_\mu(f)$ is the squared distance of f from the constant function with value $\int_E f \, d\mu$. Recall from Proposition 1.10.2 that $P_t f \to \int_E f \, d\mu$ in $L^2(\mu)$ whenever P_t is the semigroup from a compact Markov triple (this was essentially the meaning of ergodicity). A Poincaré inequality can equivalently be understood as a rate of convergence in this limit.

Proposition 2.3.1. *The compact Markov triple (E, μ, Γ) satisfies a Poincaré inequality with constant C if and only if*

$$\operatorname{Var}_\mu(P_t f) \le e^{-2t/C} \operatorname{Var}_\mu(f) \tag{2.4}$$

for every $f \in L^2(\mu)$.

Proof. Suppose first that (E, μ, Γ) satisfies a Poincaré inequality with constant C, and choose some $f \in \mathcal{A}$. Since μ is invariant for P_t, $\int_E P_t f \, d\mu$ is constant in t. On the other hand,

$$\frac{d}{dt} \int_E (P_t f)^2 \, d\mu = 2 \int_E P_t f L P_t f \, d\mu = -2\mathcal{E}(P_t f).$$

Now define $\Lambda(t) = e^{2t/C} \operatorname{Var}_\mu(P_t f)$. Then

$$\Lambda'(t) = \frac{2}{C} \operatorname{Var}_\mu(P_t f) - 2\mathcal{E}(P_t f) \le 0,$$

and so $\Lambda(t) \le \Lambda(0)$ for all t, which proves (2.4). So far we assumed $f \in \mathcal{A}$, but any $f \in L^2(\mu)$ may be reached by approximation.

For the other direction, we consider a Taylor expansion of $\operatorname{Var}(P_t f)$: if $f \in \mathcal{A}$ satisfies $\int_E f \, d\mu = 0$ then $P_t f = f + tLf + o(t)$, and so

$$\operatorname{Var}(P_t f) = \int_E f^2 \, d\mu + 2t \int_E f L f \, d\mu + o(t).$$

On the other hand,

$$e^{-2t/C} \operatorname{Var}_\mu(f) = \left(1 - \frac{2t}{C} + o(t)\right) \operatorname{Var}_\mu(f).$$

Under the assumption (2.4), we may send $t \to 0$ to obtain

$$2 \int_E f L f \, d\mu \le -\frac{2}{C} \operatorname{Var}_\mu(f),$$

which is just a re-arrangement of the Poincaré inequality. □

2.4 Examples

Example 2.4.1 (The Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$)**.** *Recall that the Ornstein-Uhlenbeck semigroup on $\mathbb{R}^n$ may be realized as the semigroup of a* $\mathrm{CD}(1, \infty)$ *compact Markov diffusion with respect to the measure γ_n. In* *particular, Theorem 2.1.1 implies that γ_n satisfies a Poincaré inequality with constant 1. The constant 1 is sharp, as may be seen by considering linear functions: the function $f(x) = x_1$ satisfies*

$$\mathrm{Var}_{\gamma_n}(f) = \int x_1^2 \, d\gamma_n(x) = 1 = \int 1 \, d\gamma_n = \int \Gamma(f) \, d\gamma_n.$$

It is of course no coincidence that the function $f(x) = x_1$ is an eigenfunction of L with eigenvalue -1.

Example 2.4.2 (The ultraspheric semigroup)**.** *Recall that the ultraspheric generator $(L_n f)(x) = (1-x^2)f''(x) - nxf'(x)$ on $(-1,1)$ satisfies the curvature-dimension condition $\mathrm{CD}(n-1,n)$. According to Theorem 2.1.1, it satisfies a Poincaré inequality with constant $\frac{1}{n}$. This constant is sharp, as can be seen from the example $f(x) = x$ which is an eigenfunction of L_n with eigenvalue $-n$.*

Example 2.4.3 (The wrapped heat semigroup)**.** *The wrapped heat semigroup satisfies $\mathrm{CD}(0,1)$; therefore Theorem 2.1.1 tells us nothing about it. However, one can nevertheless prove a Poincaré inequality by using the Fourier expansion as developed in Example 1.10.4. In particular, for any $f \in \mathcal{D}(\mathcal{E})$, we have*

$$\begin{aligned} \mathrm{Var}(f) &= \frac{1}{2} \sum_{k \geq 1} \hat{f}_k^2 + \check{f}_k^2 \\ \mathcal{E}(f) &= 2\pi^2 \sum_{k \geq 1} k^2 (\hat{f}_k^2 + \check{f}_k^2). \end{aligned}$$

In particular, $\mathrm{Var}(f) \leq \frac{1}{4\pi^2} \mathcal{E}(f)$, and so the wrapped heat semigroup satisfies a Poincaré inequality with constant $\frac{1}{4\pi^2}$. The example $f(x) = \cos(2\pi x)$ shows that this constant is sharp.

Note that the inequality above applies only to functions $f \in \mathcal{D}(\mathcal{E})$, which was obtained by starting with smooth functions satisfying $f(0) = f(1)$ and taking the appropriate closure. In particular, it does not hold for all smooth functions $f : [0,1] \to \mathbb{R}$. For example, the function $f(x) = \cos(\pi x)$ satisfies $\mathrm{Var}(f) = \frac{1}{\pi^2} \int_0^1 (f')^2 \, dx$. In order to have an inequality that applies to all smooth enough functions on $[0,1]$, we introduce a simple reflection transformation: for a smooth $f : [0,1] \to \mathbb{R}$, define $g : [0,1] \to \mathbb{R}$ by

$$g(x) = \begin{cases} f(2x) & \text{if } 0 \leq x \leq \frac{1}{2} \\ f(1-2x) & \text{if } \frac{1}{2} < x \leq 1. \end{cases}$$

Then g is smooth except possibly at the point $x = \frac{1}{2}$, where it is continuous. Moreover, $g(0) = g(1)$. Hence, by taking smooth approximations of g, one can show that $g \in \mathcal{D}(\mathcal{E})$. Then

$$\mathrm{Var}(g) \leq \frac{1}{4\pi^2}\mathcal{E}(g).$$

To relate this back to f, note that

$$\begin{aligned}\int_0^1 g\,dx &= \int_0^1 f\,dx\\ \int_0^1 g^2\,dx &= \int_0^1 f^2\,dx\\ \int_0^1 (g')^2\,dx &= 4\int_0^1 (f')^2\,dx.\end{aligned}$$

Hence,

$$\mathrm{Var}(f) \leq \frac{1}{\pi^2}\int_0^1 (f')^2\,dx.$$

This inequality holds for all smooth f (and easily extends to a larger class). The constant $\frac{1}{\pi^2}$ is sharp because of the example $f(x) = \cos(\pi x)$.

2.5 Exponential tails

Suppose that $\Gamma : \mathcal{A} \to \mathcal{A}$ is a diffusion form where $\mathcal{A}$ is dense in $L^2(\mu)$, and consider the Dirichlet form $\mathcal{E}(f) = \int_E \Gamma(f)\,d\mu$. (Note that we are not assuming here most of the properties of compact Markov triples.)

A remarkable property of Poincaré inequalities is that they automatically imply exponential concentration inequalities: if you go "far away" in the space E then the measure μ drops off exponentially. Since we did not begin with any metric structure on E, the correct notion of distance comes from Γ: say that a function $f \in \mathcal{A}$ is ℓ-Lipschitz if $\|\Gamma(f)\|_\infty \leq \ell^2$. Of course, this agrees with the usual notion of Lipschitz if $E = \mathbb{R}^n$ and $\Gamma(f) = |\nabla f|^2$.

Proposition 2.5.1. *If $(\mu, \mathcal{E})$ satisfy a Poincaré inequality with constant C then for every 1-Lipschitz $f \in \mathcal{A}$ and every $0 < s < \sqrt{\frac{4}{C}}$,*

$$\int_E e^{sf}\,d\mu \leq \frac{2 + s\sqrt{C}}{2 - s\sqrt{C}} e^{s\int_E f\,d\mu}.$$

To develop some intuition about the exponential integrability condition in Proposition 2.5.1, consider the case $E = \mathbb{R}$, $\Gamma(f,g) = f'g'$. Let μ be a mean-zero probability measure that satisfies a Poincaré inequality with constant C. Then the function $f(x) = x$ is 1-Lipschitz and has mean zero; by Proposition 2.5.1 with $s = C^{-1/2}$,

$$\int_{-\infty}^{\infty} e^{x/\sqrt{C}}\, d\mu(x) \le 3.$$

By Markov's inequality, it follows that $\mu[t,\infty) \le 3e^{-t/\sqrt{C}}$ for every $t \ge 0$. This is what we mean when we say that μ has exponential tails. For a general E and Γ, the same argument implies that

$$\mu\left\{x : f(x) \ge \int_E f\, d\mu + t\right\} \le 3e^{-t/\sqrt{C}}$$

for every 1-Lipschitz function f.

Proof. It is enough to prove the claim for bounded $f \in \mathcal{A}$; the case of arbitrary $f \in \mathcal{A}$ may then be reached by approximation and Fatou's lemma. Now assume that $f \in \mathcal{A}$ is bounded, and define $Z(s) = \int_E e^{sf}\, d\mu$. Note that

$$\mathrm{Var}_\mu(e^{sf/2}) = Z(s) - Z(s/2)^2$$

and that

$$\mathcal{E}(e^{sf/2}) = \int_E \Gamma(e^{sf/2})\, d\mu = \frac{s^2}{4}\int_E e^{sf}\Gamma(f)\, d\mu \le \frac{s^2}{4} Z(s).$$

Under the assumption that μ and $\mathcal{E}$ satisfy a Poincaré inequality with constant C, we therefore have

$$\left(1 - \frac{Cs^2}{4}\right) Z(s) \le Z(s/2)^2.$$

By induction,

$$Z(s) \le Z(s2^{-n})^{2^n} \prod_{\ell=0}^{n-1} \left(1 - \frac{Cs^2}{4^{\ell+1}}\right)^{-2^\ell} \tag{2.5}$$

for every $n \ge 1$. Now, $Z(s) = 1 + s\int_E f\, d\mu + O(s^2)$ as $s \to 0$; it follows that $Z(s2^{-n})^{2^n} \to e^{s\int_E f\, d\mu}$ as $n \to \infty$. Taking $n \to \infty$ in (2.5), we have

$$\int_E e^{sf}\, d\mu \le e^{s\int_E f\, d\mu} \prod_{\ell=0}^{\infty} \left(1 - \frac{Cs^2}{4^{\ell+1}}\right)^{-2^\ell}.$$

Finally, we claim that

$$\prod_{\ell=0}^{\infty}\left(1-\frac{x^2}{4^\ell}\right)^{-2^\ell} \leq \frac{1+x}{1-x} \tag{2.6}$$

for all $x \in [0,1)$; applying this inequality with $x = Cs^2/4$ will then complete the proof.

To check (2.6), take logarithms on both sides: on the left hand side,

$$\begin{aligned}\log\prod_{\ell=0}^{\infty}\left(1-\frac{x^2}{4^\ell}\right)^{-2^\ell} &= -\sum_{\ell=0}^{\infty}2^\ell\log\left(1-\frac{x^2}{4^\ell}\right)\\ &= \sum_{\ell=0}^{\infty}2^\ell\sum_{k=1}^{\infty}\frac{x^{2k}}{k4^{k\ell}}\\ &= \sum_{k=1}^{\infty}\frac{x^{2k}}{k(1-2^{1-2k})}.\end{aligned}$$

On the other hand, if $0 \leq x < 1$ then

$$\log\frac{1+x}{1-x} = 2\sum_{k=1}^{\infty}\frac{x^{2k-1}}{2k-1} \geq 2\sum_{k=1}^{\infty}\frac{x^{2k}}{2k-1}.$$

Then (2.6) follows from comparing these expansions term-by-term and noting that $\frac{1}{k(1-2^{1-2k})} \leq \frac{2}{2k-1}$ for all $k \geq 1$. $\square$

2.6 The exponential distribution

One might wonder whether the exponential tail behavior of the previous section is sharp. For example, the Gaussian measure satisfies a Poincaré inequality and has tails that decay at the faster rate $e^{-x^2/2}$. To see that an exponential tail decay is the best possible that can be implied by a Poincaré inequality, we consider the exponential distribution.

Proposition 2.6.1. *Let μ be the distribution $d\mu(x) = e^{-x}\,dx$ on $[0,\infty)$. Then*

$$\mathrm{Var}_\mu(f) \leq 4\int_0^\infty (f')^2\,d\mu$$

for every $f : \mathbb{R}_+ \to \mathbb{R}$ with a bounded derivative. In other words, μ satisfies a Poincaré inequality with constant 4 with respect to the usual carré du champ operator.

Note that the constant 4 in Proposition 2.6.1 is sharp: indeed, if μ were to satisfy a Poincaré inequality with constant $C < 4$, it would follow from Proposition 2.5.1 that e^x were integrable with respect to μ (which is clearly not the case). On the other hand, the exponential distribution shows that the range $0 < s < \sqrt{4/C}$ for exponential integrability in Proposition 2.5.1 is sharp.

Proof. Since f' and $\mathrm{Var}_\mu(f)$ are unchanged when we add constant functions to f, we may assume that $f(0) = 0$. Then

$$\begin{aligned}
\int_0^\infty f^2(x)\, d\mu(x) &= \int_0^\infty \left(2\int_0^x f(y)f'(y)\, dy\right) e^{-x}\, dx \\
&= 2\int_0^\infty f(y)f'(y) \left(\int_y^\infty e^{-x}\, dx\right) dy \\
&= 2\int_0^\infty f f'\, d\mu \\
&\le 2\sqrt{\int_0^\infty f^2\, d\mu \int_0^\infty (f')^2\, d\mu}.
\end{aligned}$$

Hence, $\int_0^\infty f^2\, d\mu \le 4\int_0^\infty (f')^2\, d\mu$. Since $\mathrm{Var}_\mu(f) \le \int_0^\infty f^2\, d\mu$, this proves the claim. $\square$

2.7 Log-concave measures

Definition 2.7.1. *The probability measure μ on $\mathbb{R}^n$ defined by $d\mu(x) = e^{-W(x)}\, dx$ is* log-concave *if W is convex. For $\rho > 0$, μ is ρ*-strongly log-concave *if $W(x) - \rho|x|^2$ is convex.*

From now on, assume that W is a smooth function, and let $\Gamma(f, g) = \langle \nabla f, \nabla g\rangle$ be the usual carré du champ on $\mathbb{R}^n$. Integrating by parts, we have

$$-\int_{\mathbb{R}^n} \Gamma(f, g)\, d\mu = -\int_{\mathbb{R}^n} \langle \nabla f, e^{-W}\nabla g\rangle\, dx = \int_{\mathbb{R}^n} f(\Delta g - \langle \nabla W, \nabla g\rangle)\, d\mu.$$

In particular, setting $Lg = \Delta g - \langle \nabla W, \nabla g\rangle$ gives us the integration by parts formula $\int f Lg\, d\mu = -\int \Gamma(f, g)\, d\mu$. If all derivatives of $W(x)$ grow at most polynomially fast as $|x| \to \infty$ then $(\mathbb{R}^n, \mu, \Gamma)$ is a compact Markov triple when $\mathcal{A}$ is the class of smooth, bounded functions whose derivatives all vanish super-polynomially fast. (We leave it as an exercise to check that this defines a compact Markov triple; the main difficulty is to show that $P_t\mathcal{A}_0 \subset \mathcal{A}_0$, for which one may use regularity estimates for parabolic PDE.) Of course, if $W(x) = \frac{1}{2}|x|^2 - \frac{n}{2}\log(2\pi)$ then we recover the Ornstein-Uhlenbeck triple.

By adapting the computation in Example 1.11.4 (replacing x by $\nabla W(x)$), one computes

$$\Gamma_2(f, g) = \langle \nabla^2 f, \nabla^2 g \rangle + (\nabla^2 W)(\nabla f, \nabla g),$$

where for a $n \times n$ matrix A, $A(v, w)$ means the bilinear form $\sum_{i,j} A_{ij} v_i w_j$. In particular, the triple $(\mathbb{R}^n, \mu, \Gamma)$ satisfies $\mathrm{CD}(\rho, n)$ if and only if μ is ρ-strongly log-concave and $n = \infty$.

Corollary 2.7.2. *Every ρ-strongly log-concave probability measure satisfies a Poincaré inequality (with respect to the usual carré du champ) with constant $\frac{1}{\rho}$.*

Note that Corollary 2.7.2 says nothing about measure that are log-concave but not strongly so. A theorem due to Bobkov states that every log-concave measure satisfies a Poincaré inequality, but does not give a sharp bound on the Poincaré constant. This is the content of a famous conjecture of Kannan, Lovász, and Simonovits:

Conjecture 2.7.3. *There is a universal constant C such that if μ is a log-concave probability measure on $\mathbb{R}^n$ satisfying $\int_{\mathbb{R}^n} x \, d\mu(x) = 0$ and*

$$\int_{\mathbb{R}^n} x x^T \, d\mu(x) = \mathrm{Id},$$

then μ satisfies a Poincaré inequality (with respect to the usual carré du champ) with constant C.

Note that the two conditions $\int_{\mathbb{R}^n} x \, d\mu(x) = \mathrm{Id}$ and $\int_{\mathbb{R}^n} x x^T \, d\mu(x) = \mathrm{Id}$ can always be ensured by applying an affine transformation to μ.

Chapter 3

Logarithmic Sobolev inequalities

3.1 The strong gradient bound

When we proved the Poincaré inequality under curvature conditions, the "gradient bound" $\Gamma(P_t f) \leq e^{-2t} P_t \Gamma(f)$ played a crucial role. To understand this bound better, let us consider the Ornstein-Uhlenbeck semigroup, whose carré du champ is $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$ and whose semigroup has the representation

$$(P_t f)(x) = \int_{\mathbb{R}^n} f(e^{-t} x + \sqrt{1 - e^{-2t}} y) \, d\gamma_n(y).$$

The gradient bound yields $|\nabla P_t f|^2 \leq e^{-2t} P_t |\nabla f|^2$. This is sharp when f is a linear function, because if $f(x) = \langle a, x \rangle + b$ then $(P_t f)(x) = e^{-t} \langle a, x \rangle + b$. But in this specific case, we can do better. Passing the gradient under the integral,

$$(\nabla P_t f)(x) = \int_{\mathbb{R}^n} e^{-t} (\nabla f)(e^{-t} x + \sqrt{1 - e^{-2t}} y) \, d\gamma_n(y) = e^{-t} (P_t \nabla f)(x).$$

Jensen's inequality gives $|\nabla P_t f| \leq e^{-t} P_t |\nabla f|$. In terms of Γ,

$$\sqrt{\Gamma(P_t f)} \leq e^{-t} P_t \sqrt{\Gamma(f)}.$$

This is better than our original gradient bound, which follows from this one because $x \mapsto x^2$ is convex.

Remarkably, the improved gradient bound holds not just for the Ornstein-Uhlenbeck semigroup, but under a $\mathrm{CD}(\rho, \infty)$ condition. This should be at least a little bit surprising: the $\mathrm{CD}(\rho, \infty)$ condition is equivalent to the weaker gradient bound, so how can it imply a strictly stronger bound? The key is in the diffusion assumption $\Gamma(\Psi(f), g) = \sum \partial_i \Psi \Gamma(f, g)$, which played no role in the weaker gradient bound but is crucial for the stronger one.

Theorem 3.1.1 (Strong gradient bound)**.** *Let* (E, μ, Γ) *be a compact Markov triple that satisfies* $\mathrm{CD}(\rho, \infty)$*. Then for every* $f \in \mathcal{A}$ *and every* $t \geq 0$*,*

$$\sqrt{\Gamma(P_t f)} \leq e^{-\rho t} P_t \sqrt{\Gamma(f)}.$$

We begin with a non-rigorous argument that will help motivate the real proof. Fix t and define $\Lambda(s) = P_s \sqrt{\Gamma(P_{t-s} f)}$. If the square root were a smooth function, the chain rule for L would give $L\sqrt{g} = \frac{Lg}{2\sqrt{g}} - \frac{\Gamma(g)}{4g^{3/2}}$ for any $g \in \mathcal{A}$. Then we could compute

$$\begin{aligned}
\Lambda'(s) &= P_s L \sqrt{\Gamma(P_{t-s} f)} + P_s \frac{\frac{d}{ds}\Gamma(P_{t-s} f)}{2\sqrt{\Gamma(P_{t-s} f)}} \\
&= P_s \left[\frac{L\Gamma(P_{t-s} f)}{2\sqrt{\Gamma(P_{t-s} f)}} - \frac{\Gamma(\Gamma(P_{t-s} f))}{4\Gamma(P_{t-s} f)^{3/2}} - \frac{\Gamma(P_{t-s} f, LP_{t-s} f)}{\sqrt{\Gamma(P_{t-s} f)}} \right] \\
&= P_s \left[\frac{\Gamma_2(P_{t-s} f)}{\sqrt{\Gamma(P_{t-s} f)}} - \frac{\Gamma(\Gamma(P_{t-s} f))}{4\Gamma(P_{t-s} f)^{3/2}} \right].
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{d}{ds}(e^{-\rho s}\Lambda(s)) &= e^{-\rho s}(\Lambda'(s) - \rho\Lambda(s)) \\
&= e^{-\rho s} P_s \left[\frac{\Gamma_2(P_{t-s} f)\Gamma(P_{t-s} f) - \rho\Gamma(P_{t-s} f)^2 - \frac{1}{4}\Gamma(\Gamma(P_{t-s} f))}{4\Gamma(P_{t-s} f)^{3/2}} \right].
\end{aligned}$$

Noting that the denominator is negative, the inequality $\frac{d}{ds}(e^{-\rho s}\Lambda(s)) \geq 0$ is equivalent to the inequality

$$\Gamma(g)(\Gamma_2(g) - \rho\Gamma(g)) \geq \frac{1}{4}\Gamma(\Gamma(g)), \tag{3.1}$$

for $g = P_{t-s} f$. Since $\Gamma(g)$ and $\Gamma(\Gamma(g))$ are both non-negative, (3.1) is stronger than the $\mathrm{CD}(\rho, \infty)$ condition.

In order to better understand (3.1), consider again the Ornstein-Uhlenbeck example. There, $\rho = 1$ and $\Gamma_2(g) - \Gamma(g) = \|\nabla^2 f\|_2^2$. To compute $\Gamma(\Gamma(g))$, note that

$$\partial_i |\nabla g|^2 = \sum_j 2\partial_j g \partial_i \partial_j g = 2(\nabla^2 g \cdot \nabla g)_i.$$

Hence, $|\nabla|\nabla g|^2|^2 = 4|\nabla^2 g \cdot \nabla g|^2$ and so (3.1) becomes $|v|^2\|H\|_2^2 \geq |Hv|^2$, where $H = \nabla^2 g$ and $v = \nabla g$. Since the Hilbert-Schmidt norm is larger than the operator norm, this inequality is indeed true.

To prove the inequality (3.1) under the $\mathrm{CD}(\rho, \infty)$ condition, one first needs to develop a chain rule for Γ_2. That is, by expanding the definition of Γ_2 and applying the chain rules

$$\Gamma(\Psi(f), g) = \sum_i (\partial_i \Psi)(f)\Gamma(f_i, g)$$
$$L\Psi(f) = (\sum_i \partial_i \Psi)(f) L f_i + \sum_{i,j} (\partial_i \partial_j \Psi)(f)\Gamma(f_i, f_j),$$

one can derive a similar chain rule for Γ_2. Unfortunately, the computation is extremely tedious and (in our opinion) not particularly edifying. Therefore we omit it, and only state the result (and that only in the case of the quadratic form, which is the only one we need). Let $X_i = (\partial_i \Psi)(f)$ and $Y_{ij} = (\partial_i \partial_j \Psi)(f)$. Then for any smooth $\Psi : \mathbb{R}^k \to \mathbb{R}$ and any $f = (f_1, \ldots, f_k) \in \mathcal{A}^k$,

$$\begin{aligned}\Gamma_2(\Psi(f)) = \sum_{i,j} & X_i X_j \Gamma_2(f_i, f_j) \\ & + \sum_{i,j,\ell} X_i Y_{j\ell}\big(\Gamma(f_j, \Gamma(f_i, f_\ell)) + \Gamma(f_\ell, \Gamma(f_i, f_j)) - \Gamma(f_i, \Gamma(f_j, f_\ell))\big) \\ & + \sum_{i,j,\ell,m} Y_{ij} Y_{\ell m} \Gamma(f_i, f_\ell)\Gamma(f_j, f_m). \end{aligned} \tag{3.2}$$

The $\mathrm{CD}(\rho, \infty)$ condition becomes

$$\begin{aligned} 0 \leq & \sum_{i,j} X_i X_j \Gamma_2(f_i, f_j) \\ & + \sum_{i,j,\ell} X_i Y_{j\ell}\big(\Gamma(f_j, \Gamma(f_i, f_\ell)) + \Gamma(f_\ell, \Gamma(f_i, f_j)) - \Gamma(f_i, \Gamma(f_j, f_\ell))\big) \\ & + \sum_{i,j,\ell,m} Y_{ij} Y_{\ell m} \Gamma(f_i, f_\ell)\Gamma(f_j, f_m) \\ & - \rho \sum_{i,j} X_i X_j \Gamma(f_i, f_j). \end{aligned}$$

If we fix a single point, then as Ψ ranges over all smooth functions, X_i and Y_{ij} range over all combinations of real numbers (subject to the symmetry condition $Y_{ij} = Y_{ji}$). In particular, the right hand side above may be viewed as a quadratic form in the variables X_i and Y_{ij}.

Consider the case $k = 3$, $X_1 = x$, $Y_{23} = Y_{32} = y$, and all the other variables are zero. Moreover, specialize to the functions $f = (g, g, h)$. Then

$$x^2(\Gamma_2(g) - \rho\Gamma(g)) + 2xy\Gamma(h, \Gamma(g)) + 2y^2(\Gamma(g)\Gamma(h) + \Gamma(g, h)^2)$$

is a non-negative quadratic form in x and y. In particular, its determinant must be non-negative:

$$\begin{aligned}\Gamma(h, \Gamma(g))^2 &\leq 2(\Gamma_2(g) - \rho\Gamma(g))(\Gamma(g)\Gamma(h) + \Gamma(g, h)^2) \\ &\leq 4(\Gamma_2(g) - \rho\Gamma(g))\Gamma(g)\Gamma(h),\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality for Γ. Setting $h = \Gamma(g)$ gives (3.1). By proving (3.1), we have also complete the not-quite-proof of Theorem 3.1.1, in which we assumed that the square root function was smooth. The real proof adds an approximation to the square root function.

Proof of Theorem 3.1.1. Fix $\epsilon > 0$ and let $\psi(x)$ be a smooth function satisfying $\psi(x) = \sqrt{x + \epsilon}$ for some open interval containing $\mathbb{R}_+$. Note that

$$\begin{aligned}\psi'(x) &= \frac{1}{2\psi(x)} \\ \psi''(x) &= -\frac{1}{4\psi^3(x)}\end{aligned}$$

for all $x \geq 0$. Fix t and define $\Lambda(s) = P_s\psi(e^{-2\rho s}\Gamma(P_{t-s}f))$. Similarly to our previous computation with the square root function,

$$\frac{d}{ds}\Lambda(s) = e^{-2\rho s}P_s\left[\frac{\Gamma_2 - \rho\Gamma}{\psi(e^{-2\rho s}\Gamma)} - \frac{e^{-2\rho s}\Gamma(\Gamma)}{4\psi^3(e^{-2\rho s}\Gamma)}\right],$$

where we have omitted all the occurrences of "$P_{t-s}f$" to save space. (Note that we were able to use our formulas for ψ' and ψ'' above, because we are always evaluating ψ at something non-negative.) For the second term, we apply the bound $\psi^2(x) \geq x$ in the denominator, obtaining

$$\frac{d}{ds}\Lambda(s) \geq e^{-2\rho s}P_s\left[\frac{\Gamma_2 - \rho\Gamma}{\psi(e^{-2\rho s}\Gamma)} - \frac{\Gamma(\Gamma)}{4\Gamma\psi(e^{-2\rho s}\Gamma)}\right].$$

According to (3.1), the right hand side above is non-negative. Hence, $\Lambda(s)$ is non-decreasing in s, and so

$$P_t\sqrt{e^{-2\rho t}\Gamma(f) + \epsilon} = \Lambda(t) \geq \Lambda(0) = \sqrt{\Gamma(P_t f) + \epsilon}.$$

Taking $\epsilon \to 0$ completes the proof. $\square$

3.2 The logarithmic Sobolev inequality

For a probability measure μ on E and a non-negative, measurable function $f : E \to \mathbb{R}_+$, define the *entropy* by

$$\mathrm{Ent}_\mu(f) = \int_E f \log f \, d\mu - \int_E f \, d\mu \log \int_E f \, d\mu,$$

where we adopt the convention that $0 \log 0 = 0$. By Jensen's inequality applied to the convex function $\psi(x) = x \log x$, $\mathrm{Ent}_\mu(f)$ is always non-negative. Since $\psi(x)$ is strictly convex, $\mathrm{Ent}_\mu(f) = 0$ if and only if f is constant μ-a.s. Note that the entropy is homogeneous in the sense that $\mathrm{Ent}_\mu(cf) = c \, \mathrm{Ent}_\mu(f)$ for all $c > 0$.

Definition 3.2.1. *If μ is a probability measure and $\mathcal{E}$ is a Dirichlet form, we say they satisfy a* log-Sobolev inequality with constant C *if for all $f \in \mathcal{D}(\mathcal{E})$,*

$$\mathrm{Ent}_\mu(f^2) \leq 2C\mathcal{E}(f). \tag{3.3}$$

The smallest C (which may be infinite) for which μ and $\mathcal{E}$ satisfy a log-Sobolev inequality is called the log-Sobolev constant *of μ and $\mathcal{E}$.*

Of course, it will always suffice to prove a log-Sobolev inequality for a large enough family (usually $\mathcal{A}$) of functions belonging to $\mathcal{D}(\mathcal{E})$.

The log-Sobolev inequality is often encountered in an equivalent form: if f is non-negative and bounded away from zero (i.e. there exists some $\epsilon > 0$ such that $f \geq \epsilon$) then we may apply (3.3) to $\sqrt{f}$, yielding

$$\mathrm{Ent}_\mu(f) \leq 2C\mathcal{E}(\sqrt{f}) = \frac{C}{2} \int_E \frac{\Gamma(f)}{f}, d\mu,$$

where the last equality follows in the case that $\mathcal{E}(g) = \int_E \Gamma(f) \, d\mu$ for a diffusion carré du champ Γ. Once the inequality is established for $f \in \mathcal{A}$ that are bounded away from zero, it follows immediately for all positive $f \in \mathcal{A}$ by taking limits. (Note that for an arbitrary positive $f \in \mathcal{A}$, the integral of $\frac{\Gamma(f)}{f}$ may be infinity, but this does not contradict the inequality.)

Definition 3.2.2. *If ν is a probability measure that has density f with respect to μ, define the* Fisher information *of ν with respect to μ by*

$$\mathrm{I}(\nu|\mu) = \mathrm{I}_\mu(f) = \int_E \frac{\Gamma(f)}{f} \, d\mu$$

and the entropy *of ν with respect to μ by*

$$H(\nu|\mu) = \mathrm{Ent}_\mu(f).$$

With this terminology, μ and Γ satisfy a log-Sobolev inequality with constant C if and only if

$$H(\nu|\mu) \leq \frac{C}{2} I(\nu|\mu)$$

for every probability measure ν that is absolutely continuous with respect to μ. Here, we are allowing both sides to be (positive) infinity; for example, we declare the left hand side to be infinity if $\frac{d\nu}{d\mu}$ does not belong to $\mathcal{D}(\mathcal{E})$. Note that the assumption for ν to be a probability measure is not important, since the original form (3.3) is homogeneous.

Proposition 3.2.3. *If μ and $\mathcal{E}$ satisfy a log-Sobolev inequality with constant C then they satisfy a Poincaré inequality with constant C.*

Proof. Given $f \in \mathcal{A}$ with mean zero and $\epsilon > 0$, apply the log-Sobolev inequality (3.3) to $1+\epsilon f$. On the right hand side, $2C\mathcal{E}(1+\epsilon f) = 2C\epsilon^2\mathcal{E}(f)$. On the left hand side, since $\log(1+x) = x - \frac{1}{2}x^2 + o(x^2)$, we may write the entropy of $(1+\epsilon f)^2$ as

$$\int_E (1+\epsilon f)^2 (\epsilon f - \frac{\epsilon^2}{2} f^2)\, d\mu - \int_E (1+\epsilon f)^2\, d\mu \log \int_E (1+\epsilon f)^2\, d\mu + o(\epsilon^2).$$

Recalling that $\int_E f\, d\mu = 0$, this reduces to

$$\mathrm{Ent}_\mu((1+\epsilon f)^2) = 2\epsilon^2 \int_E f^2\, d\mu + o(\epsilon^2).$$

Taking $\epsilon \to 0$, (3.3) implies that

$$\int_E f^2\, d\mu \leq C\mathcal{E}(f),$$

which is exactly the Poincaré inequality with constant C. □

3.3 Log-Sobolev inequalities under $\mathrm{CD}(\rho, n)$

The story for log-Sobolev inequalities under a $\mathrm{CD}(\rho, n)$ condition is quite similar to the story for Poincaré inequalities: there are local log-Sobolev inequalities that are equivalent to $\mathrm{CD}(\rho, \infty)$, and there is an improved (but non-local) log-Sobolev inequality for $\mathrm{CD}(\rho, n)$. We begin with the local log-Sobolev inequalities, which should be compared with Proposition 2.0.1, its Poincaré analogue.

Proposition 3.3.1. *For a compact Markov triple (E, μ, Γ) with semigroup P_t, the following are equivalent:*

(a) the $\mathrm{CD}(\rho, \infty)$ *condition holds for some* $\rho \in \mathbb{R}$*;*

(b) for every $f \in \mathcal{A}$*,*

$$\Gamma(f)(\Gamma_2(f) - \rho\Gamma(f)) \geq \frac{1}{4}\Gamma(\Gamma(f));$$

(c) for every $f \in \mathcal{A}$ *and every* $t \geq 0$*,*

$$\sqrt{\Gamma(P_t f)} \leq e^{-\rho t} P_t \sqrt{\Gamma(f)};$$

(d) for every positive $f \in \mathcal{A}$ *and every* $t \geq 0$*,*

$$P_t(f \log f) - P_t f \log P_t f \leq \frac{1 - e^{-2\rho t}}{2\rho} P_t \frac{\Gamma(f)}{f};$$

(e) for every positive $f \in \mathcal{A}$ *and every* $t \geq 0$*,*

$$P_t(f \log f) - P_t f \log P_t f \geq \frac{e^{2\rho t} - 1}{2\rho} \frac{\Gamma(P_t f)}{P_t f}.$$

We call part (d) of Proposition 3.3.1 a *local log-Sobolev* inequality, since it says that each kernel measure $p_t(x, \cdot)$ satisfies a log-Sobolev inequality with constant $\frac{1-e^{-2\rho t}}{\rho}$. Similarly, we call part (e) a *local reverse log-Sobolev* inequality. Note that by taking $t \to \infty$ and using ergodicity (Proposition 1.10.2), we obtain the following corollary:

Corollary 3.3.2. *If a compact Markov triple* (E, μ, Γ) *satisfies the* $\mathrm{CD}(\rho, \infty)$ *condition for some* $\rho > 0$ *then* μ *and* Γ *satisfy a log-Sobolev inequality with constant* $\frac{1}{\rho}$.

Proof of Proposition 3.3.1. In proving Theorem 3.1.1, we have already proved that (a) implies (b), which in turn implies (c). On the other hand, (c) and Jensen's inequality together imply that

$$\Gamma(P_t f) \leq e^{-2\rho t} P_t \Gamma(f)$$

for all $f \in \mathcal{A}$ and $t \geq 0$. According to Proposition 2.0.1, this implies the $\mathrm{CD}(\rho, \infty)$ condition (i.e., (a)).

To go from (c) to (d), fix $t > 0$ and define

$$\Lambda(s) = P_s[P_{t-s}f \log P_{t-s}f].$$

Letting $\psi(x) = x \log x$, one computes

$$\begin{aligned}\Lambda'(s) &= \frac{d}{ds} P_s \psi(P_{t-s}f) \\ &= P_s[L\psi(P_{t-s}f) - \psi'(P_{t-s}f) L P_{t-s}f] \\ &= P_s[\psi''(P_{t-s}f)\Gamma(P_{t-s}f)].\end{aligned}$$

(Of course, this computation is valid for a general smooth function ψ.) For our function ψ, $\psi''(x) = \frac{1}{x}$. Hence,

$$\Lambda'(s) = P_s \frac{\Gamma(P_{t-s}f)}{P_{t-s}f} \le e^{-2\rho(t-s)} P_s \frac{(P_{t-s}\sqrt{\Gamma(f)})^2}{P_{t-s}f}, \tag{3.4}$$

where the inequality follows from part (c) of the proposition. Now, if X and Y are positive random variables then the Cauchy-Schwarz inequality may be written as

$$\mathbb{E}X = \mathbb{E}\left[\sqrt{Y}\frac{X}{\sqrt{Y}}\right] \le \sqrt{\mathbb{E}Y\mathbb{E}\frac{X^2}{Y}}$$

Or, rearranged, $\frac{(\mathbb{E}X)^2}{\mathbb{E}Y} \le \mathbb{E}\frac{X^2}{Y}$. We may apply this to (3.4), where $X = \sqrt{\Gamma(f)}$, $Y = f$, and the expectations are taken with respect to the kernel measures $p_{t-s}(x, \cdot)$. This yields

$$\Lambda'(s) \le e^{-2\rho(t-s)} P_s P_{t-s} \frac{\Gamma(f)}{f} = e^{-2\rho(t-s)} P_t \frac{\Gamma(f)}{f}.$$

Integrating out s, we have

$$\Lambda(t) - \Lambda(0) \le P_t \frac{\Gamma(f)}{f} \int_0^t e^{-2\rho(t-s)}\, ds = \frac{1 - e^{-2\rho t}}{2\rho} P_t \frac{\Gamma(f)}{f},$$

thus proving (d).

The proof that (c) implies (d) is very similar, except that instead of applying the inequality in (3.4), we first apply the Cauchy-Schwarz inequality with $X = \sqrt{\Gamma(P_{t-s}f)}$, $Y = P_{t-s}f$, and with expectations taken according to the measures $p_s(x, \cdot)$. This gives

$$\Lambda'(s) \ge \frac{(P_s\sqrt{\Gamma(P_{t-s}f)})^2}{P_t f} \ge e^{2\rho s} \frac{\Gamma(P_t f)}{P_t f}.$$

Integrating out s implies (e).

To prove that (d) implies (a), recall that the log-Sobolev inequality with constant C implies the Poincaré inequality with the same constant. In particular, the local log-Sobolev inequality in (d) implies the local Poincaré inequality in Proposition 2.0.1, part (c). By Proposition 2.0.1, this implies $\mathrm{CD}(\rho, \infty)$.

Finally, the proof that (e) implies (a) is analogous: by applying the local reverse log-Sobolev inequality to a function of the form $(1+\epsilon f)^2$, one deduces the local reverse Poincaré inequality in Proposition 2.0.1, part (d), which in turn implies $\mathrm{CD}(\rho, \infty)$. □

3.3.1 The finite-dimensional improvement

We saw that a log-Sobolev inequality holds under a $\mathrm{CD}(\rho, \infty)$ condition. As in the case of the Poincaré inequality, the log-Sobolev constant may be sharpened if the dimension is finite:

Theorem 3.3.3. *If a compact Markov triple (E, μ, Γ) satisfies the $\mathrm{CD}(\rho, n)$ condition for $\rho > 0$ then μ and Γ satisfy a log-Sobolev inequality with constant $\frac{n-1}{n\rho}$.*

Lemma 3.3.4. *Suppose that*

$$\int_E f\Gamma(\log f)\, d\mu \le C \int_E f\Gamma_2(\log f)\, d\mu$$

for some $C > 0$ and all positive $f \in \mathcal{A}$. Then μ and Γ satisfy a log-Sobolev inequality with constant C.

Proof. Fix some positive $f \in \mathcal{A}$. Let $\Lambda(t) = \int_E P_t f \log P_t f\, d\mu$. As in the proof of Proposition 3.3.1,

$$\Lambda'(t) = -\int_E \frac{\Gamma(P_t f)}{P_t f}\, d\mu = -\int_E P_t f \Gamma(\log P_t f)\, d\mu.$$

For the second derivative,

$$\Lambda''(t) = \int_E \frac{\Gamma(P_t f) L P_t f}{(P_t f)^2} - \frac{2\Gamma(P_t f, L P_t f)}{P_t f}\, d\mu. \tag{3.5}$$

From now on, take $g = P_t f$. The change of variables formula for L gives

$$L\frac{\Gamma(g)}{g} = -\frac{\Gamma(g) Lg}{g^2} + \frac{L\Gamma(g)}{g} - \frac{2\Gamma(g, \Gamma(g))}{g^2} + \frac{2\Gamma(g)^2}{g^3}.$$

Since $\int_E Lh\, d\mu = 0$ for any $h \in \mathcal{D}(L)$,

$$\int_E \frac{\Gamma(g)Lg}{g^2} = \int_E \frac{L\Gamma(g)}{g} - 2\frac{\Gamma(g,\Gamma(g))}{g^2} + 2\frac{\Gamma(g)^2}{g^3}\, d\mu.$$

Applying this to (3.5) and recognizing that $L\Gamma(g) - 2\Gamma(g, Lg) = 2\Gamma_2(g)$, we have

$$\Lambda''(t) = 2\int_E \frac{\Gamma_2(g)}{g} - \frac{\Gamma(g,\Gamma(g))}{g^2} + \frac{\Gamma(g)^2}{g^3}\, d\mu.$$

On the other hand, the chain rule (3.2) for Γ_2 reveals that the integrand above is just $g\Gamma_2(\log g)$.

To recap,

$$\begin{aligned}
\Lambda'(t) &= -\int_E P_t f \Gamma(\log P_t f)\, d\mu \\
\Lambda''(t) &= 2\int_E P_t f \Gamma_2(\log P_t f)\, d\mu.
\end{aligned}$$

Under the assumption of the lemma, $-\Lambda'(t) \leq \frac{C}{2}\Lambda''(t)$, and hence $\Lambda'(t) \geq \Lambda'(0)e^{-2t/C}$. In other words,

$$\int_E \frac{\Gamma(P_t f)}{P_t f}\, d\mu \leq e^{-2t/C} \int_E \frac{\Gamma(f)}{f}\, d\mu.$$

Finally,

$$\Lambda(0) - \Lambda(t) = -\int_0^t \Lambda'(s)\, ds \leq -\Lambda'(0)\int_0^\infty e^{-2s/C}\, ds = \frac{C(1-e^{-2t/C})}{2}\, \mathrm{I}_\mu(f).$$

Taking $t \to \infty$ completes the proof. $\square$

Proof of Theorem 3.3.3. Our main task is to verify the condition of Lemma 3.3.4. We begin by recalling the chain rule for Γ_2:

$$\Gamma_2(\psi(g)) = (\psi')^2(g)\Gamma_2(g) + \psi'(g)\psi''(g)\Gamma(g,\Gamma(g)) + (\psi'')^2(g)\Gamma(g)^2.$$

With $\psi(g) = e^{ag}$ (for a constant a to be determined), we have

$$\begin{aligned}
\Gamma_2(e^{ag}) &= a^2 e^{2ag}[\Gamma_2(g) + a\Gamma(g,\Gamma(g)) + a^2\Gamma(g)^2] \qquad (3.6)\\
\Gamma(e^{ag}) &= a^2 e^{2ag}\Gamma(g) \\
Le^{ag} &= ae^{ag}[Lg + a\Gamma(g)].
\end{aligned}$$

Hence, the CD(ρ, n) condition may be written as

$$\Gamma_2(g) + a\Gamma(g, \Gamma(g)) + a^2\Gamma(g)^2 \geq \rho\Gamma(g) + \frac{1}{n}[Lg + a\Gamma(g)]^2.$$

Multiplying by e^g and integrating,

$$\int_E e^g \left[\Gamma_2(g) + a\Gamma(g, \Gamma(g)) + a^2\Gamma(g)^2 - \rho\Gamma(g) - \frac{1}{n}[Lg + a\Gamma(g)]^2\right] d\mu \geq 0. \tag{3.7}$$

By various manipulations, we will turn this into an inequality involving only Γ_2 and Γ. To that end, note that $4(Le^{g/2})^2 = e^g[Lg + \frac{1}{2}\Gamma(g)]^2$. On the other hand, $\int_E (Lf)^2\, d\mu = \int_E \Gamma_2(f)\, d\mu$ for any f. Hence the term involving L in (3.7) may be expanded as follows:

$$\begin{aligned}
&\int_E [Lg + a\Gamma(g)]^2\, d\mu \\
&= \int_E 4(Le^{g/2})^2 + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right] d\mu \\
&= \int_E 4\Gamma_2(e^{g/2}) + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right] d\mu \\
&= \int_E 4\Gamma_2(e^{g/2}) + e^g\left[(2a-1)\Gamma(g)Lg + \frac{4a^2-1}{4}\Gamma(g)^2\right] d\mu \\
&= \int_E e^g\left[\Gamma_2(g) + \frac{1}{2}\Gamma(g, \Gamma(g)) + (2a-1)\Gamma(g)Lg + a^2\Gamma(g)^2\right] d\mu,
\end{aligned} \tag{3.8}$$

where the last line follows from the chain rule for Γ_2 (i.e., (3.6) in the case $a = \frac{1}{2}$). To remove the last remaining L, note that

$$\int_E e^g\Gamma(g)Lg\, d\mu = -\int_E \Gamma(g, e^g\Gamma(g))\, d\mu = -\int_E e^g\Gamma(g, \Gamma(g)) + e^g\Gamma(g)^2\, d\mu.$$

Plugging this back into (3.8) and the result back into (3.7), we see that everything is written in terms of $\Gamma_2(g)$, $\Gamma(g, \Gamma(g))$, $\Gamma(g)$ and $\Gamma(g)^2$. It only remains to check the coefficients:

$$\int_E e^g\left[\frac{n-1}{n}\Gamma_2(g) + b_n\Gamma(g, \Gamma(g)) + c_n\Gamma(g)^2 - \rho\Gamma(g)\right] d\mu \geq 0 \tag{3.9}$$

where $b_n = \frac{2an+4a-3}{2n}$ and $c_n = \frac{na^2-(a-1)^2}{n}$. If we choose $a = \frac{3}{2n+4}$ then $b_n = 0$ and $c_n \leq 0$ for all $n \geq 1$. Hence, we can remove b_n and c_n from the inequality above, concluding that

$$\int_E e^g\Gamma_2(g)\, d\mu \geq \frac{n\rho}{n-1}\int_E e^g\Gamma(g)\, d\mu$$

for all bounded $g \in \mathcal{A}$. This implies that the condition of Lemma 3.3.4 holds for all bounded, positive $f \in \mathcal{A}$ that are also bounded away from zero. For a bounded, positive $f \in \mathcal{A}$, approximating it by $f + \epsilon$ allows us to verify the condition of Lemma 3.3.4 completely. □

3.4 Examples

Example 3.4.1 (The Ornstein-Uhlenbeck semigroup)**.** *We have already mentioned that the standard Gaussian measure satisfies the log-Sobolev inequality with constant* 1 *(since the Ornstein-Uhlenbeck semigroup satisfies* $\mathrm{CD}(1, \infty)$*. This is the sharp constant: it can be seen from the fact that (by Proposition 3.2.3) it yields the sharp Poincaré constant, or from the fact that Proposition 3.6.1 yields sharp tail bounds.*

Example 3.4.2 (The ultraspheric semigroup)**.** *Since the ultraspheric semigroup satisfies* $\mathrm{CD}(n-1, n)$*, Theorem 3.3.3 implies that it satisfies a log-Sobolev inequality with constant* $\frac{1}{n}$*. This is the sharp constant, since it implies (by Proposition 3.2.3) the sharp Poincaré constant of* $\frac{1}{n}$*.*

Example 3.4.3 (The wrapped heat semigroup)**.** *Theorem 3.3.3 has nothing to say about the wrapped heat semigroup on* $[0, 1]$*, which satisfies* $\mathrm{CD}(0, 1)$*. However, we already saw that the wrapped heat semigroup satisfies a Poincaré inequality (despite Theorem 2.1.1 having nothing to say about it), and so it is perhaps not so surprising that there is a log-Sobolev inequality also. In fact, we can show something a little stronger: under* $\mathrm{CD}(0, 1)$*, a Poincaré inequality with constant* C *is equivalent to a log-Sobolev inequality with constant* C*. Indeed, under a Poincaré inequality with constant* C*, Lemma 2.1.2 implies that*

$$\int_E \Gamma(f)\, d\mu \le C \int_E \Gamma_2(f)\, d\mu \tag{3.10}$$

for all $f \in \mathcal{A}$*. We will deduce from this the condition*

$$\int_E e^g \Gamma(g)\, d\mu \le C \int_E e^g \Gamma_2(g)\, d\mu \tag{3.11}$$

of Lemma 3.3.4, which will then lead to a log-Sobolev inequality. We begin by applying (3.10) *with* $f = e^{g/2}$*. Using the chain rule for* Γ_2*, this leads to*

$$\int_E e^g \left[\Gamma_2(g) + \frac{1}{2}\Gamma(g, \Gamma(g)) + \frac{1}{4}\Gamma(g)^2\right] d\mu \ge \int_E e^g \Gamma(g)\, d\mu.$$

In order to arrive at (3.11) *it suffices to show that*

$$\int_E e^g[\frac{1}{2}\Gamma(g,\Gamma(g)) + \frac{1}{4}\Gamma(g)^2]\,d\mu \leq 0 \tag{3.12}$$

for all $g \in \mathcal{A}$. But this follows from a computation that we already carried out in the proof of Theorem 3.3.3: by (3.9) *with $n = 1$ and $\rho = 0$,*

$$\int_E e^g\left[\frac{6a-3}{2}\Gamma(g,\Gamma(g)) + (2a-1)\Gamma^2(g)\right]\,d\mu \geq 0$$

for every a. With $a = \frac{1}{3}$,

$$\int_E e^g\left[\frac{1}{2}\Gamma(g,\Gamma(g)) + \frac{1}{3}\Gamma^2(g)\right]\,d\mu \leq 0,$$

which is stronger than (3.12)*. Hence, the wrapped heat semigroup on* $[0,1]$ *satisfies a log-Sobolev inequality with (sharp) constant $\frac{1}{4\pi^2}$.*

3.5 Decay of entropy

Recall that a Poincaré inequality was equivalent to exponential decay of the variance under P_t (Proposition 2.3.1. A similar equivalence holds between a log-Sobolev inequality and the decay of entropy. Indeed, the proofs in the Poincaré and log-Sobolev settings are analogous: in the case of the Poincaré inequality, the derivative of the variance along the semigroup was the negation of the Dirichlet form. Hence, the Poincaré inequality is equivalent to an exponential-type differential inequality for the variance along the semigroup. In this section (and as we have essentially already observed), the derivative of the entropy along the semigroup is the negation of the Fisher information. The Poincaré inequality is hence equivalent to an exponential-type differential inequality for the entropy along the semigroup.

Proposition 3.5.1. *The compact Markov triple (E, μ, Γ) satisfies a log-Sobolev inequality with constant C if and only if*

$$\mathrm{Ent}_\mu(P_t f) \leq e^{-2t/C}\,\mathrm{Ent}_\mu(f)$$

for every $t \geq 0$ and every $f \in L^1(\mu)$ with finite entropy.

Proof. It is enough to consider $f \in \mathcal{A}$: fix $f \in \mathcal{A}$ with finite entropy and define

$$\Lambda(t) = \mathrm{Ent}_\mu(P_t f) = \int_E P_t f \log P_t f\,d\mu - \int_E P_t f\,d\mu \log \int_E P_t f\,d\mu.$$

Note that the second term is actually constant in t. That is, the $\Lambda(t)$ here and the one from the proof of Lemma 3.3.4 differ only by an additive constant. In particular, we already computed there that

$$\Lambda'(t) = -\int_E \frac{\Gamma(P_t f)}{P_t f}\, d\mu = -I_\mu(P_t f).$$

Under the log-Sobolev inequality with constant C, $\Lambda'(t) \leq -\frac{2}{C}\Lambda(t)$, and it follows that $\Lambda(t) \leq e^{-2t/C}\Lambda(0)$.

In the other direction,

$$\Lambda(t) = \Lambda(0) + t\Lambda'(0) + o(t) = \Lambda(0) - tI_\mu(f) + o(t).$$

If $\Lambda(t) \leq e^{-2t/C}\Lambda(0) = (1 - \frac{2t}{C} + o(t))\Lambda(0)$ then as $t \to 0$ it follows that

$$I_\mu(f) \geq \frac{2}{C}\Lambda(0),$$

which is the log-Sobolev inequality with constant C. □

We may reformulate Proposition 3.5.1 in terms of probability distributions: for a probability distribution ν_0 that has a density f with respect to μ, let ν_t be the probability distribution $d\nu_t = P_t f\, d\mu$. If μ and Γ satisfy a log-Sobolev inequality with constant C, Proposition 3.5.1 implies that

$$H(\nu_t \mid \mu) \leq e^{-2t/C} H(\nu_0 \mid \mu). \tag{3.13}$$

In applications, convergence in total variation is often of interest: the total variation distance between probability distributions μ and ν is defined as

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

The well-known Pinsker-Csizsár-Kullback inequality relates total variation distance with entropy:

Proposition 3.5.2. *For any probability measures μ and ν on the same space,*

$$d_{\mathrm{TV}}(\mu, \nu)^2 \leq \frac{1}{2} H(\nu \mid \mu)$$

(where we declare $H(\nu \mid \mu)$ to be infinity if ν does not have a density with respect to μ).

Combining Proposition 3.5.2 with (3.13),

$$d_{\mathrm{TV}}(\mu,\nu_t)^2 \le \frac{1}{2} e^{-2t/C} H(\nu_0 \mid \mu).$$

That is, if we start with a probability measure that has finite entropy, then as we apply the semigroup it converges exponentially fast in total variation.

Note that the Poincaré inequality may also be used to deduce convergence in total variation: based on the equivalent (at least, when $\frac{d\nu}{d\mu}$ exists) form

$$d_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\int_E \left|1 - \frac{d\nu}{d\mu}\right| d\mu$$

for total variation distance, one can bound $d_{\mathrm{TV}}(\mu,\nu)^2 \le \frac{1}{4}\operatorname{Var}_\mu(\frac{d\nu}{d\mu})$ and then deduce exponential convergence of $d_{\mathrm{TV}}(\mu,\nu_t)$ from Proposition 2.3.1. The disadvantage with that approach is that one needs to start with $\frac{d\nu}{d\mu} \in L^2(\mu)$, which is a stronger condition than the assumption that ν has finite entropy with respect to μ.

Proof of Proposition 3.5.2. Let $f = \frac{d\nu}{d\mu}$ (if the density doesn't exist, the claim is trivial). It suffices to show that

$$\left(\int_E |1-f|\, d\mu\right)^2 \le 2\operatorname{Ent}_\mu(f).$$

For this, define $f_s = 1 + s(f-1)$ for $s \in [0,1]$ and set

$$\Lambda(s) = 2\operatorname{Ent}_\mu(f_s) - \left(\int_E |1-f_s|\, d\mu\right)^2 = 2\operatorname{Ent}_\mu(f_s) - s^2\left(\int_E |1-f|\, d\mu\right)^2.$$

Since $\int f_s\, d\mu = 1$ for all s, it follows that

$$\frac{d}{ds}\operatorname{Ent}_\mu(f_s) = \frac{d}{ds}\int_E f_s \log f_s\, d\mu = \int_E (f-1)(1+\log f_s)\, d\mu$$

and

$$\frac{d^2}{ds^2}\operatorname{Ent}_\mu(f_s) = \int_E \frac{(f-1)^2}{f_s}\, d\mu.$$

In particular, $\Lambda(0) = \Lambda'(0) = 0$ and

$$\Lambda''(s) = 2\int_E \frac{(f-1)^2}{f_s}\, d\mu - 2\left(\int_E |1-f|\, d\mu\right)^2 \ge 0,$$

where the inequality follows from the Cauchy-Schwarz inequality. Hence, $\Lambda(s) \ge 0$ for all $s \in [0,1]$ $\square$

3.6 Exponential integrability and Gaussian tails

By Proposition 2.5.1, a Poincaré inequality implies sub-exponential tail behavior. As we will see here, a log-Sobolev inequality implies sub-Gaussian tail behavior. Here, we assume that $\mathcal{E}$ takes the form $\mathcal{E}(f) = \int_E \Gamma(f)\, d\mu$ for a diffusion carré du champ Γ. Recall that a function f is said to be ℓ-Lipschitz if $\Gamma(f) \le \ell^2$ pointwise.

Proposition 3.6.1. *If μ and Γ satisfy a log-Sobolev inequality with constant $C > 0$ then for every 1-Lipschitz function f and every $s \in \mathbb{R}$,*

$$\int_E e^{sf}\, d\mu \le e^{s\int_E f\, d\mu + Cs^2/2}.$$

Moreover, for every $s < \frac{1}{\sqrt{C}}$,

$$\int_E e^{s^2 f^2/2}\, d\mu < \infty.$$

Proof. Let f be a bounded, 1-Lipschitz function in $\mathcal{A}$ (it suffices to check this case). Define $Z(s) = \int_E e^{sf}\, d\mu$, which is differentiable and satisfies

$$Z'(s) = \int_E f e^{sf}\, d\mu.$$

The log-Sobolev inequality $\mathrm{Ent}_\mu(e^{sf}) \le 2C\mathcal{E}(e^{sf/2})$ may be written in terms of Z: for the entropy,

$$\mathrm{Ent}_\mu(e^{sf}) = s\int_E f e^{sf}\, d\mu - Z(s)\log Z(s) = sZ'(s) - Z(s)\log Z(s).$$

On the other hand,

$$\mathcal{E}(e^{sf/2}) = \frac{s^2}{4}\int_E e^{sf}\Gamma(f)\, d\mu \le \frac{s^2}{4} Z(s).$$

Hence, the log-Sobolev inequality implies that

$$sZ'(s) \le Z(s)\log Z(s) + \frac{Cs^2}{2} Z(s)$$

for all $s \in \mathbb{R}$. Using the change of variables $F(s) = \frac{1}{s}\log Z(s)$ (where we set $F(0) = Z'(0) = \int_E f d\mu$ for continuity), we have $F'(s) \le \frac{C}{2}$. Integrating this inequality, we have

$$F(s) \le F(0) + \frac{Cs}{2} = \int_E f\, d\mu + \frac{Cs}{2}$$

for all $s \geq 0$ (and the reverse inequality for negative s). Hence, for all $s \in \mathbb{R}$

$$Z(s) = e^{sF(s)} \leq e^{s \int_E f\, d\mu + Cs^2/2}.$$

For the second claim, we integrate the first claim against a Gaussian measure: since $\int_{\mathbb{R}} e^{\sigma s}\, d\gamma_1(\sigma) = e^{s^2/2}$, Fubini's theorem implies

$$\begin{aligned}
\int_E e^{s^2 f^2/2}\, d\mu &= \int_E \int_{\mathbb{R}} e^{\sigma s f}\, d\gamma_1(\sigma)\, d\mu \\
&= \int_{\mathbb{R}} \int_E e^{\sigma s f}\, d\mu\, d\gamma_1(\sigma) \\
&\leq \int_{\mathbb{R}} e^{\sigma s \int_E f\, d\mu + C\sigma^2 s^2/2}\, d\gamma_1(\sigma).
\end{aligned}$$

This last integral is finite whenever $Cs^2 < 1$. □

As a consequence of Proposition 3.6.1, measures satisfying the log-Sobolev inequality have the property that Lipschitz functions are concentrated around their means with sub-Gaussian tails: if f is 1-Lipschitz with mean zero and μ, Γ satisfy a log-Sobolev inequality with constant C then by Propositon 3.6.1 and Markov's inequality,

$$\begin{aligned}
\mu(f \geq t) &= \mu\left(e^{sf} \geq e^{st}\right) \\
&\leq e^{-st} \int_E e^{sf}\, d\mu \\
&\leq e^{Cs^2/2 - st}.
\end{aligned}$$

Choosing $s = t/C$ gives $\mu(f \geq t) \leq e^{-\frac{t^2}{2C}}$. By scaling and shifting, this may be applied to an ℓ-Lipschitz function of arbitrary mean, yielding

$$\mu\left(f \geq \int_E f\, d\mu + t\right) \leq e^{-\frac{t^2}{2C\ell^2}}. \tag{3.14}$$

To test this estimate, consider the Gaussian measure. By Theorem 3.3.3 and the fact that the Ornstein-Uhlenbeck semigroup satisfies $\mathrm{CD}(1, \infty)$, the standard Gaussian measure on $\mathbb{R}$ satisfies a log-Sobolev inequality with constant 1. Choosing $f(x) = x$ (which is 1-Lipschitz), (3.14) implies that $\gamma_1(x \geq t) \leq e^{-t^2/2}$. On the other hand, classical bounds for the Gaussian tail imply that

$$\frac{t}{t^2+1} \leq \sqrt{2\pi} e^{t^2/2} \gamma_1(x \geq t) \leq \frac{1}{t}$$

whenever $t > 0$. This shows that the exponent in (3.14) is sharp, although there may be some missing polynomial factors.

Example 3.6.2 (The exponential measure). *Recall the exponential distribution $d\mu(x) = e^{-x}dx$ on $[0, \infty)$. In Section 2.6, we showed that the Poincaré constant of μ (with respect to the usual carré du champ $\Gamma(f, g) = f'g'$) is 4. On the other hand, Proposition 3.6.1 implies that μ does not have a finite log-Sobolev constant: if it did, then e^{sx} would be integrable against μ for every $s > 0$, which is clearly not the case. This is our first example of a measure that satisfies a Poincaré inequality but not a log-Sobolev inequality; in particular, it shows that Proposition 3.2.3 is only a one-directional implication.*

3.7 Hypercontractivity

One of the first things we proved about Markov semigroups was that they are contractions: if μ is an invariant measure then $P_t : L^p(\mu) \to L^p(\mu)$ is a contraction for every $1 \le p \le \infty$. One important consequence of a log-Sobolev inequality is that the associated semigroup is *hypercontractive*, meaning that it even gives contractions from $L^p(\mu)$ to $L^q(\mu)$ for some $q > p$.

Theorem 3.7.1. *For a compact Markov triple with semigroup P_t, the following are equivalent:*

(a) the log-Sobolev inequality holds with constant C; and

(b) for some (or every) $1 < p < \infty$, every $t \ge 0$ and every $f \in L^p(\mu)$,

$$\|P_t f\|_{q(t)} \le \|f\|_p,$$

where $q(t)$ is the solution of $\frac{q(t)-1}{p-1} = e^{2t/C}$.

(c) for some (or every) $-\infty < p < 1$, every $t \ge 0$ and every positive, bounded $f \in \mathcal{A}$,

$$\|P_t f\|_{q(t)} \ge \|f\|_p,$$

where $q(t)$ is as above.

Part (b) above is called a *hypercontractive inequality*, while part (b) is a *reverse hypercontractive inequality.*

Proof. Fix $1 < p < \infty$ and a bounded, positive $f \in \mathcal{A}$. Since $\mathcal{A}$ is dense and $P_t|f| \ge |P_t f|$, it suffices to check (b) for such functions. By the (usual calculus) chain rule,

$$\begin{aligned}\frac{\partial}{\partial q}\|f\|_q^q &= \frac{\partial}{\partial q}\int_E f^q \, d\mu \\ &= \int_E f^q \log f \, d\mu.\end{aligned}$$

The chain rule also gives

$$\frac{d}{dq}u(q)^{1/q} = \frac{1}{q}u(q)^{1/q-1}u'(q) - \frac{1}{q^2}u(q)^{1/q}\log u(q);$$

applying this to $u(q) = \|f\|_q^q$ gives

$$\frac{\partial}{\partial q}\|f\|_q = \frac{\|f\|_q^{1-q}}{q^2}\operatorname{Ent}_\mu(f^q).$$

On the other hand,

$$\begin{aligned}\frac{\partial}{\partial t}\|P_t f\|_q &= \frac{\|P_t f\|_q^{1-q}}{q}\frac{\partial}{\partial t}\int_E (P_t f)^q \, d\mu \\ &= \|P_t f\|_q^{1-q}\int_E (P_t f)^{q-1} L P_t f \, d\mu \\ &= -\frac{4(q-1)\|P_t f\|_q^{1-q}}{q^2}\int_E \Gamma((P_t f)^{q/2}) \, d\mu.\end{aligned}$$

Now define $\Lambda(t,q) = \|P_t f\|_q$. For any function $q(t)$ that is differentiable in t, the chain rule yields

$$\frac{d}{dt}\Lambda(t,q(t)) = \frac{q'\Lambda^{1-q}}{q^2}\operatorname{Ent}_\mu((P_t f)^q) - \frac{4(q-1)\Lambda^{1-q}}{q^2}\mathcal{E}((P_t f)^{q/2}). \quad (3.15)$$

Under a log-Sobolev inequality with constant C,

$$\frac{d}{dt}\Lambda(t,q(t)) \le (2Cq' - 4(q-1))\frac{\Lambda^{1-q}}{q^2}\mathcal{E}((P_t f)^{q/2}).$$

For our choice of q, $2Cq' = 4(q-1)$ and so $\frac{d}{dt}\Lambda(t,q(t)) \le 0$. It follows that

$$\|P_t f\|_{q(t)} = \Lambda(t) \le \Lambda(0) = \|f\|_p.$$

The previous argument may also be used to show that the log-Sobolev inequality implies the reverse hypercontractive inequality: the only difference is that we begin with $p < 1$, which implies that $q'(t) < 0$. Hence, the log-Sobolev inequality applied to (3.15) produces an inequality in the other direction.

On the other hand, suppose that the hypercontractive (or reverse hypercontractive) inequality holds for some p and all t. Then $\frac{d}{dt}\big|_{t=0} \Lambda(t, q(t)) \leq 0$, which by (3.15) implies the log-Sobolev inequality with constant C. $\square$

An immediate consequence of Theorem 3.7.1 is that all eigenfunctions of the generator belong to $L^q(\mu)$ for every $2 < q < \infty$. Indeed, if $Lf = -\lambda f$ for $\lambda > 0$ then $P_t f = e^{-\lambda t} f$. Applying Theorem 3.7.1 with $p = 2$ and $t = \frac{C}{2}\log(q-1)$,

$$e^{-\lambda t}\|f\|_q = \|P_t f\|_q \leq \|f\|_2,$$

which yields

$$\|f\|_q \leq (q-1)^{C\lambda/2}\|f\|_2.$$

Chapter 4

Isoperimetric inequalities

4.1 Surface area

One of our applications of the log-Sobolev inequality was a tail estimate that was *almost* sharp for the Gaussian measure. Since the Gaussian measure is our model space for the $\mathrm{CD}(1,\infty)$ condition, it is natural to ask whether we can get an exact tail estimate under this assumption. This is one motivation for introducing the notion of an isoperimetric inequality.

We begin with the notion of bounded variation: for the remainder of this section, we fix a compact Markov triple (E,μ,Γ).

Definition 4.1.1. *For $f \in L^1(\mu)$, define*

$$\|f\|_{\mathrm{BV}} = \inf_{f_n} \liminf_{n\to\infty} \int_E \sqrt{\Gamma(f_n)}\, d\mu,$$

where the infimum ranges over all sequences $f_n \in \mathcal{A}$ satisfying $f_n \to f$ in $L^1(\mu)$. Note that $\|f\|_{\mathrm{BV}}$ may be infinite.

This defines a seminorm on $\{f \in L^1(\mu) : \|f\|_{\mathrm{BV}} < \infty\}$ (not a norm, because if f is constant then $\|f\|_{\mathrm{BV}} = 0$). Indeed, the triangle inequality for $\|\cdot\|_{\mathrm{BV}}$ follows from the Cauchy-Schwarz inequality for Γ, since

$$\Gamma(f+g) = \Gamma(f) + \Gamma(g) + 2\Gamma(f,g) \le (\sqrt{\Gamma(f)} + \sqrt{\Gamma(g)})^2.$$

Taking square roots and integrating proves the triangle inequality for $\|f\|_{\mathrm{BV}}$.

Definition 4.1.2. *For a measurable subset A of E, $\mu^+(A) = \|1_A\|_{\mathrm{BV}}$. We call this the μ-*surface area *of A.*

Example 4.1.3. *Let $E = \mathbb{R}$, $\Gamma(f,g) = f'g'$ (with $\mathcal{A} = C_c^\infty(\mathbb{R})$), and suppose that μ has a continuous density ϕ with respect to the Lebesgue measure. Then for any $a < b$ and an interval $A = [a,b]$, $\mu^+(A) = \phi(a) + \phi(b)$.*

To see that $\mu^+(A) \le \phi(a) + \phi(b)$, it suffices to find a sequence of smooth f_ϵ such that $f_\epsilon \to 1_A$ as $\epsilon \to 0$ and

$$\liminf_{\epsilon \to 0} \int_{\mathbb{R}} |f'_\epsilon(x)| d\mu(x) \le \phi(a) + \phi(b).$$

This can be easily done by convolving 1_A with compactly supported, smooth mollifiers. For example, in this way one can construct functions $f_\epsilon : \mathbb{R} \to [0,1]$ satisfying

- *$f_\epsilon = 0$ outside of $[a-\epsilon, b+\epsilon]$,*
- *$f_\epsilon = 1$ on $[a+\epsilon, b-\epsilon]$,*
- *f_ϵ is non-decreasing on $[a-\epsilon, a+\epsilon]$, and*
- *f_ϵ is non-increasing on $[b-\epsilon, b+\epsilon]$.*

For such a function f_ϵ, we have

$$\|f_\epsilon - 1_A\|_{L^1(\mu)} \le \mu([a-\epsilon, a+\epsilon]) + \mu([b-\epsilon, b+\epsilon]),$$

which converges to zero since μ has a continuous density. On the other hand,

$$\int_{\mathbb{R}} |f'_\epsilon| \, d\mu = \int_{a-\epsilon}^{a+\epsilon} f'_\epsilon(x)\phi(x)\, dx - \int_{b-\epsilon}^{b+\epsilon} f'_\epsilon(x)\phi(x)\, dx.$$

Since ϕ is continuous, as $\epsilon \to 0$ one can approximate the right hand side above by

$$\phi(a) \int_{a-\epsilon}^{a+\epsilon} f'_\epsilon(x)\, dx - \phi(b) \int_{b-\epsilon}^{b+\epsilon} f'_\epsilon(x)\, dx = \phi(a) + \phi(b).$$

In order to check that $\mu^+(A) \ge \phi(a) + \phi(b)$, one needs to show that every sequence of smooth functions f_n approximating 1_A satisfies

$$\liminf_{n \to \infty} \int_{\mathbb{R}} |f'_\epsilon(x)| \, d\mu(x) \ge \phi(a) + \phi(b).$$

Suppose that $\phi(a)$ and $\phi(b)$ are both strictly positive (otherwise, a small modification of the following argument will also work). Then for any $\epsilon > 0$ there exists a $\delta > 0$ (depending on $\phi(a)$, $\phi(b)$, and the continuity of ϕ near *a and b) such that if $\|f - 1_A\|_{L^1(\mu)} \le \delta$ then there $a^- < a^+ \in [a-\epsilon, a+\epsilon]$ and $b^- < b^+ \in [b-\epsilon, b+\epsilon]$ with $f(a^-), f(b^+) \le \epsilon$ and $f(a^+), f(b^-) \ge 1-\epsilon$. Then*

$$\int_{\mathbb{R}} |f'(x)|\phi(x)\,dx \ge \int_{a^-}^{a^+} |f'(x)|\phi(x)\,dx + \int_{b^-}^{b^+} |f'(x)|\phi(x)\,dx.$$

By taking ϵ small, we can ensure that $\phi(x)$ is nearly constant on $[a-\epsilon, a+\epsilon]$ and $[b-\epsilon, b+\epsilon]$. In particular,

$$\begin{aligned}\int_{\mathbb{R}} |f'(x)|\phi(x)\,dx &\ge \phi(a)\int_{a^-}^{a^+} f'(x)\,dx + \phi(b)\int_{b^-}^{b^+} f'(x)\,dx - o_\epsilon(1)\\ &\ge (1-2\epsilon)(\phi(a)+\phi(b)) - o_\epsilon(1),\end{aligned}$$

where $o_\epsilon(1)$ is some quantity converging to zero as $\epsilon \to 0$.

Exercise 4.1.1. *Extend the previous example to the case where the density ϕ is not necessarily continuous.*

4.1.1 Minkowski content

When there is a metric on the space E, the Minkowski content is another common method of defining surface area. For simplicity, we restrict here to the setting $E = \mathbb{R}^n$.

Definition 4.1.4. *Let μ be a measure on $\mathbb{R}^n$. For a measurable set A, define*

$$\tilde{\mu}^+(A) = \liminf_{\epsilon \to 0} \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon},$$

where $A_\epsilon = \{x \in \mathbb{R}^n : d(x, A) \le \epsilon\}$. We call $\tilde{\mu}^+$ the Minkowski content*.*

Even when $\Gamma(f) = |\nabla f|^2$, the Minkowski content and the μ-surface area we defined above are not always the same. As an example, in $\mathbb{R}^2$ with μ the Lebesgue measure, consider the set $A = ([0,1]\times[0,1]) \cup ([1,2]\times\{1\})$. This set looks like a square with a line sticking out of it, and one can easily check that $\mu(A_\epsilon) = 6\epsilon + O(\epsilon^2)$. On the other hand, one can approximate A by a smooth version of the indicator function of $[0,1]\times[0,1]$ (because the piece $[1,2]\times\{1\}$ has measure zero. Hence, one sees that $\mu^+(A) \le 4$. This is actually a demonstration of a more general phenomenon: μ^+ is unaffected by changes of measure zero, whereas $\tilde{\mu}^+$ might be.

Nevertheless, there is a relationship between μ^+ and $\tilde{\mu}^+$.

Proposition 4.1.5. *If $\Gamma(f) = |\nabla f|^2$ then for every measurable $A \subset \mathbb{R}^n$, $\mu^+(A) \le \tilde{\mu}^+(A)$.*

Proof. For simplicity, we will assume that A is bounded. The general case may be attained by an approximation argument.

For $\epsilon > 0$, let $f_\epsilon = (1 - (\epsilon + 2\epsilon^2)^{-1} d(x, A_{\epsilon^2}))_+$. Then f_ϵ is not smooth, but it is $(\epsilon+2\epsilon^2)^{-1}$-Lipschitz. Moreover, f_ϵ is 1 on A_{ϵ^2} and $f_\epsilon = 0$ outside of $A_{\epsilon-\epsilon^2}$. By convolving f_ϵ with a suitable smooth mollifier, one can construct a smooth function $g_\epsilon : \mathbb{R}^n \to [0,1]$ with the following properties:

- $g_\epsilon = 1$ on A and 0 outside A_ϵ,
- $g_\epsilon \in C_c^\infty(\mathbb{R}^n)$ and is $(\epsilon + 2\epsilon^2)^{-1}$-Lipschitz.

From the first property and the fact that g_ϵ takes values in $[0,1]$, $\|g_\epsilon - 1_A\| \le \mu(A_\epsilon) - \mu(A)$, which we may assume converges to zero for some sequence of $\epsilon \to 0$ (otherwise $\tilde{\mu}^+(A) = \infty$, so there is nothing to prove). On the other hand, $\sqrt{\Gamma(g_\epsilon)}$ is zero on A and outside A_ϵ; hence,

$$\int_{\mathbb{R}^n} \sqrt{\Gamma(g_\epsilon)}\, d\mu \le \frac{\mu(A_\epsilon) - \mu(A)}{\epsilon + 2\epsilon^2}$$

and so $\mu^+(A) \le \liminf_{\epsilon \to 0} \int \sqrt{\Gamma(g_\epsilon)}\, d\mu \le \tilde{\mu}^+(A)$. $\square$

Exercise 4.1.2. *Take $E = \mathbb{R}^n$, $\Gamma(f) = |\nabla f|^2$, and suppose that μ has a smooth density ϕ.*

(a) *Show that*

$$\mu^+(A) = \sup\left\{ \int_A \operatorname{div}(\phi v)\, dx : v \in C_c^\infty(\mathbb{R}^n; \mathbb{R}^n), \sup_x |v(x)| \le 1 \right\}.$$

(b) *We say that $A \subset \mathbb{R}^n$ has a smooth boundary if there is some smooth function $\psi : \mathbb{R}^n \to \mathbb{R}$ such that $A = \{x : \psi(x) \le 0\}$ and $\nabla\psi(x) \ne 0$ on ∂A. Show that if A is compact with smooth boundary then $\tilde{\mu}^+(A) = \mu^+(A)$.*

4.2 Isoperimetric inequalities

An isoperimetric inequality is a lower bound on the surface area of a set in terms of its measure. The most famous isoperimetric inequality is the classical Euclidean isoperimetric inequality, which says that the Lebesgue surface area of $A \subset \mathbb{R}^n$ is at least as large as the surface area of a Euclidean ball with the same volume.

Definition 4.2.1. *Given a measure μ and a diffusion carré du champ Γ, an* isoperimetric function *is a function $\mathcal{I} : [0, \mu(E)] \to \mathbb{R}_+$ such that for every measurable A,*

$$\mu^+(A) \geq \mathcal{I}(\mu(A)).$$

The largest possible isoperimetric function is called the isoperimetric profile*: the isoperimetric profile of μ and Γ is the function $\mathcal{I} : [0, \mu(E)] \to \mathbb{R}_+$ defined by*

$$\mathcal{I}(a) = \inf\{\mu^+(A) : A \subset E \textit{ measurable, } \mu(A) = a\}.$$

On $\mathbb{R}^n$ with μ the Lebesgue measure, the classical Euclidean isoperimetric inequality is equivalent to the statement that the isoperimetric profile for $\mathbb{R}^n$ and μ is

$$\mathcal{I}_n(a) = n\mu(B)^{1/n} a^{\frac{n-1}{n}},$$

where B is the unit Euclidean ball.

Here, however, we will focus on Gaussian-type isoperimetric inequalities. Hence, the isoperimetric profile of the standard Gaussian measure will be of central importance. From now on, let $\phi : \mathbb{R} \to \mathbb{R}$ be the density of the standard Gaussian measure on $\mathbb{R}$, and let Φ be its cumulative distribution function:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
$$\Phi(x) = \int_{-\infty}^{x} \phi(y)\, dy.$$

We note that both ϕ and Φ may be naturally extended to $\pm\infty$: set $\phi(\pm\infty) = 0$ and $\Phi(\infty) = 1$, $\Phi(-\infty) = 0$. Note also that Φ is strictly increasing, and so it has a inverse function $\Phi^{-1} : [0, 1] \to \mathbb{R} \cup \{\pm\infty\}$.

Theorem 4.2.2 (Gaussian isoperimetric inequality)**.** *The isoperimetric profile for γ_n is $\mathcal{I} = \phi \circ \Phi^{-1}$.*

Based on Theorem 4.2.2, we say that a measure has a *Gaussian-type* isoperimetric inequality if it's isoperimetric function is larger than the Gaussian one: we say that μ has Gaussian-type isoperimetry if there is some $c > 0$ such that $c\phi \circ \Phi^{-1}$ is an isoperimetric function for μ.

There are of course two directions of Theorem 4.2.2 that need to be proven. The more difficult direction is to show that $\mathcal{I}$ is an isoperimetric function; we will do this later. In order to show that $\mathcal{I}$ is the best possible isoperimetric function, however, it suffices to give an example showing equality. For this, we begin with the case $n = 1$: for $a \in [0, 1]$, define $A = (-\infty, \Phi^{-1}(a)]$. By the definition of Φ,

$$\gamma_1(A) = \int_{-\infty}^{\Phi^{-1}(a)} \phi(y)\, dy = a.$$

In order to find the surface area of A, we consult Example 4.1.3. Technically, that example only considered the surface area of finite intervals, but it can easily be extended to the semi-infinite case. Hence, $\mu^+(A) = \phi(\Phi^{-1}(a)) = \mathcal{I}(a)$. In particular, the isoperimetric profile of γ_1 is at most $\mathcal{I}$. For $n > 1$, it suffices to consider sets of the form $A = \{x : x_1 \le \Phi^{-1}(a)\}$; such a set has $\mu(A) = a$ and $\mu^+(A) = \mathcal{I}(a)$. Since γ_n is rotationally invariant, one can of course also take $A = \{x : \langle x, \theta\rangle \le \Phi^{-1}(a)\}$ for any $\theta \in S^{n-1}$.

4.3 Isoperimetry and expansion

For now, fix $E = \mathbb{R}^n$ and $\Gamma(f) = |\nabla f|^2$. Recall that in this setting the Minkowski content $\tilde{\mu}^+$ is always at least as large as the μ-surface area μ^+, and that an isoperimetric function $\mathcal{I}$ is a lower bound on the μ-surface area. Therefore it is also a lower bound on the Minkowski content. Since the Minkowski content is a sort of derivative, we may integrate out this inequality in order to obtain bounds on $\mu(A_r)$ (recall that $A_r = \{x : d(x, A) \le r\}$).

Proposition 4.3.1. *Let $\mathcal{I}$ be an isoperimetric function for μ, which we assume to be strictly positive on $(0, \mu(\mathbb{R}^n))$. Suppose that $g : \mathbb{R} \to (0, \mu(\mathbb{R}^n))$ solves $g'(r) = \mathcal{I}(g(r))$. Then for every measurable $A \subset \mathbb{R}^n$ and every $r > 0$, $\mu(A_r) \ge g(g^{-1}(\mu(A_r)) + r)$.*

Proof. For the duration of this proof, we write f' for the lower derivative of f; i.e., $f'(x) = \liminf_{h\to 0} \frac{f(x+h)-f(x)}{h}$. Let $f(r) = \mu(A_r)$; then the definition of Minkowski content implies that $f'(r) = \tilde{\mu}^+(A_r)$ (here, we are also using the fact that $(A_r)_s = A_{r+s}$). Since $\mathcal{I}$ is an isoperimetric function, $f'(r) \ge \mathcal{I}(f(r))$ for all $r \ge 0$.

Since $\mathcal{I}$ is strictly positive, g is strictly increasing. Hence its inverse is defined on $(0, \mu(\mathbb{R}^n))$; we may extend it by continuity to $[0, \mu(\mathbb{R}^n)]$. Let $h = g^{-1} \circ f$. By the chain rule (appropriately extended for lower derivatives),

$$h'(r) = \frac{f'(r)}{g'(g^{-1}(f(r)))} = \frac{f'(r)}{\mathcal{I}(f(r))} \ge 1.$$

It follows that $g^{-1}(f(r)) \ge g^{-1}(f(0)) + r$. Since $f(0) = \mu(A)$, the claim follows. $\square$

Combining Proposition 4.3.1 with Theorem 4.2.2 and the function $g(r) = \Phi(r)$, we obtain a lower bound on the Gaussian measure of an expansion.

Corollary 4.3.2. *For any measurable $A \subset \mathbb{R}^n$, $\gamma(A_r) \geq \Phi(\Phi^{-1}(A) + r)$.*

Note that the bound of Corollary 4.3.2 is sharp for sets of the form $A = \{x : x_1 \leq \Phi^{-1}(a)\}$. Indeed, for this set $A_r = \{x : x_1 \leq \Phi^{-1}(a) + r\}$. In one dimension, applying Corollary 4.3.2 to the set $A = (-\infty, 0]$ yields $\gamma_1([t, \infty)) \leq 1 - \Phi(t)$ for $t \geq 0$. Of course, this is a sharp inequality since it is essentially the definition of Φ. The point, however, is that an isoperimetric inequality can give a sharp bound for tail decay.

4.3.1 The coarea formula

Of course, expansions as we defined them above make sense when we have a metric that is compatible with Γ in some sense. In our general setting, however, it is more natural to work with thresholds of Lipshitz functions instead of expansions of sets. To this end, we introduce the *co-area formula.* From now on, we are in the setting of a compact Markov diffusion (E, μ, Γ). In what follows, $\int^*$ denotes the outer integral

$$\int_E^* f \, d\mu = \inf \left\{ \int_E g \, d\mu : f \geq g \text{ is measurable} \right\}.$$

Proposition 4.3.3. *For any $f \in \mathcal{A}$, let $A_r(f) \subset E$ be the set $\{x : f(x) > r\}$. Then*

$$\int_{\mathbb{R}}^* \mu^+(A_r(f)) \, dr \leq \int_E \sqrt{\Gamma(f)} \, d\mu.$$

In nice settings (e.g. $\mathbb{R}^n$ with the usual carré du champ), the inequality of Proposition 4.3.3 is actually an equality.

Before proving Proposition 4.3.3, it will be convenient to extend Γ from $\mathcal{A}$ to $\mathcal{D}(\mathcal{E})$. Indeed, if $f \in \mathcal{D}(\mathcal{E})$ then there exists a sequence $f_n \in \mathcal{A}$ such that $f_n \to f$ in $L^2(\mu)$ and f_n is Cauchy with respect to $\mathcal{E}$. Since

$$\mathcal{E}(f_n - f_m) = \int_E \Gamma(f_n) + \Gamma(f_m) - 2\Gamma(f_m, f_n) \, d\mu \geq \int_E (\sqrt{\Gamma(f_n)} - \sqrt{\Gamma(f_m)})^2 \, d\mu,$$

it follows that $\sqrt{\Gamma(f_n)}$ is Cauchy (and hence converges) in $L^2(\mu)$. We may therefore define $\sqrt{\Gamma(f)} \in L^2(\mu)$ to be this limit. Of course, this procedure also defines $\Gamma(f) \in L^1(\mu)$.

Exercise 4.3.1. *Verify that the chain rule $\Gamma(\psi(f)) = (\psi'(f))^2 \Gamma(f)$ remains valid for $f \in \mathcal{D}(\mathcal{E})$ and for smooth functions ψ with bounded derivatives.*

Exercise 4.3.2. *Recall that if $f \in \mathcal{D}(\mathcal{E})$ then $a \wedge f \vee b \in \mathcal{D}(\mathcal{E})$ for all $a < b$. Show that $\Gamma(a \wedge f \vee b) = 1_{\{a \le f \le b\}}\Gamma(f) = 1_{\{a < f < b\}}\Gamma(f)$.*

Proof of Proposition 4.3.3. For $k \in \mathbb{N}$ and $j \in \mathbb{Z}$, let

$$f_{j,k} = 0 \vee k\big(f - \frac{j}{k}\big) \wedge 1.$$

By Exercise 4.3.2,

$$\Gamma(f_{j,k}) = \Gamma(f) 1_{\{\frac{j}{k} \le f < \frac{j+1}{k}\}}.$$

Hence, for every $k \in \mathbb{N}$

$$\int_E \sqrt{\Gamma(f)}\, d\mu = \sum_{j \in \mathbb{Z}} \int_E \sqrt{\Gamma(f_{j,k})}\, d\mu.$$

Now fix some $r \in \mathbb{R}$. If $\{f = r\}$ has measure zero (which happens for all but at most countably many r) then as $k \to \infty$, $f_{\lfloor rk \rfloor, k}$ converges in $L^1(\mu)$ to $1_{\{f > r\}}$. It follows by the definition of μ^+ that

$$\mu^+(\{f > r\}) \le \liminf_{k \to \infty} \int_E \sqrt{\Gamma(f_{\lfloor rk \rfloor, k})}.$$

Hence, define the function $\psi_k : \mathbb{R} \to \mathbb{R}$ by

$$\psi_k(r) = \int_E \sqrt{\Gamma(f_{\lfloor rk \rfloor, k})}\, d\mu$$

Then ψ_k is a measurable function such that $\mu^+(\{f > r\}) \le \liminf_{k \to \infty} \psi_k(r)$ for all but countably many r. It follows from Fatou's lemma that

$$\begin{aligned} \int_{\mathbb{R}}^{*} \mu^+(\{f > r\})\, dr &\le \liminf_{k \to \infty} \int_{\mathbb{R}} \psi_k(r)\, dr \\ &= \liminf_{k \to \infty} \sum_{j \in \mathbb{Z}} \int_E \sqrt{\Gamma(f_{j,k})}\, d\mu \\ &= \int_E \sqrt{\Gamma(f)}\, d\mu. \end{aligned}$$ □

By a similar argument to the one in Proposition 4.3.3, we get an analogue of Proposition 4.3.1 in the setting of a compact Markov triple.

Corollary 4.3.4. *Suppose that $\mathcal{I}$ is a positive isoperimetric function for μ and Γ, and let g be as in Proposition 4.3.1. Then for any 1-Lipschitz $f \in \mathcal{A}$ and any $t \in \mathbb{R}$, $r > 0$,*

$$\mu(\{f > t + r\}) \ge g(g^{-1}(\mu(\{f > t\})) + r).$$

Proof. Consider the function $\psi(s) = \mu(\{f \le s\})$. Then

$$\begin{aligned}
\psi(t) - \psi(s) &= \int_E 1_{\{s<f\le t\}}\, d\mu \\
&= \int_E \sqrt{\Gamma(s \vee f \wedge t)}\, d\mu \\
&\ge \int_{\mathbb{R}}^{*} \mu^+(A_r(s \vee f \wedge t))\, dr \\
&\ge \int_s^t \mathcal{I}(\mu(\{f \le r\}))\, dr \\
&= \int_s^t \mathcal{I}(\psi(r))\, dr,
\end{aligned}$$

where the first inequality follows from Proposition 4.3.3 and the second from the fact that $\mathcal{I}$ is an isoperimetric function for μ. Hence, $\psi'(s) \ge \mathcal{I}(\psi(s))$, where ψ' denotes the lower derivative. We conclude by applying the same argument as in Proposition 4.3.1. □

4.4 Gaussian-type isoperimetry under $\mathrm{CD}(\rho, \infty)$

For the rest of this section, let $\mathcal{I} = \phi \circ \Phi^{-1}$ be the Gaussian isoperimetric profile. Our goal here is to prove the Gaussian isoperimetric inequality, which we will do by seeing how $P_s\mathcal{I}(P_{t-s}f)$ changes with s. We will begin, however, with a bound in the opposite direction (i.e., an upper bound on $\Gamma(f)$). This inequality should be compared to the reverse Poincaré or reverse log-Sobolev inequalities.

Proposition 4.4.1. *Let (E, μ, Γ) be a compact Markov triple that satisfies $\mathrm{CD}(\rho, \infty)$ for some $\rho \in \mathbb{R}$. For any $f \in \mathcal{A}$ taking values in $(0, 1)$ and any $t > 0$,*

$$\Gamma(P_t f) \le \frac{\rho}{e^{2\rho t} - 1}(\mathcal{I}^2(P_t f) - (P_t\mathcal{I}(f))^2).$$

Proof. Assume that f takes values in $[\epsilon, 1-\epsilon]$ for some $\epsilon > 0$ (this assumption may be removed at the end by taking limits). Then $\mathcal{I}$ is smooth on an open set strictly containing the range of f, and so the following computations are justified.

Fix $t > 0$ and define $\Lambda(s) = P_{t-s}\mathcal{I}(P_s f)$. By a direct computation, $\mathcal{I}'' = -\frac{1}{\mathcal{I}}$; hence,

$$\Lambda'(s) = P_{t-s}\frac{\Gamma(P_s f)}{\mathcal{I}(P_s f)}.$$

By the Cauchy-Schwarz inequality applied in the form $P_{t-s}\frac{g^2}{h} \geq \frac{(P_{t-s}g)^2}{P_{t-s}h}$ and the strong gradient bound,

$$\Lambda'(s) \geq \frac{(P_{t-s}\sqrt{\Gamma(P_s f)})^2}{P_{t-s}\mathcal{I}(P_s f)} \geq e^{2\rho(t-s)}\frac{\Gamma(P_t f)}{\Lambda(s)}.$$

Then the chain rule yields

$$\frac{d}{ds}(\Lambda(s))^2 = 2\Lambda(s)\Lambda'(s) \geq 2e^{2\rho(t-s)}\Gamma(P_t f).$$

Integrating,

$$\Lambda^2(t) - \Lambda^2(0) \geq 2\Gamma(P_t f)\int_0^t e^{2\rho(t-s)}\,ds = \frac{e^{2\rho t}-1}{\rho}\Gamma(P_t f). \qquad \square$$

Now we are prepared to prove the Gaussian isoperimetric inequality. Actually, we will prove something stronger: an isoperimetric inequality for diffusions satisfying $\mathrm{CD}(\rho,\infty)$.

Theorem 4.4.2. *Let (E,μ,Γ) be a compact Markov triple that satisfies $\mathrm{CD}(\rho,\infty)$ for some $\rho > 0$. Then for any $f \in \mathcal{A}$ with values in $(0,1)$,*

$$P_t\sqrt{\Gamma(f)} \geq \frac{\sqrt{\rho}}{\sqrt{1-e^{-2\rho t}}}(\mathcal{I}(P_t f) - P_t\mathcal{I}(f)).$$

Before proving Theorem 4.4.2, let us see how it implies an isoperimetric inequality. First, take $t \to \infty$ and integrate Theorem 4.4.2 with respect to μ, yielding

$$\int_E \sqrt{\Gamma(f)}\,d\mu \geq \sqrt{\rho}\left(\mathcal{I}\left(\int_E f\,d\mu\right) - \int \mathcal{I}(f)\,d\mu\right). \tag{4.1}$$

For a measurable A, let $f_n \in \mathcal{A}$ be a sequence approaching 1_A in $L^1(\mu)$. By truncating, we may assume that f_n take values in $(0,1)$ (because truncating f_n only reduces $\int_E \sqrt{\Gamma(f_n)}\,d\mu$). Therefore, we may apply (4.1) to each f_n. In doing so, note that the convergence $f_n \to 1_A$ implies that $\int_E f_n\,d\mu \to \mu(A)$ and that $\mathcal{I}(f_n) \to 0$ in $L^1(\mu)$. Hence,

$$\liminf_{n\to\infty}\int_E \sqrt{\Gamma(f_n)} \geq \sqrt{\rho}\mathcal{I}(\mu(A)),$$

which implies that $\sqrt{\rho}\mathcal{I}$ is an isoperimetric function for μ.

Proof of Theorem 4.4.2. Take $\Lambda(s) = P_{t-s}\mathcal{I}(P_s f)$. As in the proof of Proposition 4.4.1, $\Lambda'(s) = P_{t-s}\frac{\Gamma(P_s f)}{\mathcal{I}(P_s f)}$. Now, Proposition 4.4.1 implies that

$$\sqrt{\Gamma(P_s f)} \le \frac{\sqrt{\rho}}{\sqrt{1-e^{-2\rho s}}}\mathcal{I}(P_s f),$$

and so

$$\Lambda'(s) \le \frac{\sqrt{\rho}}{\sqrt{e^{2\rho s}-1}} P_{t-s}\sqrt{\Gamma(P_s f)} \le \frac{\sqrt{\rho}e^{-\rho s}}{\sqrt{e^{2\rho s}-1}} P_t\sqrt{\Gamma(f)}.$$

Integrating out s,

$$\Lambda(t) - \Lambda(0) \le \sqrt{\rho} P_t \sqrt{\Gamma(f)} \int_0^t \frac{1}{\sqrt{e^{2\rho s}-1}}\, ds = \frac{\sqrt{1-e^{-2\rho t}}}{\sqrt{\rho}} P_t\sqrt{\Gamma(f)}. \quad \square$$

4.5 Bobkov's inequality

Although we have already investigated Gaussian-type isoperimetric inequalities, it turns out that there is a stronger functional inequality than Theorem 4.2.2. This inequality is known as Bobkov's inequality, and we will investigate it in a way that bears strong similarities to the local Poincaré and log-Sobolev inequalities.

Theorem 4.5.1. *Let (E, μ, Γ) be a compact Markov triple with semigroup P_t. Then the following are equivalent:*

(a) the curvature-dimension condition $\mathrm{CD}(\rho, \infty)$ holds for $\rho \in \mathbb{R}$; and

(b) for every $f \in \mathcal{A}$ taking values in $[0, 1]$ and every $t \ge 0$,

$$\sqrt{\rho\mathcal{I}^2(P_t f) + \Gamma(P_t f)} \le P_t\sqrt{\rho\mathcal{I}^2(f) + \Gamma(f)}. \tag{4.2}$$

If $\rho > 0$ in Theorem 4.5.1 then we may take $t \to \infty$; the gradient bound ensures that $\Gamma(P_t f) \to 0$, and hence

$$\sqrt{\rho}\mathcal{I}\left(\int_E f\, d\mu\right) \le \int_E \sqrt{\rho\mathcal{I}^2(f) + \Gamma(f)}\, d\mu. \tag{4.3}$$

This last equation is known as Bobkov's inequality, while we call (4.2) a local Bobkov inequality. Note that (4.3) implies (4.1) via the inequality $\sqrt{a^2+b^2} \le |a| + |b|$. In particular, it also implies the Gaussian isoperimetric inequality.

4.5.1 Equality cases for Bobkov's inequality

Recall that the Gaussian isoperimetric inequality is an equality for the standard Gaussian measure γ_n and a set of the form $\{x \in \mathbb{R}^n : \langle x, \theta \rangle \le b\}$. Hence, the functional inequality (4.1) almost achieves equality when f is the smoothed indicator of a half-space. However, when $E = \mathbb{R}^n$ and $\mu = \gamma_n$ then (4.1) actually never achieves equality for any smooth, non-constant function. This may be deduced from the inequality (4.3), since the inequality $\sqrt{a^2 + b^2} \le |a| + |b|$ is strict unless either a or b is zero. In particular, in order to maintain equality in going from (4.3) to (4.1) we must have

$$\gamma_n(\{x : f(x) \in \{0, 1\} \text{ or } \nabla f = 0\}) = 1,$$

which is incompatible with f being a smooth, non-constant function.

On the other hand, there is a family of smooth functions achieving equality in (4.3). Indeed, consider $f : \mathbb{R}^n \to [0, 1]$ defined by $f(x) = \Phi(\langle x, a \rangle + b)$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Taking random variables $X \sim \gamma_n$ and $Z \sim \gamma_1$,

$$\begin{aligned}
\int_{\mathbb{R}^n} f(x)\, d\gamma_n(x) &= \mathbb{E}\Phi(\langle X, a \rangle + b) \\
&= \Pr(Z \le \langle X, a \rangle + b) \\
&= \Pr(Z\sqrt{1 + |a|^2} \le b) \\
&= \Phi\left(\frac{b}{\sqrt{1 + |a|^2}}\right),
\end{aligned}$$

where we have used the fact that $\langle X, a \rangle$ has the same distribution as $|a|Z$, and that the sum of independent Gaussian variables is again Gaussian.

Applying (4.3) (with $\rho = 1$) to the function f above, the left hand side becomes

$$\phi\left(\frac{b}{\sqrt{1 + |a|^2}}\right).$$

On the right hand side, $\mathcal{I}(f) = \phi(\langle x, a \rangle + b)$ and $|\nabla f| = |a|\phi(\langle x, a \rangle + b)$; hence the right hand side becomes

$$\sqrt{1 + |a|^2} \int_{\mathbb{R}^n} \phi(\langle a, x \rangle + b)\, d\gamma_n(x) = \sqrt{1 + |a|^2} \int_{\mathbb{R}} \phi(|a|z + b)\, d\gamma_1(z).$$

By expanding the definition of ϕ and completing the square in the exponent, this last integral may be computed explicitly, showing that this function f attains equality in (4.3).

4.5.2 From Bobkov to log-Sobolev

By applying Bobkov's inequality to functions of the form ϵf and then sending $\epsilon \to 0$, we may recover the log-Sobolev inequality with constant $\frac{1}{\rho}$. In order to show this, however, we first consider the asymptotics of the function $\mathcal{I}$ near zero.

Lemma 4.5.2. *For all $x > 0$,*

$$\frac{1}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} \leq 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}.$$

Proof. By definition, $1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\, dy$. Integrating by parts twice,

$$\int_x^\infty e^{-y^2/2}\, dy = \frac{e^{-y^2/2}}{y} - \int_x^\infty \frac{e^{-y^2/2}}{y^2}\, dy = \frac{e^{-y^2/2}}{y} - \frac{e^{-y^2/2}}{y^3} + \int_x^\infty \frac{e^{-y^2/2}}{y^4}\, dy.$$

The first equality above gives the claimed upper bound on $1 - \Phi(x)$, while the second equality gives the claimed lower bound. □

In what follows, we write $f(x) \sim g(x)$ if $\frac{f(x)}{g(x)} \to 1$ as x converges to some prescribed limit. According to Lemma 4.5.2 (and since $1 - \Phi(x) = \Phi(-x)$), $\Phi(x) \sim \frac{1}{|x|}\phi(x)$ as $x \to -\infty$. On the other hand, as $x \to 0$

$$\Phi\left(-(1+\epsilon)\sqrt{2\log(1/x)}\right) \ll x$$
$$\Phi\left(-(1-\epsilon)\sqrt{2\log(1/x)}\right) \gg x;$$

and hence $\Phi^{-1}(x) \sim -\sqrt{2\log(1/x)}$ as $x \to 0$.

For the asymptotics of $\mathcal{I}$, note that if $y = \Phi^{-1}(x)$ then

$$x = \Phi(y) \sim -\frac{1}{y}\phi(y) = -\frac{1}{\Phi^{-1}(x)}\mathcal{I}(x)$$

as $x \to 0$. In particular, $\mathcal{I}(x) \sim -x\Phi^{-1}(x) \sim x\sqrt{2\log(1/x)}$ as $x \to 0$.

Now let $f \in \mathcal{A}$ be bounded above and away from zero. Assume also that $\int_E f\, d\mu = 1$, and note that in order to establish the log-Sobolev inequality in general, it suffices to do it for such f. For small enough $\epsilon > 0$, ϵf takes values in $[0, 1]$ and we may apply (4.3). On the left hand side,

$$\sqrt{\rho}\mathcal{I}\left(\int_E \epsilon f\, d\mu\right) = \sqrt{\rho}\mathcal{I}(\epsilon) \sim \epsilon\sqrt{2\rho\log(1/\epsilon)}$$

as $\epsilon \to 0$. On the right hand side,

$$\begin{aligned}
&\int_E \sqrt{\rho \mathcal{I}^2(\epsilon f) + \Gamma(\epsilon f)}\, d\mu \\
&\sim \epsilon \int_E \sqrt{2\rho f^2 \log \frac{1}{\epsilon f} + \Gamma(f)}\, d\mu \\
&\sim \epsilon \int_E \sqrt{2\rho f^2 \log \frac{1}{\epsilon} + 2\rho f^2 \log \frac{1}{f} + \Gamma(f)}\, d\mu \\
&= \epsilon \int_E \sqrt{2\rho \log(1/\epsilon)} f + \frac{2\rho f^2 \log(1/f) + \Gamma(f)}{2\sqrt{2\rho \log(1/\epsilon)} f}\, d\mu + o\left(\frac{\epsilon}{\sqrt{\log(1/\epsilon)}}\right),
\end{aligned}$$

where the last line follows from $\sqrt{y+\delta} = \sqrt{y} + \delta/(2\sqrt{y}) + o(\delta)$. Comparing the asymptotics for the two sides of Bobkov's inequality, the terms of order $\epsilon\sqrt{\log(1/\epsilon)}$ cancel out. Examining the term of order $\epsilon \log^{-1/2}(1/\epsilon)$, we see that

$$2\rho \int_E f \log f\, d\mu \le \int_E \frac{\Gamma(f)}{f}\, d\mu,$$

which is equivalent to the log-Sobolev inequality with constant $\frac{1}{\rho}$. Recalling that the log-Sobolev inequality itself implies the Poincaré inequality, we have established a heirarchy of three inequalities: Bobkov implies log-Sobolev implies Poincaré.

4.5.3 Proof of Bobkov's inequality

We will prove Theorem 4.5.1 in much the same way as we proved the local Poincaré and local log-Sobolev inequalities. Note that we already know one direction: if the local Bobkov inequality holds then the local log-Sobolev inequality holds, and we already observed that the local log-Sobolev inequality implies $\mathrm{CD}(\rho, \infty)$. Hence, it only remains to prove the local Bobkov inequality under a $\mathrm{CD}(\rho, \infty)$ condition.

Lemma 4.5.3. *For any smooth function on $\mathbb{R}^2$, $f \in \mathcal{A}$, and $t > 0$,*

$$\begin{aligned}
&\frac{d}{ds} P_s \Psi(P_{t-s} f, \Gamma(P_{t-s} f)) = \\
&\qquad P_s \left[2\partial_2 \Psi \Gamma_2(g) + \partial_1^2 \Psi \Gamma(g) + 2\partial_1\partial_2 \Psi \Gamma(g, \Gamma(g)) + \partial_2^2 \Psi \Gamma(\Gamma(g))\right],
\end{aligned}$$

where $g = P_{t-s} f$ and Ψ stands for $\Psi(g, \Gamma(g))$.

Proof. We begin with

$$\begin{aligned}\frac{d}{ds}P_s\Psi(P_{t-s}f,\Gamma(P_{t-s}f)) &= P_s\left[L\Psi + \frac{d}{ds}\Psi(P_{t-s}f,\Gamma(P_{t-s}f))\right]\\ &= P_s\left[L\Psi - \partial_1\Psi Lg - 2\partial_2\Psi\Gamma(g,Lg)\right].\end{aligned}$$

The chain rule for L yields

$$L\Psi = \partial_1\Psi Lg + \partial_2\Psi L\Gamma(g) + \partial_1^2\Psi\Gamma(f) + 2\partial_1\partial_2\Psi\Gamma(g,\Gamma(g)) + \partial_2^2\Psi\Gamma(\Gamma(g)).$$

Gathering terms and recalling the definition of Γ_2 completes the proof. $\square$

Proof of Theorem 4.5.1. Let $\Psi(u,v) = \sqrt{\rho\mathcal{I}^2(u)+v}$. Then we may compute

$$\begin{aligned}\partial_1\Psi &= \frac{\rho\mathcal{I}\mathcal{I}'}{\Psi}\\ \partial_2\Psi &= \frac{1}{2\Psi}\\ \partial_1^2\Psi &= -\frac{(\rho\mathcal{I}\mathcal{I}')^2}{\Psi^3} + \rho\frac{(\mathcal{I}')^2-1}{\Psi}\\ \partial_1\partial_2\Psi &= -\frac{\rho\mathcal{I}\mathcal{I}'}{2\Psi^3}\\ \partial_2^2\Psi &= -\frac{1}{4\Psi^3}.\end{aligned}$$

Now fix t and define $\Lambda(s) = P_s\Psi(P_{t-s}f,\Gamma(P_{t-s}f))$. We will aim to show that Λ is non-decreasing in s; according to Lemma 4.5.3, it is enough to check that

$$(*) = \Psi^3\left[2\partial_2\Psi\Gamma_2(g) + \partial_1^2\Psi\Gamma(g) + 2\partial_1\partial_2\Psi\Gamma(g,\Gamma(g)) + \partial_2^2\Psi\Gamma(\Gamma(g))\right] \geq 0$$

pointwise for all $g \in \mathcal{A}$ taking values in $[0,1]$. Plugging in the derivatives of Ψ,

$$(*) = \Psi^2\Gamma_2(g) - (\rho\mathcal{I}\mathcal{I}')^2\Gamma(g) + \Psi^2\rho((\mathcal{I}')^2-1)\Gamma(g) - \rho\mathcal{I}\mathcal{I}'\Gamma(g,\Gamma(g)) - \frac{1}{4}\Gamma(\Gamma(g)),$$

where $\mathcal{I}$ and $\mathcal{I}'$ stand for $\mathcal{I}(g)$ and $\mathcal{I}'(g)$. Now recall that $\Psi^2 = \rho\mathcal{I} + \Gamma(g)$. Making this expansion and recalling also the reinforced $\mathrm{CD}(\rho,\infty)$ condition $\Gamma(g)\Gamma_2(g) - \rho\Gamma(g)^2 \geq \frac{1}{4}\Gamma(\Gamma(g))$, we have

$$(*) \geq \rho\mathcal{I}^2\Gamma_2(g) + \rho(\mathcal{I}')^2\Gamma(g)^2 - \rho^2\mathcal{I}^2\Gamma(g) - \rho\mathcal{I}\mathcal{I}'\Gamma(g,\Gamma(g)). \tag{4.4}$$

By applying the reinforced $\mathrm{CD}(\rho, \infty)$ condition and the Cauchy-Schwarz inequality, one can directly show this to be non-negative. Another argument uses the chain rule for Γ_2: since $(\Phi^{-1})' = 1/\mathcal{I}$ and $(\Phi^{-1})'' = \Phi^{-1}/\mathcal{I}^2 = -\mathcal{I}'/\mathcal{I}^2$, we have

$$\mathcal{I}^4\Gamma_2(\Phi^{-1} \circ g) = \mathcal{I}^2\Gamma_2(g) - \mathcal{I}\mathcal{I}'\Gamma(g, \Gamma(g)) + (\mathcal{I}')^2\Gamma(g)^2.$$

Comparing this to (4.4), we have

$$(*) \geq \rho\mathcal{I}^4\Gamma_2(\Phi^{-1} \circ g) - \rho^2\mathcal{I}^4\Gamma(\Phi^{-1} \circ g),$$

which is non-negative by the $\mathrm{CD}(\rho, \infty)$ condition. □

4.6 Exponential-type isoperimetry

Recall the exponential measure $d\mu(x) = e^{-x}dx$ on $[0, \infty)$. We may define a surface area μ^+ by applying our standard definition to the carré du champ $\Gamma(f, g) = f'g'$ on the algebra $\mathcal{A}$ of smooth functions whose derivatives vanish at infinity. Note that with this definition of surface area, a set $[a, b]$ with $a > 0$ has surface area $e^{-a} + e^{-b}$, but the set $[0, b]$ has surface area e^{-b}.

Proposition 4.6.1. *The isoperimetric profile of μ is $\mathcal{I}(a) = \min\{a, 1 - a\}$.*

Proof. For any $f \in \mathcal{A}$ and any $t \geq 0$, integration by parts yields

$$\begin{aligned}\int_t^\infty f' d\mu &= \int_t^\infty f'(x)e^{-x}\,dx \\ &= -e^{-t}f(t) + \int_0^\infty f(x)e^{-x}\,dx \\ &= \int_0^\infty f\,d\mu - f(0).\end{aligned}$$

Then, by the triangle inequality

$$\int_t^\infty \sqrt{\Gamma(f)}\,d\mu \geq \left|\int_0^\infty f\,d\mu - e^{-t}f(t)\right| \tag{4.5}$$

Now take any measurable $A \subset [0, \infty)$, and suppose that f_n is a sequence of functions in $\mathcal{A}$ converging to 1_A in $L^1(\mu)$. Without loss of generality, f_n takes values in $[0, 1]$.

We will consider two cases, depending on whether or not 0 belongs to the support of A. If it does, then for any $\epsilon > 0$, $\mu(A \cap [0, \epsilon)) > 0$. It follows that for sufficiently large n, there must be $x_n \in (0, \epsilon)$ such that $f_n(x_n) \geq 1 - \epsilon$.

By (4.5), if n is sufficiently large then

$$\begin{aligned}\int_0^\infty \sqrt{\Gamma(f_n)}\,d\mu &\geq \int_{x_n}^\infty \sqrt{\Gamma(f_n)}\,d\mu \\ &\geq e^{-\epsilon}(1-\epsilon) - \int_{x_n}^\infty f_n\,d\mu \\ &\geq e^{-\epsilon}(1-\epsilon) - \int_0^\infty f_n\,d\mu - \epsilon,\end{aligned}$$

where the last line follows because $x_n \leq \epsilon$ and $f_n \leq 1$. Taking $n \to \infty$ and then $\epsilon \to 0$, it follows that

$$\liminf_{n\to\infty} \int_0^\infty \sqrt{\Gamma(f_n)}\,d\mu \geq 1 - \mu(A),$$

and so $\mu^+(A) \geq 1 - \mu(A)$.

When 0 does not belong to the support of A, the argument is similar except that $f_n(x_n) \leq \epsilon$. In this case, we obtain

$$\liminf_{n\to\infty} \int_0^\infty \sqrt{\Gamma(f_n)}\,d\mu \geq \mu(A).$$

Putting these two directions together, $\mu^+(A) \geq \min\{\mu(A), 1-\mu(A)\}$. Hence, $\mathcal{I}(a) = \min\{a, 1-a\}$ is an isoperimetric function for μ.

To show that $\mathcal{I}$ is the isoperimetric profile, it suffices to give examples. In the case $a \leq \frac{1}{2}$, the set $[\log\frac{1}{a}, \infty)$ has measure a and surface area a; when $a > \frac{1}{2}$, the set $[0, \log\frac{1}{1-a})$ has measure a and surface area $1-a$. □

Based on Proposition 4.6.1, we make the following definition:

Definition 4.6.2 (Exponential-type isoperimetry)**.** *Say that μ and Γ satisfy an exponential-type isoperimetric inequality if there is a constant $c > 0$ such that $a \mapsto c\min\{a, 1-a\}$ is an isoperimetric function for μ, Γ.*

Note that by a change of variables in Proposition 4.6.1, $a \mapsto c\min\{a, 1-a\}$ is the isoperimetric function of the scaled exponential measure $d\mu(x) = ce^{-cx}\,dx$.

4.6.1 Exponential-type isoperimetry implies Poincaré

Proposition 4.6.3. *Let μ be a probability measure and Γ be a diffusion carré du champ. If $a \mapsto c\min\{a, 1-a\}$ is an isoperimetric function for μ and Γ then μ and Γ satisfy a Poincaré inequality with constant $\frac{4}{c^2}$.*

Note that the constant $\frac{4}{c^2}$ is best possible: the measure $d\mu(x) = e^{-x}\,dx$ has isoperimetric profile $a \mapsto \min\{a, 1-a\}$ and (by a change of variables in Proposition 2.6.1) Poincaré constant $\frac{4}{c^2}$.

Proof. Choose $f \in \mathcal{A}$ such that $\Gamma(f)$ is bounded (just to ensure that all integrals converge – this restriction can be removed at the end by taking limits). Let m be a median of f, meaning that $\mu(\{f \geq m\}) \geq \frac{1}{2}$ and $\mu(\{f \leq m\}) \geq \frac{1}{2}$. Let $g_+ = (f-m)_+$ and let $g_- = (f-m)_-$. Then $(f-m)^2 = g_+^2 + g_-^2$. By Fubini's theorem,

$$\begin{aligned}\int_E (f-m)^2\,d\mu &= \int_E g_+^2\,d\mu + \int_E g_-^2\,d\mu \\ &= \int_0^\infty \mu(\{g_+^2 \geq r\})\,dr + \int_0^\infty \mu(\{g_-^2 \geq r\})\,dr.\end{aligned}$$

Since $c\min\{a, 1-a\}$ is an isoperimetric function and $\mu(\{g_+^2 \geq r\}) \leq \frac{1}{2}$ for all $r \geq 0$, we have $c\mu(\{g_+^2 \geq r\}) \leq \mu^+(\{g_+^2 \geq r\})$ for all $r \geq 0$ (and similarly for g_-). Hence,

$$\begin{aligned}c\int_E (f-m)^2\,d\mu &\leq \int_{\mathbb{R}}^* \mu^+(\{g_+^2 \geq r\})\,dr + \int_{\mathbb{R}}^* \mu^+(\{g_-^2 \geq r\})\,dr \\ &\leq \int_E \sqrt{\Gamma(g_+^2)}\,d\mu + \int_E \sqrt{\Gamma(g_-^2)}\,d\mu, \end{aligned} \tag{4.6}$$

where the second inequality follows from the co-area formula (Proposition 4.3.3). By the diffusion property of Γ and the Cauchy-Schwarz inequality,

$$\int_E \sqrt{\Gamma(g_+^2)}\,d\mu = 2\int_E g_+\sqrt{\Gamma(g_+)}\,d\mu \leq 2\left(\int_E g_+^2\,d\mu \int_E \Gamma(g_+)\,d\mu\right)^{1/2}.$$

By the analogous inequality for g_- and the fact that (by the Cauchy-Schwarz inequality) $\sqrt{ab}+\sqrt{cd} \leq \sqrt{(a+c)(b+d)}$ for non-negative a, b, c, d, the right hand side of (4.6) is bounded by

$$2\sqrt{\int_E g_+^2 + g_-^2\,d\mu \int_E \Gamma(g_+) + \Gamma(g_-)\,d\mu} = 2\sqrt{\int_E (f-m)^2\,d\mu \int_E \Gamma(f)\,d\mu}$$

Going back to (4.6) and re-arranging, we have

$$\int_E (f-m)^2\,d\mu \leq \frac{4}{c^2}\int_E \Gamma(f)\,d\mu.$$

Since $\mathrm{Var}_\mu(f) \leq \int_E (f-a)^2\,d\mu$ for any $a \in \mathbb{R}$, it follows that μ and Γ satisfy a Poincaré inequality with constant $\frac{4}{c^2}$. $\square$

4.6.2 Milman's theorem: from integrability to isoperimetry

Definition 4.6.4. *Let μ be a probability measure and Γ be a diffusion carré du champ. Say that the non-increasing function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is a* concentration function *for μ, Γ if for every 1-Lipschitz function f, every median m of f, and every $r \geq 0$,*

$$\mu(\{f \geq m + r\}) \leq \alpha(r).$$

Note that the requirement that a concentration function be non-decreasing is without loss of generality, since we can always replace $\alpha(r)$ by $\inf\{\alpha(s) : s \leq r\}$.

When we were discussing exponential tails (for the Poincaré inequality) and exponential-squared tails (for the log-Sobolev inequality), it was more convenient to use the mean instead of the median. However, the mean and the median are basically the same as long as the concentration function decays quickly enough.

Exercise 4.6.1.

(a) Suppose that α is a concentration function for μ, Γ. Show that for every 1-Lipschitz function f and every $r \geq r_0$,

$$\mu\left(\left\{f \geq \int_E f\, d\mu + r\right\}\right) \leq \alpha(r - r_0),$$

where $r_0 = \int_0^\infty \alpha(r)\, dr$.

(b) Suppose that $\mu(\{f \geq \int_E f\, d\mu + r\}) \leq \alpha(r)$ for every 1-Lipschitz f and every $r \geq 0$. Show that

$$\beta(r) = \begin{cases} \alpha(r_0) & r \leq r_0 \\ \alpha(r - r_0) & r > r_0 \end{cases}$$

is a concentration function for μ, Γ, where

$$r_0 = \inf\{r \geq 0 : \alpha(r) \leq 1/2\}.$$

In particular, if $\alpha(r) = Ce^{-cr}$ or $\alpha(r) = Ce^{-cr^2}$ in Exercise 4.6.1 then we can swap the median with the mean in Definition 4.6.4 at the cost of changing the constants C and c.

Theorem 4.6.5 (Milman's theorem)**.** *Suppose that* (E, μ, Γ) *is a compact Markov triple satisfying* $\mathrm{CD}(0, \infty)$*, and suppose that* α *is a concentration function for* μ *and* Γ*. Then there exists some* $c > 0$ *depending only on* α *such that*

$$\mathcal{I}(a) = \begin{cases} \frac{c}{\alpha^{-1}(a/2)} a \log \frac{1}{a} & a \le \frac{1}{2} \\ \frac{c}{\alpha^{-1}((1-a)/2)} (1-a) \log \frac{1}{1-a} & a > \frac{1}{2} \end{cases}$$

is an isoperimetric function for μ *and* Γ*.*

By plugging in $\alpha(r) = Ce^{-cr}$ and $\alpha(r) = Ce^{-cr^2}$, we see that under a $\mathrm{CD}(0, \infty)$ condition, exponential-type concentration implies an exponential-type isoperimetric inequality while exponential-squared-type concentration implies a Gaussian-type isoperimetric inequality.

Before beginning the proof of Theorem 4.6.5, recall the local reverse log-Sobolev inequality of Proposition 3.3.1. In the case that f takes values in $(0, 1]$, we have $f \log f \le 0$ and hence

$$\Gamma(P_t f) \le \frac{1}{t}(P_t f)^2 \log \frac{1}{P_t f}. \tag{4.7}$$

Taking $\psi(x) = \sqrt{\log(1/x)}$, we have $\psi'(x) = (2x\sqrt{\log(1/x)})^{-1}$ and so the chain rule for Γ implies that

$$\Gamma\left(\sqrt{\log \frac{1}{P_t f}}\right) \le \frac{1}{4t}.$$

We summarize this as a lemma:

Lemma 4.6.6. *Under a* $\mathrm{CD}(0, \infty)$ *condition, if* $f \in \mathcal{A}$ *takes values in* $(0, 1]$ *then* $\sqrt{\log(1/P_t f)}$ *is* $\frac{1}{2\sqrt{t}}$*-Lipschitz.*

Under the additional assumption of a concentration function, Lemma 4.6.6 may be used to obtain a tail bound on $P_t f$: let m be a median of $P_t f$; since $\sqrt{\log(1/x)}$ is monotonic in x, $\sqrt{\log(1/m)}$ is a median of $\sqrt{\log(1/P_t f)}$. Then

$$\begin{aligned} \alpha(r) &\ge \mu\left(\left\{\sqrt{\log(1/P_t f)} \le \sqrt{\log(1/m)} - \frac{r}{2\sqrt{t}}\right\}\right) \\ &= \mu\left(\left\{P_t f \ge m e^{r^2/(4t)}\right\}\right). \end{aligned} \tag{4.8}$$

Recall from our proof of the log-Sobolev and reverse log-Sobolev inequalities that

$$P_t(f \log f) - P_t f \log P_t f = \int_0^t P_{t-s} \frac{\Gamma(P_s f)}{P_s f}\, ds.$$

Now, (4.7) implies that $\sqrt{\Gamma(P_s f)} \le P_s f \sqrt{s^{-1}\log(1/\epsilon)}$. Hence,

$$\begin{aligned} P_t f \log \frac{1}{P_t f} - P_t\left(f \log \frac{1}{f}\right) &\le \int_0^t P_{t-s}\sqrt{s^{-1}\log(1/\epsilon)\Gamma(P_s f)}\, ds \\ &\le \int_0^t P_t \sqrt{s^{-1}\log(1/\epsilon)\Gamma(f)}\, ds, \end{aligned}$$

where the second inequality follows from the gradient bound under $\mathrm{CD}(0,\infty)$. Rearranging this and integrating over E,

$$\int_E P_t f \log \frac{1}{P_t f}\, d\mu - \int_E f \log \frac{1}{f}\, d\mu \le 2\sqrt{t \log \frac{1}{\epsilon}} \int_E \sqrt{\Gamma(f)}\, d\mu.$$

This is going in the right direction for an isoperimetric inequality, since we are lower-bounding $\int_E \sqrt{\Gamma(f)}\, d\mu$ in terms of an integral of f. Moreover, if f is approximating an indicator function then the second term on the left hand side is very small; the remaining step is to relate $\int P_t f \log(1/P_t f)\, d\mu$ to $a = \int f\, d\mu$. For this, take some $\delta > 0$. Then

$$\int_E P_t \log \frac{1}{P_t f}\, d\mu \ge \log \frac{1}{\delta} \int_{\{P_t f < \delta\}} P_t f\, d\mu \ge \log \frac{1}{\delta}\left[a - \mu(\{P_t f \ge \delta\})\right].$$

If $\delta \ge m e^{t^2/(4t)}$ then we may apply (4.8) to bount $\mu(\{P_t f \ge \delta\})$ by $\alpha(r)$. Combining the argument so far, we have the following lemma.

Lemma 4.6.7. *Suppose that $f \in \mathcal{A}$ takes values in $[\epsilon, 1]$. Then for any $r, t > 0$ and any $\delta \ge m e^{r^2/(4t)}$ (where m is a median of $P_t f$),*

$$\log \frac{1}{\delta}\left(\int_E f\, d\mu - \alpha(r)\right) - \int_E f \log \frac{1}{f}\, d\mu \le 2\sqrt{t \log \frac{1}{\epsilon}} \int_E \sqrt{\Gamma(f)}\, d\mu. \tag{4.9}$$

To turn this into an isoperimetric inequality, we will apply Lemma 4.6.7 to a sequence of functions approaching an indicator function, and we will need to specify the parameters r, t, ϵ, and δ. We will consider two cases, depending on how large the measure of the set is.

First, let $A \subset E$ be a measurable set. Set, $\mu(A) = a$ and assume first that $0 \le a \le \eta$, where η is some constant to be determined. Let $f_n \in \mathcal{A}$ be a sequence converging to 1_A in $L^1(\mu)$, and suppose without loss of generality that f_n take values in $[0, 1]$. Now take $\epsilon = a^2$ and let $g_n = \epsilon \wedge f_n$. Then $\int_E g_n\, d\mu \ge \int_E f_n\, d\mu$ and $\int_E \sqrt{\Gamma(g_n)}\, d\mu \le \int_E \sqrt{\Gamma(f_n)}\, d\mu$ for every n. Choose $r = \alpha^{-1}(a/2)$, and let t solve $r^2 = t\log(1/\epsilon) = 2t\log(1/a)$. Set

$$\delta = 2a e^{r^2/(4t)} = 2a e^{\frac{1}{2}\log\frac{1}{a}} = 2\sqrt{a},$$

which satisfies the conditions of Lemma 4.6.7 because (by Markov's inequality) $m \leq \int P_t f \, d\mu = 2a$. Plugging all these parameters back into (4.9),

$$\log \frac{1}{2\sqrt{a}} \left(\int_E f_n \, d\mu - \frac{a}{2} \right) - \int_E g_n \log \frac{1}{g_n} \, d\mu \leq 2\alpha^{-1}(a/2) \int_E \sqrt{\Gamma(f_n)} \, d\mu. \tag{4.10}$$

Now, $g_n \to a^2 \wedge 1_A$ in $L^1(\mu)$ and hence also pointwise μ-almost everywhere. Since $g_n \log \frac{1}{g_n}$ is also bounded, it follows that

$$\limsup_{n \to \infty} \int_E g_n \log \frac{1}{g_n} \, d\mu \leq a^2 \log \frac{1}{a^2}.$$

Since $\int_E f_n \, d\mu \to a$, the left hand side of (4.10) is of order $a \log \frac{1}{a}$ for small a. In particular, there are constants $c, \eta > 0$ such that for all A with $\mu(A) = a \leq \eta$,

$$\mu^+(A) \geq \frac{c}{\alpha^{-1}(a/2)} a \log \frac{1}{a}. \tag{4.11}$$

This proves Theorem 4.6.5 in the case that $\mu(A)$ is small.

For $\eta \leq \mu(A) \leq \frac{1}{2}$, we apply essentially the same argument but with some different parameters. First, choose $r = \alpha^{-1}(\eta/2)$, so that $\alpha(r) = \eta/2 \leq a/2$. Next, note that (4.7) and the fact that $x^2 \log(1/x) \leq 2$ for $x \in [0, 1]$ together imply that $P_t f$ is $\sqrt{2/t}$-Lipschitz. This may be used to find a better bound on $a = \int_E f \, d\mu$ and m, the median of $P_t f$. Indeed,

$$\begin{aligned} a = \int_E P_t f \, d\mu &\leq m + \int_E (f - m)_+ \, d\mu \\ &= m + \int_0^1 \mu(\{f \geq m + r\}) \, dr \\ &\leq m + \alpha(\sqrt{t/2}). \end{aligned}$$

For t large enough (depending on α and η), we may ensure that $a \leq \frac{11}{10} m$. By increasing t if necessary, we may also ensure that $e^{r^2/(4t)} \leq \frac{12}{11}$. This implies that $\delta = \frac{3}{5} \geq \frac{6a}{5} \geq \frac{11}{10} a e^{r^2/(4t)}$ satisfies the conditions of Lemma 4.6.7. Finally, choose ϵ so that $\epsilon \log(1/\epsilon) \leq \frac{\eta \log(5/3)}{4}$. Then (4.9) implies that

$$\frac{a \log \frac{5}{3}}{2} - \int_E f \log \frac{1}{f} \, d\mu \leq 2\sqrt{t \log \frac{1}{\epsilon}} \int_E \sqrt{\Gamma(f)}$$

whenever f is a function of mean $a \in [\eta, 1/2]$ taking values in $[\epsilon, 1]$. Note that t and ϵ depend only on η and α. By the same argument for the isoperimetric inequality above, we see that

$$\mu^+(A) \geq c\mu(A)$$

for every A with $\eta \leq \mu(A) \leq \frac{1}{2}$, where c is a constant depending only on η and α. Combining this with (4.11) proves Theorem 4.6.5.

Chapter 5

Riemannian manifolds

5.1 Riemannian manifolds

This section contains a brief introduction to Riemannian manifolds, in which we introduce just enough in order to show how to put compact Markov triples on compact, connected Riemannian manifolds. Readers who have not yet seen the main concepts may prefer to consult a real book (for example, [Lee06]) on the topic. Those who are already familiar with Riemannian manifolds will probably only care about Section 5.1.9 and Proposition 5.1.29.

Definition 5.1.1. *A* smooth n-manifold *is a topological space M together with a collection of open sets U_i and maps $\psi_i : U_i \to \mathbb{R}^n$ satisfying*

(a) M is a second countable Hausdorff space;

(b) each ψ_i is a homeomorphism onto its image;

(c) every $p \in M$ belongs to some U_i; and

(d) for every U_i and U_j that have a non-empty intersection, the map

$$\psi_j \circ \psi_i^{-1} : \psi_i(U_i \cap U_j) \to \mathbb{R}^n$$

belongs to C^∞.

The maps ψ_i are called charts *and the collection of charts is called an* atlas*.*

The simplest example of a smooth n-manifold is $\mathbb{R}^n$ itself, with the only chart being the identity map (whose domain is all of $\mathbb{R}^n$). Similarly, an open subset of $\mathbb{R}^n$ is an n-manifold.

For a slightly less trivial example, consider $M = S^n = \{x \in \mathbb{R}^{n+1} : \sum_i x_i^2 = 1\}$. We may equip this with charts defined on open hemispheres. For example, let $U_i^+ = \{p \in M : p_i > 0\}$ and let $U_i^- = \{p \in M : p_i < 0\}$. Then let $\psi_i^\pm : U_i^\pm$ be the map that leaves out the ith coordinate. It is easy to check that this defines an atlas for M.

Usually, when we do a local computation (such as differentiating a function) on a manifold it is more convenient to fix a chart ψ and do the computation on the image of ψ (which is a subset of $\mathbb{R}^n$). However, it would be extremely inconvenient to write ψ all the time; therefore, we will usually do computations on the image of a chart but without explicitly writing down the chart. We adopt the convention that p denotes a point on the manifold M and x (or y) denotes its image under some chart. For example, if f is a function on M then we will write $f(x)$ when we really mean $f(\psi^{-1}(x))$ for some chart ψ. We say that x are *local coordinates* on M. When we need to deal with two intersecting charts (say ψ_1 and ψ_2) simultaneously, we will let x be the image under ψ_1 and y be the image under ψ_2. We will then write $y(x)$ for the map $\psi_2 \circ \psi_1^{-1}$ that changes from x coordinates to y coordinates.

Definition 5.1.2. *Let M be a smooth m-manifold and N be a smooth n-manifold. A map $f : M \to N$ is* smooth *if it is smooth when viewed in local coordinates – that is, when the map $\psi_N \circ f \circ \psi_M^{-1}$ is smooth whenever ψ_M is a chart for M and ψ_N is a chart for N.*

We write $C^\infty(M, N)$ for the set of smooth maps from $M \to N$. We will use the abbreviation $C^\infty(M) = C^\infty(M, \mathbb{R})$.

If f is smooth, bijective, and has a smooth inverse then we call it a diffeomorphism.

One of the points of a manifold is that many things that are locally true in Euclidean space are also true on manifolds. In particular, the following lemma is very useful for imitating Euclidean arguments on manifolds.

Lemma 5.1.3 (Extension lemma)**.** *If $A \subset M$ is a closed set and $f : A \to \mathbb{R}$ is a smooth function (meaning that there exists an open set containing M on which f can be smoothly extended), then for every open $U \supset A$ there is a smooth extension of f to all of M such that $f = 0$ outside of U.*

In particular, Lemma 5.1.3 implies that manifolds have smooth “bump” functions: for every closed A contained in an open U, there is a smooth function that is 1 on A and 0 outside of U.

5.1.1 Tangent spaces

Definition 5.1.4. *Given a point $p \in M$, a* tangent vector *to M at p is a map $X : C^\infty(M) \to \mathbb{R}$ such that for every $f, g \in C^\infty(M)$ and $a \in \mathbb{R}$,*

(a) $X(f+g) = X(f) + X(g)$ and $X(af) = aX(f)$, and

(b) $X(fg) = g(p)X(f) + f(p)X(g)$.

The set of all tangent vectors to M at p is denoted T_pM.

The definition of tangent vectors may seem abstract, but it turns out that they can be used in a fairly concrete way. The following exercise should help to make some sense of them.

Exercise 5.1.1. *Let M be a smooth n-manifold. Let f and g be smooth functions and take $X \in T_pM$.*

(a) If f is constant then $X(f) = 0$.

(b) If f and g agree on some open set containing p then $X(f) = X(g)$. (Hint: use a bump function.)

(c) If $f(p) = g(p) = 0$ then $X(fg) = 0$.

(d) If f has gradient zero at p in local coordinates (i.e. $f \circ \psi^{-1}$ has gradient zero for some chart ψ around p) then $X(f) = 0$. (Hint: use the previous point and a Taylor expansion of $f \circ \psi^{-1}$.)

(e) T_pM is an n-dimensional vector space. Given local coordinates x around p, T_pM is spanned by

$$\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^n}\right|_p,$$

where $\left.\frac{\partial}{\partial x^i}\right|_p(f)$ is defined in the usual calculus way. (This time I leave it to you to make sense of this statement in terms of a chart.)

In particular, the last part of Exercise 5.1.1 implies that for every tangent vector X at p (and every choice of local coordinates), there are real numbers $X^1, \dots, X^n$ such that

$$X = \sum_{i=1}^n X^i \left.\frac{\partial}{\partial x^i}\right|_p.$$

We will adopt the *Einstein summation convention*, which says that whenever an index i appears in both an "upper" position and a "lower" position then we implicitly sum over it. In other words, we abbreviate the equation above by

$$X = X^i \left.\frac{\partial}{\partial x^i}\right|_p .$$

When there is no ambiguity about the choice of coordinates x or the point p, we will abbreviate $\left.\frac{\partial}{\partial x^i}\right|_p$ by ∂_i.

Note that the representation of X in terms of $X^1, \ldots, X^n$ depends on the choice of local coordinates. However, there is a simple formula for changing coordinates.

Exercise 5.1.2. *Let x and $\tilde{x}$ be two sets of local coordinates around the same point p. Show that*

$$\left.\frac{\partial}{\partial x^i}\right|_p = \left.\frac{\partial \tilde{x}^j}{\partial x^i}\right|_p \left.\frac{\partial}{\partial \tilde{x}^j}\right|_p .$$

(The Einstein convention considers a upper index in the denominator as a lower index, so there is an implicit summation on the right hand side above.) As a consequence, if $X^1, \ldots, X^n$ represent X in the x coordinates and $\tilde{X}^1, \ldots, \tilde{X}^n$ represent X in the $\tilde{x}$ coordinates then

$$\tilde{X}^j = \left.\frac{\partial \tilde{x}^j}{\partial x^i}\right|_p X^i.$$

In other words, the Jacobian matrix of the coordinate change map represents the change of basis for T_pM.

Note that since T_pM is a finite-dimensional vector space, it comes with a natural topology.

5.1.2 The tangent bundle

Definition 5.1.5. *The* tangent bundle *TM of M is the disjoint union $\bigsqcup_{p\in M} T_pM$.*

If M is an n-manifold then we may define a topology and an atlas for TM such that TM is a $2n$-manifold. Indeed, if $(x^1, \ldots, x^n)$ are local coordinates for M on the neighborhood $U \subset M$, then we may define a map $\bigsqcup_{p\in U} T_pM \to \mathbb{R}^n \times \mathbb{R}^n$ by

$$\left(v^i \left.\frac{\partial}{\partial x_i}\right|_p \right) \mapsto (x^1(p), \ldots, x^n(p), v^1, \ldots, v^n).$$

Exercise 5.1.3. *Consider the topology on TM generated by the maps above. Show that these maps and this topology turn TM into a smooth $2n$-manifold.*

Definition 5.1.6. *A* vector field *on M is a smooth map $X : M \to TM$ (which we write as $p \mapsto X_p$) with the property that $X_p \in T_pM$ for every $p \in M$. The set of vector fields on M is denoted $\mathcal{T}M$.*

Note that for every choice of local coordinates around a point p, we may write X_p in terms of the standard basis of T_pM. This means that locally there exist real-valued functions $X^1, \ldots, X^n$ such that

$$X = X^i \partial_i.$$

Also, if $f : M \to \mathbb{R}$ is any smooth function then we can define another function $Xf : M \to \mathbb{R}$ by $(Xf)(p) = X_p(f)$.

Exercise 5.1.4.

(a) The coordinate functions $p \mapsto X_p^i$ are (locally) smooth functions from M to $\mathbb{R}$.

(b) The function $Xf : M \to \mathbb{R}$ is smooth.

5.1.3 The cotangent bundle

For any $p \in M$, T_pM is a finite-dimensional vector space. Therefore we may also consider its dual space (i.e. the set of linear functionals on T_pM), which be denote by T_p^*M and call the *cotangent space*. With respect to any local coordinates, ∂_i form a basis for T_pM; hence, it induces a dual basis for T_p^*M. We denote this dual basis by $dx^1, \ldots, dx^n$. That is, dx^i is defined by

$$dx^i(\partial_j) = \delta_i^j.$$

Definition 5.1.7. *The* cotangent bundle T^*M *of M is the disjoint union $\bigsqcup_{p\in M} T_p^*M$.*

We may endow the cotangent bundle with a smooth manifold structure in the same way we did with the tangent bundle.

Definition 5.1.8. *A* covector field *on M is a smooth map $\omega : M \to T^*M$ (which we write as $p \mapsto \omega_p$) with the property that $\omega_p \in T_p^*M$ for every $p \in M$. The set of covector fields on M is denoted $\mathcal{T}^*M$.*

As we did for vector fields, we may of course write ω locally in terms of the basis coordinate covector basis as $\omega = \omega_i dx^i$ for smooth functions ω_i. There is also a change of coordinates formula for covectors: if $\omega = \omega_i dx^i = \omega_j d\tilde{x}^j$ then

$$\omega_i = \frac{\partial \tilde{x}^j}{\partial x^i}\tilde{\omega}_j$$

Note that this change of coordinates formula is different from the one for vectors. This is important, since it gives us a coordinate-independent way of writing the derivative of a function.

Definition 5.1.9. *For a smooth function $f : M \to \mathbb{R}$, its* differential *df is the covector field defined by $df(X) = X(f)$.*

Exercise 5.1.5.

(a) In local coordinates, $df = \frac{\partial f}{\partial x^i}dx^i$. In particular, $\frac{\partial f}{\partial x^i}dx^i$ is invariant under changes of coordinates.

(b) Show that $\frac{\partial f}{\partial x^i}\partial_i$ is not invariant under changes of coordinates.

5.1.4 Vector bundles

The tangent and cotangent bundles share a structure that is useful to generalize: they are manifolds built by "glueing" together in a smooth way a vector space at every point of the manifold M. This construction may be made in a more general way, although we will not give the full definition here. (In all the cases we consider, the construction will be analogous to the one for the tangent bundle.) When E is built from M by gluing together k-dimensional vector spaces at each point of M, then we call E a k-dimensional vector bundle. We equip E with the projection map $\pi : E \to M$ (so that $\pi^{-1}(p)$ is the k-dimensional vector space that we attached to p).

The natural generalization of vector fields and covector fields are called sections.

Definition 5.1.10. *Let E be a vector bundle over M with projection map π. A* section *of E is a smooth function $F : M \to E$ such that $F(p) \in \pi^{-1}(p)$ for all $p \in M$.*

Besides the tangent bundle and the cotangent bundle that we have already seen, one can easily visualize the *normal bundle* of a manifold embedded in $\mathbb{R}^n$; here, we attach to every point the space of vectors that point in a normal direction. A section of the normal bundle (called a normal vector field) consists of a choice of a normal vector at every point.

5.1.5 Tensors and tensor bundles

Definition 5.1.11. *Let V be a finite-dimensional vector space. A for k and ℓ non-negative integers, a (k, ℓ)-tensor on V is a multilinear map*

$$V^k \times (V^*)^\ell \to \mathbb{R}.$$

We write $T^k_\ell(V)$ for the set of all (k, ℓ)-tensors on V; when either k or ℓ is zero, we will omit it.

Note that there is a natural vector space structure on $T^k_\ell(V)$. Also, if we fix a basis $e_1, \dots, e_n$ of V (and let $\phi^1, \dots, \phi^n$) be the dual basis) then a (k, ℓ)-tensor F may be represented by a $(k+\ell)$-tuple of numbers: define

$$F^{i_1,\dots,i_k}_{j_1,\dots,j_\ell} = F(e_{i_1}, \dots e_{i_k}, \phi^{j_1}, \dots, \phi^{j_\ell}).$$

Since F was assumed to be multilinear, this completely defines F.

Definition 5.1.12. *If F is a (k, ℓ)-tensor and G is a (k', ℓ')-tensor then the* tensor product *$F \otimes G$ is the $(k+k', \ell+\ell')$-tensor defined by*

$$\begin{aligned}(F \otimes G)&(x_1, \dots, x_{k+k'}, y^1, \dots, y^{\ell+\ell'}) \\ &= F(x_1, \dots, x_k, y^1, \dots, y^\ell) G(x_{k+1}, \dots, x_{k+k'}, y^{\ell+1}, y^{\ell+\ell'}).\end{aligned}$$

Some spaces of tensors are already familiar: $T^0_0(V)$ is just $\mathbb{R}$, $T^1(V) = V^*$, and $T_1(V)$ is canonically isomorphic to V. Moreover, $T^1_1(V)$ is canonically isomorphic to the space of linear maps from V to itself.

Exercise 5.1.6. *Let $e_1, \dots, e_n$ be a basis for V and take $\phi^1, \dots, \phi^n$ to be the dual basis for V^*. Show that*

$$\{e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes \phi^{j_1} \otimes \cdot \otimes \phi^{j_\ell} : i_1, \dots, i_k, j_1, \dots, j_\ell = 1, \dots, n\}$$

is a basis for $T^k_\ell(V)$. In particular, $T^k_\ell(V)$ has dimension $n^{k+\ell}$.

Definition 5.1.13. *Let $e_1, \dots, e_n$ be a basis for V and let $\phi^1, \dots, \phi^n$ be its dual basis for V^*. The* trace, *or* contraction *of a (k, ℓ)-tensor F is the $(k-1, \ell-1)$-tensor given by*

$$(\operatorname{tr} F)(X_1, \dots, X_{k-1}, \omega^1, \dots, \omega^{\ell-1}) = F(e_i, X_1, \dots, X_{k-1}, \phi^i, \omega^1, \dots, \omega^{\ell-1}).$$

Note that the trace of a $(1,1)$-tensor is the usual matrix trace.

Exercise 5.1.7. *Show that the trace does not depend on the choice of basis $e_1, \dots, e_n$.*

Definition 5.1.14. *Given an n-manifold M, the* tensor bundle $T^k_\ell(M)$ *is the vector bundle given by*

$$\bigsqcup_{p\in M} T^k_\ell(T_pM).$$

A section of a tensor bundle is called a tensor field*, and the set of (k,ℓ)-tensor fields is denoted $\mathcal{T}^k_\ell(M)$.*

Note that $T_1(M)$ is just the tangent bundle, while $T^1(M)$ is the cotangent bundle. Also, $T^0_0(M) = C^\infty(M)$.

A tensor field $F \in \mathcal{T}^k_\ell(M)$ naturally induces a multilinear map $(\mathcal{T}M)^k \times (\mathcal{T}^*M)^\ell \to C^\infty(M)$, defined by simply doing everything pointwise. It is a fundamental property of tensor fields that this map is not only linear over $\mathbb{R}$, but also over $C^\infty(M)$. Specializing to the case $F \in \mathcal{T}^1_1(M)$ for brevity, this means that for every $X \in \mathcal{T}M$, every $Y \in \mathcal{T}^*M$, and every $f, g \in C^\infty(M)$,

$$F(fX, gY) = fgF(X,Y).$$

In fact, this property characterizes tensor fields.

Lemma 5.1.15 (Tensor field characterization)**.** *A function $F : (\mathcal{T}M)^k \times (\mathcal{T}^*M)^\ell \to C^\infty(M)$ belongs to $\mathcal{T}^k_\ell(M)$ if and only if it is $C^\infty(M)$-linear in every coordinate.*

5.1.6 Riemannian metrics

Definition 5.1.16 (Riemannian metric)**.** *A Riemannian metric g on a manifold M is a $(2,0)$-tensor satisfying*

(a) symmetry: for every $X, Y \in \mathcal{T}M$, $g(X,Y) = g(Y,X)$; and

(b) positivity: for every $p \in M$ and $X \in T_pM$, if $X \neq 0$ then $g_p(X,X) > 0$.

In other words, a Riemannian metric specifies an inner product on every tangent space T_pM. We will usually write $\langle X, Y\rangle$ instead of $g(X,Y)$. As with any $(2,0)$-tensor, we may express a Riemannian metric in local coordinates as

$$g = g_{ij}dx^i \otimes dx^j.$$

The symmetry condition ensures that $g_{ij} = g_{ji}$, while the positivity condition ensures that the $n \times n$ matrix (g_{ij}) is positive definite at every $p \in M$.

Exercise 5.1.8. *Write down the change of coordinates formula for g_{ij}.*

An inner product on T_pM induces an isomorphism between T_pM and its dual T_p^*M by mapping $X \in T_pM$ to the linear functional $Y \mapsto \langle X, Y\rangle$. We denote this map by $X \mapsto X^\flat$ and its inverse by $\omega \mapsto \omega^\sharp$. In coordinates, $X^\flat = X_j dx^j$ where $X_j = g_{ij}X^i$ and $\omega^\sharp = \omega^j \partial_j$ where $\omega^j = g^{ij}\omega_i$ and g^{ij} is the inverse matrix of g_{ij}.

We will extend the notation $\langle , \rangle$ to covectors using this isomorphism:

$$\langle \omega, \phi\rangle = \langle \omega^\sharp, \phi^\sharp\rangle = \omega_i \phi_j g^{ij}.$$

5.1.7 Connections

A connection on a manifold is essentially a local correspondence between tangent spaces; for example, it allows one to say what it means for two vectors in different tangent spaces to be "parallel." Formally, we define a connection to be something that allows us to differentiate vector fields.

Definition 5.1.17 (Connection)**.** *A* Riemannian connection *on a Riemannian manifold is a map $\nabla : \mathcal{T}M \times \mathcal{T}M \to \mathcal{T}M$, written $(X, Y) \mapsto \nabla_X Y$, that satisfies*

(a) $\nabla_X Y$ is $C^\infty(M)$-linear in X, i.e.

$$\nabla_{fX_1+gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y \quad \text{for } f, g \in C^\infty(M);$$

(b) $\nabla_X Y$ is $\mathbb{R}$-linear in Y, i.e.

$$\nabla_X (aY_1 + bY_2) = a\nabla_X Y_1 + b\nabla_X Y_2 \quad \text{for } a, b \in \mathbb{R};$$

(c) for $f \in C^\infty(M)$, we have the product rule

$$\nabla_X (fY) = f\nabla_X Y + (Xf)Y;$$

(d) the connection is compatible with the metric*, meaning that*

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle;$$

(e) the connection is torsion-free*, meaning that*

$$(\nabla_X Y - \nabla_Y X)f = X(Yf) - Y(Xf).$$

We interpret $\nabla_X Y$ as the derivative of Y along the directions given by X.

Exercise 5.1.9. *On $\mathbb{R}^n$ with the Euclidean metric, show that $\nabla_X Y = (XY^i)\partial_i$ defines a Riemannian connection.*

Theorem 5.1.18 (Levi-Civita theorem)**.** *Every Riemannian manifold admits a unique Riemannian connection.*

It turns out that knowing how to differentiate functions and vector fields gives us a canonical way to differentiate all tensor fields:

- For $f \in \mathcal{T}_0^0(M) = C^\infty(M)$, define $\nabla_X f = Xf$.
- For $Y \in \mathcal{T}_1(M) = \mathcal{T}(M)$, $\nabla_X Y$ is defined by Theorem 5.1.18.
- For $\omega \in \mathcal{T}^1(M)$, $\nabla_X \omega$ is defined by the equation

$$X(\omega(Y)) = \nabla_X(\omega(Y)) = (\nabla_X \omega)(Y) + \omega(\nabla_X Y).$$

- For $F \in \mathcal{T}_\ell^k(M)$, $\nabla_X F$ is defined by the equation

$$\begin{aligned} X(F(Y_1, \ldots, Y_k, \omega^1, \ldots, \omega^\ell)) = {} & (\nabla_X F)(Y_1, \ldots, Y_k, \omega^1, \ldots, \omega^\ell) \\ & + \sum_i F(\ldots, \nabla_X Y_i, \ldots) \\ & + \sum_j F(\ldots, \nabla_X \omega^j, \ldots). \end{aligned}$$

Exercise 5.1.10. *The map ∇ defined above satisfies the following properties:*

(a) $\nabla_X(Y^\flat) = (\nabla_X Y)^\flat$ *and* $\nabla_X(\omega^\sharp) = (\nabla_X \omega)^\sharp$.

(b) $\nabla_X(F \otimes G) = F \otimes (\nabla_X G) + (\nabla_X F) \otimes G$.

(c) $\nabla_X \operatorname{tr}(F) = \operatorname{tr}(\nabla_X F)$.

(d) *For $F \in \mathcal{T}_\ell^k(M)$, the map $\nabla F : (\mathcal{T}M)^{k+1} \times (\mathcal{T}^*\mathcal{M})^\ell \to \mathcal{C}^\infty(\mathcal{M})$ defined by*

$$(\nabla F)(Y_1, \ldots, Y_k, X, \omega^1, \ldots, \omega^\ell) = (\nabla_X F)(Y_1, \ldots, Y_k, \omega^1, \ldots, \omega^\ell)$$

is a $(k+1, \ell)$-tensor field.

(e) For $f \in C^\infty(M)$, $\nabla f = df$.

Definition 5.1.19 (Hessian)**.** *The* Hessian *of $f \in C^\infty(M)$ is the $(2,0)$-tensor $\nabla^2 f = \nabla(\nabla f)$.*

Note that $(\nabla^2 f)(X,Y)$ is *not* the same as $\nabla_X(\nabla_Y f)$.

Exercise 5.1.11. *Show that $(\nabla^2 f)(X,Y) = Y(Xf) - (\nabla_Y X)(f)$. Conclude that $\nabla^2 f$ is* symmetric*, in the sense that $(\nabla^2 f)(X,Y) = (\nabla^2 f)(Y,X)$.*

So far, we defined the trace only for tensors that take both a vector and a covector. However, using the isomorphism induced by the Riemannian metric between vectors and covectors, we may also define the trace of other tensors (although we will stick to the $(2,0)$ case because the notation is simpler).

Definition 5.1.20 (Trace)**.** *If F is a $(2,0)$-tensor, written as $F_{ij}dx^i dx^j$ in local coordinates, then its trace is $\operatorname{tr} F = F^{ij} g_{ji}$.*

Exercise 5.1.12. *Show that the trace defined above is invariant under changes of coordinates.*

Definition 5.1.21 (Laplace-Beltrami operator)**.** *The Laplace-Beltrami operator $\Delta : C^\infty(M) \to C^\infty(M)$ is defined by*

$$\Delta f = \operatorname{tr}(\nabla^2 f).$$

Exercise 5.1.13. *Show that the Laplace-Beltrami operator is given in local coordinates by*

$$\Delta f = \frac{1}{\sqrt{|\det g|}} \partial_i \left(\sqrt{|\det g|} \cdot g^{ij} \partial_j f \right).$$

(Hint: first, figure out how to write $\nabla_{\partial_i} \partial_j$ in terms of g_{ij} and its derivatives.)

5.1.8 Integration on a manifold

There are two different ways to define integration on manifolds. The more common one defines a "volume form," which depends on orientation and requires an oriented manifold. We will instead define a "volume element," which is more natural for the applications we have in mind.

We define the integral in two steps: first, if $f \in C^\infty(M)$ is supported on a single chart U then we define

$$\int_M f\, dV = \int_U f \sqrt{|\det g|}\, dx.$$

Exercise 5.1.14. *If $f \in C^\infty(M)$ is supported on the intersection of two charts, show that the definition above does not depend on the chart.*

Next, if $f \in C^\infty(M)$ has compact support then we write $f = \sum_{i=1}^k f_i$ where each f_i is smooth and supported on a single chart (the fact that this can be done is not completely trivial, but it can be done by using a standard tool known as a *partition of unity*). Then we define

$$\int_M f \, dV = \sum_{i=1}^k \int_M f_i \, dV.$$

Of course, this may be extended beyond smooth functions of compact support by taking limits.

Exercise 5.1.15. *Show that the definition above does not depend on the choice of f_i.*

The usual integration by parts formulas transfer over from Euclidean space to manifolds:

Theorem 5.1.22 (Green's theorem)**.** *For any compactly supported $f, g \in C^\infty(M)$,*

$$\int_M f \Delta g \, dV = -\int_M \langle \nabla f, \nabla g \rangle \, dV = \int_M g \Delta f \, dV.$$

Exercise 5.1.16. *Prove Theorem 5.1.22. (Hint: do it first for functions supported on a single chart.)*

5.1.9 Markov triple on a compact Riemannian manifold

Let M be a compact Riemannian manifold and take $W \in C^\infty(M)$. Then $\int_M e^{-W} \, dV < \infty$, and so by adding a constant to W we may assume without loss of generality that $\int_M e^{-W} \, dV = 1$. By approximating indicators of open sets with smooth functions, one can define a probability measure μ on on the Borel sets of M using the formula

$$\mu(A) = \int_M 1_A e^{-W} \, dV.$$

Let $\mathcal{A} = C^\infty(M)$ and define $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ by

$$\Gamma(f, g) = \langle \nabla f, \nabla g \rangle = g^{ij} \partial_i f \partial_j g.$$

Because the Riemannian metric is a non-negative bilinear form, Γ is also a non-negative bilinear form. Moreover, Γ is a diffusion carré du champ in the sense of Definition 1.9.1.

Now, Theorem 5.1.22 implies that for $f, g \in C^\infty(M)$

$$\begin{aligned}\int_M f\Delta g\,d\mu &= \int_M (e^{-W}f)\Delta g\,dV \\ &= -\int_M \Gamma(e^{-W}f, g)\,dV \\ &= -\int_M \Gamma(f,g) - f\Gamma(W,g)\,d\mu.\end{aligned}$$

Setting $Lg = \Delta g - \langle \nabla W, \nabla g\rangle$ for $g \in C^\infty(M)$, we have the integration by parts formula

$$\int_M fLg\,d\mu = -\int_M \Gamma(f,g)\,d\mu.$$

Hence, we obtain the bound

$$\int_M \Gamma(f,g)\,d\mu \leq \|Lf\|_{L^2(\mu)}\|g\|_{L^2(\mu)}.$$

This implies that the Dirichlet form $\mathcal{E}$ is closable. All the required properties of Definition 1.9.5 follow immediately, except possibly for (h). For this last property, we appeal to a classical result on parabolic regularity:

Theorem 5.1.23 (Parabolic regularity theorem)**.** *Suppose that* $u : M\times\mathbb{R} \to \mathbb{R}$ *is a bounded solution to*

$$\begin{cases} \dfrac{\partial u(x,t)}{\partial t} = \Delta u - \langle \nabla W, \nabla u\rangle \\ u(x,0) = f(x) \end{cases} \tag{5.1}$$

where $f, W \in C^\infty(M)$*. Then* $u \in C^\infty(M)$*.*

With Theorem 5.1.23 in hand, note that the function $u(x,t) = (P_tf)(x)$ solves (5.1). It follows that for any fixed t, $P_tf = u(\cdot, t) \in C^\infty(M) = \mathcal{A}$. It follows now that (M, μ, Γ) is a compact Markov triple.

5.1.10 Curvature

Definition 5.1.24 (Lie bracket)**.** *Given* $X, Y \in \mathcal{T}M$*, their* Lie bracket *is the vector field* $[X, Y]$ *defined by*

$$[X,Y]f = X(Yf) - Y(Xf).$$

Exercise 5.1.17. *Show that* $[X, Y]$ *defines a vector field.*

Definition 5.1.25 (Riemannian curvature)**.** *Define the map* $R : \mathcal{T}M \times \mathcal{T}M \times \mathcal{T}M \to \mathcal{T}M$ *by*

$$R(X, Y, Z) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

The Riemannian curvature endomorphism *is the* $(3, 1)$*-tensor defined by*

$$R(X, Y, Z, \omega) = \omega(R(X, Y, Z)).$$

The Riemannian curvature tensor *is the* $(4, 0)$ *tensor defined by raising the last index of* R*:*

$$\mathrm{Rm}(X, Y, Z, W) = \langle R(X, Y, Z), W\rangle.$$

Exercise 5.1.18. *Show that* Rm *is a* $(4, 0)$*-tensor.*

Exercise 5.1.19. *On* $\mathbb{R}^n$ *with the Euclidean metric, the Riemannian curvature is identically zero.*

Definition 5.1.26 (Ricci curvature)**.** *The* Ricci curvature *is the* $(2, 0)$*-tensor* Ric *obtained by contracting the first and last positions of the Riemannian curvature endomorphism:*

$$\mathrm{Ric}(X, Y) = R(e_i, X, Y, \phi^i)$$

for some basis e_i *and its dual basis* ϕ^i*.*

We write R_{ijk}^{ℓ} for the coordinates of R and R_{ij} for the coordinates of Ric, so that

$$R_{ij} = R_{kij}^{k}.$$

Proposition 5.1.27. *The Ricci curvature is a symmetric* $(2, 0)$*-tensor, in the sense that* $\mathrm{Ric}(X, Y) = \mathrm{Ric}(Y, X)$*. Or in coordinates,* $R_{ij} = R_{ji}$*.*

The symmetry of the Ricci tensor follows from some (not completely trivial) symmetries of the Riemannian curvature tensor. See, for example, [Lee06, Proposition 7.4].

When our manifold M comes with a probability density e^{-W}, it turns out to be natural to consider a modified definition of curvature.

Definition 5.1.28 (Weighted Ricci curvature)**.** *Given* $W \in C^\infty(M)$*, its* weighted Ricci curvature *is the* $(2, 0)$*-tensor* Ric_W *defined by*

$$\mathrm{Ric}_W(X, Y) = \mathrm{Ric}(X, Y) + (\nabla^2 W)(X, Y).$$

The connection between curvature and the Γ-calculus is given by Bochner's formula. We define $\operatorname{grad} f = (\nabla f)^\sharp$ and

$$|\nabla^2 f|^2 = \sum_i |\nabla_{E_i} \nabla f|^2,$$

where E_i is an orthonormal basis.

Proposition 5.1.29. *Let M be a Riemannian n-manifold and take $W \in C^\infty(M)$. Define*

$$\begin{aligned} Lf &= \Delta f - \langle \nabla W, \nabla g \rangle \\ \Gamma(f,g) &= \langle \nabla g, \nabla g \rangle \\ \Gamma_2(f) &= \frac{1}{2} L\Gamma(f) - \Gamma(f, Lf). \end{aligned}$$

Then

$$\Gamma_2(f) = \operatorname{Ric}_W(\operatorname{grad} f, \operatorname{grad} f) + |\nabla^2 f|^2.$$

In proving Proposition 5.1.29, the computations will be much less messy if we work in a special set of coordinates known as *normal coordinates.* At a fixed point p, normal coordinates around p have two nice properties at the point p:

$$\langle \partial_i|_p, \partial_j|_p \rangle = \delta_{ij} \text{ and } (\nabla_{\partial_i} \partial_j)|_p = 0.$$

(Note that these properties only hold at exactly the point p, and not on a neighborhood of it.) For a construction of these coordinates, see [Lee06, Chapter 5].

Lemma 5.1.30. *In normal coordinates at p, $(\Delta f)(p) = \sum_{i=1}^n (\partial_i^2 f)(p)$.*

Proof. For this entire proof, everything is evaluated at p. By the definition of the trace, $\Delta f = \operatorname{tr} \nabla^2 f = (\nabla^2 f)(\partial_i, \partial_j) g^{ij}$. In normal coordinates, g^{ij} is the identity matrix at p, and hence $\Delta f = \sum_i (\nabla^2 f)(\partial_i, \partial_i)$. Now,

$$(\nabla^2 f)(\partial_i, \partial_i) = (\nabla_{\partial_i} f)(\partial_i) = \partial_i(\nabla_{\partial_i} f) - \nabla f(\nabla_{\partial_i} \partial_i),$$

where the last equality follows from the definition of what ∇_{∂_i} does to the covector field ∇f. Finally, note that in normal coordinates $\nabla_{\partial_i} \partial_i = 0$. Hence,

$$(\nabla^2 f)(\partial_i, \partial_i) = \partial_i(\nabla_{\partial_i} f) = \partial_i^2 f. \qquad \square$$

Proof of Proposition 5.1.29. Fix a point p and fix normal coordinates around p. For this entire proof, everything is evaluated at p.

First, we note a simple consequence of the symmetry of Hessians. By definition,

$$(\nabla^2 f)(X, Y) = (\nabla_X \nabla f)(Y) = \langle \nabla_X \operatorname{grad} f, Y\rangle.$$

Then the symmetry of $\nabla^2 f$ implies that

$$\langle \nabla_X \operatorname{grad} f, Y\rangle = \langle \nabla_Y \operatorname{grad} f, X\rangle.$$

Beginning with Lemma 5.1.30,

$$\begin{aligned}
\frac{1}{2}\Delta|\nabla f|^2 &= \frac{1}{2}\sum_{i=1}^{n} \partial_i\partial_i\langle \operatorname{grad} f, \operatorname{grad} f\rangle \\
&= \sum_{i=1}^{n} \partial_i\langle \nabla_{\partial_i} \operatorname{grad} f, \operatorname{grad} f\rangle \\
&= \sum_{i=1}^{n} \partial_i\langle \nabla_{\operatorname{grad} f} \operatorname{grad} f, \partial_i\rangle \\
&= \sum_{i=1}^{n} \langle \nabla_{\partial_i}\nabla_{\operatorname{grad} f} \operatorname{grad} f, \partial_i\rangle + \langle \nabla_{\operatorname{grad} f}, \nabla_{\partial_i}\partial_i\rangle \\
&= \sum_{i=1}^{n} \langle \nabla_{\partial_i}\nabla_{\operatorname{grad} f} \operatorname{grad} f, \partial_i\rangle,
\end{aligned}$$

since $\nabla_{\partial_i}\partial_i = 0$ at p. We continue by recalling the definition of the Riemannian curvature tensor.

$$\begin{aligned}
\frac{1}{2}\Delta|\nabla f|^2 = \sum_{i=1}^{n} &\operatorname{Rm}(\partial_i, \operatorname{grad} f, \operatorname{grad} f, \partial_i) &(5.2)\\
&+ \sum_{i=1}^{n} \langle \nabla_{\operatorname{grad} f}\nabla_{\partial_i} \operatorname{grad} f, \partial_i\rangle &(5.3)\\
&+ \sum_{i=1}^{n} \langle \nabla_{[\partial_i, \operatorname{grad} f]} \operatorname{grad} f, \partial_i\rangle. &(5.4)
\end{aligned}$$

By definition, (5.2) is $\operatorname{Ric}(\operatorname{grad} f, \operatorname{grad} f)$. To simplify (5.3), note that $\nabla_{\operatorname{grad} f}\partial_i =$

0 at p; hence, the compatibility of ∇ with the metric implies that

$$
\begin{aligned}
(5.3) &= \sum_{i=1}^{n} (\operatorname{grad} f)(\langle \nabla_{\partial_i} \operatorname{grad} f, \partial_i \rangle) \\
&= (\operatorname{grad} f) \sum_{i=1}^{n} (\nabla^2 f)(\partial_i, \partial_i) \\
&= (\operatorname{grad} f) \Delta f = \langle \nabla f, \nabla \Delta f \rangle.
\end{aligned}
$$

For part (5.4), note that because ∇ is torsion-free and we are working with normal coordinates, $[\partial_i, \operatorname{grad} f] = \nabla_{\partial_i} \operatorname{grad} f - \nabla_{\operatorname{grad} f} \partial_i = \nabla_{\partial_i} \operatorname{grad} f$. Hence,

$$
\begin{aligned}
(5.4) &= \sum_{i=1}^{n} \langle \nabla_{[\partial_i, \operatorname{grad} f]} \operatorname{grad} f, \partial_i \rangle \\
&= \sum_{i=1}^{n} \langle \nabla_{\nabla_{\partial_i} \operatorname{grad} f} \operatorname{grad} f, \partial_i \rangle \\
&= \sum_{i=1}^{n} \langle \nabla_{\partial_i} \operatorname{grad} f, \nabla_{\partial_i} \operatorname{grad} f \rangle \\
&= |\nabla^2 f|^2.
\end{aligned}
$$

Putting these computations together,

$$
\frac{1}{2} \Delta |\nabla f|^2 = \operatorname{Ric}(\operatorname{grad} f, \operatorname{grad} f) + \langle \nabla f, \nabla \Delta f \rangle + |\nabla^2 f|^2. \tag{5.5}
$$

This is in fact Bochner's original formula, which corresponds to the claim of Proposition 5.1.29 when W is a constant function. As a consequence of (5.5),

$$
\begin{aligned}
\Gamma_2(f) &= \frac{1}{2} L |\nabla f|^2 - \langle \nabla f, \nabla L f \rangle \\
&= \operatorname{Ric}(\operatorname{grad} f, \operatorname{grad} f) + |\nabla^2 f|^2 - \frac{1}{2} \langle \nabla W, \nabla |\nabla f|^2 \rangle + \langle \nabla f, \nabla \langle \nabla W, \nabla f \rangle \rangle.
\end{aligned}
$$

To simplify these last two terms, note that

$$
\frac{1}{2} \langle \nabla W, \nabla |\nabla f|^2 \rangle = \frac{1}{2} \nabla_{\operatorname{grad} W} |\nabla f|^2 = \langle \nabla f, \nabla_{\operatorname{grad} W} f \rangle,
$$

while

$$
\begin{aligned}
\langle \nabla f, \nabla \langle \nabla W, \nabla f \rangle \rangle &= \nabla_{\operatorname{grad} f} \langle \nabla W, \nabla f \rangle \\
&= \langle \nabla_{\operatorname{grad} f} \nabla W, \nabla f \rangle + \langle \nabla W, \nabla_{\operatorname{grad} f} \nabla f \rangle \\
&= \nabla^2 W(\operatorname{grad} f, \operatorname{grad} f) + \langle \nabla f, \nabla_{\operatorname{grad} W} \nabla f \rangle.
\end{aligned}
$$

It follows then that

$$\begin{aligned}\Gamma_2(f) &= \mathrm{Ric}(\operatorname{grad} f, \operatorname{grad} f) + |\nabla^2 f|^2 + \nabla^2 W(\operatorname{grad} f, \operatorname{grad} f)\\ &= \mathrm{Ric}_W(\operatorname{grad} f, \operatorname{grad} f) + |\nabla^2 f|^2. \qquad \square\end{aligned}$$

Let (M, g) be a compact Riemannian n-manifold and let $d\mu = e^{-W} dV$ be a probability measure on M, where $W \in C^\infty(M)$. Define $\Gamma(f, g) = \langle \nabla f, \nabla g\rangle$ on $\mathcal{A} = C^\infty(M)$, so that (M, μ, Γ) is a compact Markov triple. As consequence of Proposition 5.1.29, lower bounds on the Ricci tensor in terms of the metric tensor imply curvature-dimension conditions on (M, μ, Γ).

Corollary 5.1.31.

(a) If W is constant and $\mathrm{Ric} \geq \rho g$ then (M, μ, Γ) satisfies $\mathrm{CD}(\rho, n)$.

(b) If $\mathrm{Ric}_W \geq \rho g$ then (M, μ, Γ) satisfies $\mathrm{CD}(\rho, \infty)$.

(c) If $\mathrm{Ric}_W \geq \rho g$ and $m \in (n, \infty]$ satisfies $\mathrm{Ric}_W \geq \rho g + \frac{1}{m-n} \nabla W \otimes \nabla W$ then (M, μ, Γ) satisfies $\mathrm{CD}(\rho, m)$.

Proof. Note that the first two points are just special cases of the last point. For the last point, note by Proposition 5.1.29 that

$$\begin{aligned}\Gamma_2(f) &= \mathrm{Ric}_W(\operatorname{grad} f, \operatorname{grad} f) + |\nabla^2 f|^2\\ &\geq \rho\Gamma(f) + \frac{1}{m-n}\langle \nabla W, \nabla f\rangle^2 + |\nabla^2 f|^2.\end{aligned}$$

Now, under local coordinates with $\{\partial_1, \ldots, \partial_n\}$ an orthogonal basis at p, $|\nabla^2 f|^2$ is the sum of squared entries of an $n \times n$ matrix, while Δf is the sum of the diagonal entries. By the Cauchy-Schwarz inequality, $|\nabla^2 f|^2 \geq \frac{1}{n}(\Delta f)^2$. Hence,

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{1}{m-n}\langle \nabla W, \nabla f\rangle^2 + \frac{1}{n}(\Delta f)^2.$$

Thanks to the inequality $\frac{a^2}{m} + \frac{b^2}{m-n} \geq \frac{1}{m}(a+b)^2$, the $\mathrm{CD}(\rho, m)$ condition follows. $\square$

5.2 The n-sphere

Consider the n-sphere S^n; that is, the set $S^n = \{p \in \mathbb{R}^{n+1} : |p|^2 = 1\}$ endowed with the metric coming from $\mathbb{R}^{n+1}$. To be precise, we identify T_pS^n with the subspace $p^\perp \subset \mathbb{R}^{n+1}$; then the metric on $T_pS^n \subset \mathbb{R}^{n+1}$ is defined to be the restriction of the usual Euclidean inner product.

Clearly S^n is a compact Riemannian manifold, and so we may apply our earlier machinery in order to construct a compact Markov triple (S^n, μ, Γ), where μ is the normalized Riemannian volume element on S^n. However, in order to apply any of our inequalities to this Markov triple, we must first compute the Ricci tensor on S^n. There are many different methods of carrying out this computation, although most of them require a more thorough introduction to methods in Riemannian geometry than these notes provide.

We will do our computation in the coordinates

$$(p^1, \dots, p^{n+1}) = (x^1, \dots, x^n, \sqrt{1 - |x|^2}).$$

The final step, for simplicitly, will be carried out only at the "north pole," where $x = 0$ and $p = (0, \dots, 0, 1)$. The rotational symmetry of S^n, however, will imply that the conclusion applies at every point.

As a first step, we compute the metric in our coordinates. On the one hand, $\langle \frac{\partial}{\partial p^i}, \frac{\partial}{\partial p^j} \rangle = \delta_{ij}$. On the other hand,

$$\frac{\partial p^i}{\partial x^j} = \begin{cases} \delta_{ij} & i \leq n \\ \frac{x^j}{\sqrt{1-|x|^2}} & i = n. \end{cases}$$

From this, it follows that

$$g_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle = \langle \frac{\partial}{\partial p^k} \frac{\partial p^k}{\partial x^i}, \frac{\partial}{\partial p^\ell} \frac{\partial p^\ell}{\partial x^j} \rangle = \delta_{ij} + \frac{x^i x^j}{1 - |x|^2}.$$

It will be convenient also to compute the inverse of the metric: in local coordinates,

$$g^{ij} = \delta_{ij} - x^i x^j;$$

this may be verified by checking that $g_{ij} g^{jk} = \delta_{ik}$. From now on, we write ∂_i instead of $\frac{\partial}{\partial x_i}$.

Next, we will compute the connection in local coordinates. Note that a connection is determined by by the vector fields $\nabla_{\partial_i} \partial_j$: once we know these, $\nabla_X Y = \nabla_{X^i \partial_i} (Y^i \partial_i)$ is determined by properties (a)–(c) of Definition 5.1.17. We claim that the Riemannian connection on S^n is determined by

$$\nabla_{\partial_i} \partial_j = \frac{x^i x^j}{1 - |x|^2} x^k \partial_k + \delta_{ij} x^k \partial_k;$$

to check this, it suffices to show that the quantity above is torsion-free and compatible with the metric. The torsion-free property holds because $\nabla_{\partial_i} \partial_j$

is symmetric in i and j, and because $[\partial_i, \partial_j] = 0$. For compatibility with the metric, we need to check that

$$\nabla_{\partial_i}\langle \partial_j, \partial_k\rangle = \langle \nabla_{\partial_i}\partial_j, \partial_k\rangle + \langle \partial_j, \nabla_{\partial_i}\partial_k\rangle.$$

On the left hand side,

$$\nabla_{\partial_i}\langle \partial_j, \partial_k\rangle = \partial_i\left(\delta_{jk} + \frac{x^j x^k}{1-|x|^2}\right) = \frac{2x^i x^j x^k}{(1-|x|^2)^2} + \frac{\delta_{ij}x^k + \delta_{ik}x^j}{1-|x|^2}.$$

On the other hand,

$$\begin{aligned}
\langle \nabla_{\partial_i}\partial_j, \partial_k\rangle &= \left(\frac{x^i x^j}{1-|x|^2} + \delta_{ij}\right) x^\ell \langle \partial_\ell, \partial_k\rangle \\
&= \left(\frac{x^i x^j}{1-|x|^2} + \delta_{ij}\right) \sum_\ell x^\ell \left(\delta_{k\ell} + \frac{x^k x^\ell}{1-|x|^2}\right) \\
&= \frac{x^k}{1-|x|^2}\left(\frac{x^i x^j}{1-|x|^2} + \delta_{ij}\right).
\end{aligned}$$

It follows then that ∇ is compatible with the metric.

Next, we will calculate the coordinates of the Riemannian curvature tensor at the north pole (i.e., where $x = 0$). Since $[\partial_i, \partial_j] = 0$, we have

$$R_{ijk\ell} = \langle \partial_\ell, \nabla_{\partial_i}\nabla_{\partial_j}\partial_k - \nabla_{\partial_j}\nabla_{\partial_i}\partial_k\rangle.$$

Now,

$$\begin{aligned}
\nabla_{\partial_i}\nabla_{\partial_j}\partial_k &= \nabla_{\partial_i}\left(\left(\frac{x^j x^k x^\ell}{1-|x|^2} + \delta_{jk}x^\ell\right)\partial_\ell\right) \\
&= \left(\nabla_{\partial_i}\left(\frac{x^j x^k x^\ell}{1-|x|^2} + \delta_{jk}x^\ell\right)\right)\partial_\ell + \left(\frac{x^j x^k x^\ell}{1-|x|^2} + \delta_{jk}x^\ell\right)\nabla_{\partial_i}\partial_\ell.
\end{aligned}$$

At the point $x = 0$, various parts of this expression vanish. The second term vanishes entirely, since $\nabla_{\partial_i}\partial_\ell = 0$. For the first term, $\nabla_{\partial_i}\frac{x^j x^k x^\ell}{1-|x|^2}$ also vanishes at $x = 0$. Hence,

$$\nabla_{\partial_i}\nabla_{\partial_j}\partial_k|_{x=0} = \delta_{jk}\partial_i$$

Since $g_{ij} = \delta_{ij}$ at $x = 0$, we have

$$R_{ijk\ell} = \langle \partial_\ell, \delta_{jk}\partial_i - \delta_{ik}\partial_j\rangle = \delta_{jk}\delta_{i\ell} - \delta_{ij}\delta_{j\ell}.$$

Hence, the components of the Ricci tensor at $x = 0$ are

$$R_{jk} = R_{ijk\ell}g^{i\ell} = (n-1)\delta_{jk} = (n-1)g_{jk}.$$

It follows then that on S^n, $\mathrm{Ric} = (n-1)g$.

5.2.1 The ultraspheric triple

Recall the ultraspheric Markov triple from Example 1.11.5: $\mathcal{A} = C^\infty([-1,1])$ with $\Gamma_U(f)(x) = (1-x^2)f'(x)^2$ and $d\mu_n(x) = C_n(1-x^2)^{n/2-1}\,dx$. To see the connection between this triple and the compact Markov triple on S^n, suppose that we have a function h on S^n that only depends on the last coordinate: $h(p) = f(p_{n+1})$. Using our earlier computation for Γ_{S^n} in local coordinates,

$$\Gamma_{S^n}(h) = \sum_{i=1}^n (\partial_i h)^2 - \sum_{i,j=1}^n x^i x^j \partial_i h \partial_j h.$$

Now, $h(x) = f(\sqrt{1-x^2})$ and so $\partial_i h = f'(\sqrt{1-x^2})\frac{x_i}{\sqrt{1-|x|^2}}$. Hence

$$\Gamma_{S^n}(h) = f'(\sqrt{1-x^2})^2 \left[\frac{|x|^2}{1-|x|^2} - \frac{|x|^4}{1-|x|^2}\right] = f'(p_{n+1})^2(1-p_{n+1}^2).$$

In other words, $(\Gamma_{S^n}(h))(p) = (\Gamma_U(f))(p_{n+1})$. This begins to clarify the correspondence between the triples (S^n, dV, Γ_{S^n}) and $((-1,1), \mu_n, \Gamma_U)$: the ultraspheric carré du champ acts on f in the same way that the spherical carré du champ acts on $p \mapsto f(p_{n+1})$.

The same correspondence holds at the level of the measures. Suppose that f is a function supported on $(0,1)$, so that $p \mapsto f(p_{n+1})$ is supported on a single chart. If $B = \{x \in \mathbb{R}^n : |x| \le 1\}$ then

$$\begin{aligned}\int_{S^n} f(p_{n+1})\,dV &= \int_B \sqrt{|\det g|} f(\sqrt{1-x^2})\,dx \\ &= \int_B \frac{f(\sqrt{1-|x|^2})}{\sqrt{1-|x|^2}}\,dx,\end{aligned}$$

where we used Sylvester's determinant theorem to compute

$$\det g = \det\left(I_n - \frac{xx^T}{1-|x|^2}\right) = 1 - \frac{|x|^2}{1-|x|^2} = \frac{1}{1-|x|^2}.$$

Moving to polar coordinates,

$$\int_{S^n} f(p_{n+1})\,dV = s_{n-1}\int_0^1 r^n \frac{f(\sqrt{1-r^2})}{\sqrt{1-r^2}}\,dr$$

where s_{n-1} is the surface area of the $(n-1)$-dimensional unit sphere. With the change of variables $t = \sqrt{1-r^2}$,

$$\int_{S^n} f(p_{n+1})\,dV = s_{n-1}\int_0^1 (1-t^2)^{n/2-1} f(t)\,dt.$$

By a similar argument for the lower hemisphere, if f is any smooth function on $[0,1]$ then

$$\int_{S^n} f(p_{n+1})\,dV = s_{n-1}\int_{-1}^{1}(1-t^2)^{n/2-1}f(t)\,dt.$$

Up to the difference in normalizing constants, this is the same as $\int_{-1}^{1} f\,d\mu$. In this sense, the Markov triple $((-1,1),\mu_n,\Gamma_U)$ may be seen as the Markov triple (S^n, dV, Γ_{S^n}) restricted to functions only depending on one coordinate.

Chapter 6

Gaussian spaces

6.1 Infinite-dimensional Gaussian measures

In this section, we develop another class of spaces on which we can place compact Markov triples with curvature lower bounds: the class of Gaussian measures on Banach spaces. In the end, the construction of such Markov triples will be essentially trivial – it will essentially be an extension of the Ornstein-Uhlenbeck example. However, the construction of the spaces themselves is interesting in its own right, and we will give a self-contained introduction. For more background on infinite-dimensional Gaussian measures, see [Bog98]; to see more general diffusion processes on infinite-dimensional spaces, see [DPZ14].

6.1.1 Integration on Banach spaces

Let $(E, \|\cdot\|)$ be a separable Banach space; recall that *separable* means that there exists a countable, dense subset of E. We write E^* for the dual space of E: the set of bounded linear functionals on E with norm

$$\|\phi\|_* = \sup\{\phi(x) : x \in E, \|x\| = 1\}.$$

We write $\mathcal{B}(E)$ for the Borel σ-algebra on E; from now on, when we say that a X is an E-valued random variable, we mean that it is $\mathcal{B}(E)$-measurable.

Exercise 6.1.1.

(a) If E is separable then there exists a countable set $\{\phi_n\} \subset E$ such that $\|x\| = \sup_n \phi_n(x)$ for every $x \in E$.

(b) $\mathcal{B}(E)$ is generated by sets of the form $\{x \in E : \phi(x) \leq a\}$, for $\phi \in E^*$ and $a \in \mathbb{R}$.

For a random variable $X : \Omega \to E$ on a probability space $(\Omega, \mathcal{F}, \mu)$, we say that X is *integrable* if $\mathbb{E}\|X\| < \infty$, where $\mathbb{E}\|X\| = \int_\Omega \|X\| \, d\mu$; note here that $\|\cdot\| : E \to \mathbb{R}$ is a continuous, and hence Borel measurable, function. Hence, $\|X\|$ is $\mathcal{B}(\mathbb{R})$-measurable.

When a random variable is integrable, we may define its integral, also known as the *Bochner integral*. We do this in two stages: first, if $X = \sum_{i=1}^N x_i 1_{A_i}$ is simple (i.e. takes only finitely many values) then we define

$$\mathbb{E}X = \sum_{i=1}^N x_i \mu(A_i).$$

In the general case, we will use the fact that any random variable may be approximated in norm by simple ones:

Lemma 6.1.1. *If E is a separable Banach space and $X : \Omega \to E$ is a random variable then there exists a sequence $X_n : \Omega \to E$ of simple random variables such that $\|X - X_n\| \searrow 0$ pointwise.*

Proof. Let $\{x_1, x_2, \dots\}$ be a countable dense subset of E. Define

$$X_n(\omega) = \operatorname{argmin}_{x \in \{x_1, \dots, x_n\}} \|X(\omega) - x\|,$$

where if there are multiple x_i that achieve the minimum then we choose the one with the smallest i. Clearly, each X_n is measurable and simple. Moreover,

$$\|X_n - X\| = \min_{i=1,\dots,n} \|X - x_i\|,$$

which is non-increasing in n. Finally, the density of the set $\{x_i\}$ ensures that $\|X_n - X\| \to 0$ pointwise. $\square$

Given an integrable random variable X, we will use Lemma 6.1.1 to define $\int_E X \, d\mu$: let X_n be a sequence of simple random variables such that $\|X_n - X\| \searrow 0$ pointwise. Then

$$\|\mathbb{E}X_n - \mathbb{E}X_m\| \leq \mathbb{E}\|X_n - X\| + \mathbb{E}\|X_m - X\|,$$

which converges to zero by the dominated convergence theorem and the fact that X is integrable. Hence, $\mathbb{E}X_n$ is a Cauchy sequence in E; then we define

$$\mathbb{E}X = \lim_{n \to \infty} \mathbb{E}X_n.$$

Note that this definition does not depend on the choice of approximating sequence X_n: given another such sequence $\tilde{X}_n$,

$$\|\mathbb{E}X_n - \mathbb{E}\tilde{X}_n\| \le \mathbb{E}\|X_n - X\| + \mathbb{E}\|\tilde{X}_n - X\| \to 0;$$

hence $\mathbb{E}X_n$ and $\mathbb{E}\tilde{X}_n$ have the same limit.

Exercise 6.1.2. *The Bochner integral has the usual properties of integrals: if X and Y are integrable E-valued random variables then*

(a) for $a \in \mathbb{R}$, aX is integrable and $\int_E aX\,d\mu = a\int_E X\,d\mu$;

(b) $X + Y$ is integrable and $\int_E X + Y\,d\mu = \int_E X\,d\mu + \int_E Y\,d\mu$;

(c) $\|\mathbb{E}X\| \le \mathbb{E}\|X\|$; and

(d) if $A : E \to F$ is a bounded linear operator between two separable Banach spaces then then AF is integrable and $\mathbb{E}[AX] = A\mathbb{E}X$.

6.1.2 Gaussian measures

A Gaussian measure μ on $\mathbb{R}$ takes one of two forms: it is either a unit mass at a point, or it has a density of the form $\frac{1}{\sqrt{2\pi}\sigma}\exp(-(x-m)^2/(2\sigma^2))$ for some $m, \sigma \in \mathbb{R}$. Recall that the push-forward of a measure μ on E under a measurable map $f : E \to F$ is the measure $f_*\mu$ on F defined by $(f_*\mu)(A) = \mu(f^{-1}(A))$.

Definition 6.1.2. *Let E be a separable Banach space. A (Borel) probability measure μ on E is* Gaussian *if for every $\phi \in E^*$, $\phi_*\mu$ is a Gaussian measure on $\mathbb{R}$. If every $\phi_*\mu$ has mean zero, we say that μ is* centered*.*

In the case that E is a finite-dimensional space, this definition of Gaussian measures coincides with the usual one: a probability measure on $\mathbb{R}^n$ is Gaussian if it can be written as a linear push-forward of the standard Gaussian measure $d\gamma_n(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)\,dx$.

Exercise 6.1.3.

(a) Prove the claim of the previous paragraph.

(b) Let μ be a Gaussian measure on E. For a finite collection $\phi_1, \ldots, \phi_k \in E^*$, show that $(\phi_1, \ldots, \phi_k)_*\mu$ is a Gaussian measure on $\mathbb{R}^k$.

It turns out that many well-known facts about finite-dimensional Gaussian measures extend to Gaussian measures on Banach spaces. We begin with a straight-forward one:

Lemma 6.1.3. *Let μ be a centered Gaussian measure on E, and let $X, Y \sim \mu$ be independent. Then $\frac{X+Y}{\sqrt{2}}$ and $\frac{X-Y}{\sqrt{2}}$ are independent and have distribution μ.*

Proof. Let $U = \frac{X+Y}{\sqrt{2}}$ and $V = \frac{X-Y}{\sqrt{2}}$. First, note that $\phi(U) + \psi(V) = \frac{1}{\sqrt{2}}(\phi(X) + \psi(X) + \phi(Y) - \psi(Y))$ for any $\phi, \psi \in E^*$. Since $\phi(X) + \psi(X)$ and $\phi(Y) - \psi(Y)$ are independent real-valued Gaussian variables, it follows that $\phi(U) + \psi(V)$ is also a real-valued Gaussian variable. Hence, U has a Gaussian distribution on E. Moreover, for any finite collection of ϕ_i and ψ_j, the finite-dimensional random variable

$$Z := (\phi_1(U), \dots, \phi_k(U), \psi_1(V), \dots, \psi_\ell(V))$$

is Gaussian, since any linear functional of Z may be written in the form $\phi(U) + \psi(V)$.

Now, recall that finite-dimensional Gaussian variables are independent if and only if they are uncorrelated. Applying this to X and Y, it follows that for any $\phi, \psi \in E^*$,

$$\mathbb{E}[\phi(U)\psi(V)] = \frac{\mathbb{E}[\phi(X)\psi(X)] - \mathbb{E}[\phi(Y)\psi(Y)]}{2} = 0.$$

It follows then that the first k components of Z are independent of the other ℓ components.

Recall (from Exercise 6.1.1) that sets of the form $\{x : \phi(x) \le a\}$ generate $\mathcal{B}(E)$ (where $\phi \in E^*$). By the monotone class theorem, if we want to show that U and V are independent, it suffices to show that $\Pr(U \in A, V \in B) = \Pr(U \in A)\Pr(V \in B)$ for sets A and B of the form $\{x : \phi_1(x) \le a_1, \dots, \phi_k(x) \le a_k\}$. However, this follows from the previous paragraph. Hence, U and V are independent. □

The following theorem shows that any Gaussian measure on a separable Banach space has strong integrability properties.

Theorem 6.1.4 (Fernique)**.** *For every Gaussian random variable X on a separable Banach space E, there exists some $\epsilon > 0$ such that $\mathbb{E}\exp(\epsilon\|x\|^2) < \infty$.*

Proof. We first consider the case that μ is centered; then let X and Y be independent random variables with distribution μ. By Lemma 6.1.3,

$$\begin{aligned}
\Pr(\|X\| \le s)\Pr(\|Y\| > t) &= \Pr(\|X - Y\| \le \sqrt{2}s, \|X + Y\| > \sqrt{2}t) \\
&\le \Pr(\sqrt{2}\|X\| \ge t - s, \sqrt{2}\|Y\| > t - s) \\
&= \Pr(\sqrt{2}\|X\| \ge t - s)^2,
\end{aligned}$$

where the inequality follows because $2\|X\| \geq \|X+Y\| - \|X-Y\|$, and similarly for Y. Since X and Y have the same distribution, we may rearrange this to read

$$\Pr(\|X\| > t) \leq \frac{\Pr\left(\|X\| \geq \frac{t-s}{\sqrt{2}}\right)^2}{\Pr(\|X\| \leq s)}. \tag{6.1}$$

Since $\Pr(\|X\| < \infty) = 1$, there exists some t_0 such that $\Pr(\|X\| \leq t_0) > \frac{1}{2}$. We then recursively define $t_n = t_0 + \sqrt{2} t_{n-1}$; equivalently, $t_n = t_0(1+\sqrt{2})(2^{(n+1)/2} - 1)$. Define

$$p_n = \frac{\Pr(\|X\| > t_n)}{\Pr(\|X\| \leq t_0)}.$$

By (6.1) with $s = t_0$ and $t = t_n$, $p_n \leq p_{n-1}^2$. Then $p_n \leq p_0^{2^n}$, recalling that we chose t_0 so that $p_0 < 1$.

Now,

$$\begin{aligned}
\mathbb{E}\exp(\epsilon\|X\|^2) &\leq \exp(\epsilon t_0^2) + \sum_{n=0}^{\infty} \exp(\epsilon t_{n+1}^2) \Pr(t_n < \|X\| \leq t_{n+1}) \\
&\leq \exp(\epsilon t_0^2) + \sum_{n=0}^{\infty} \exp(\epsilon 36 t_0^2 2^n) p_0^{2^n} \\
&= \exp(\epsilon t_0^2) + \sum_{n=0}^{\infty} \exp(\epsilon 36 t_0^2 2^n - 2^n \log(1/p_0)),
\end{aligned}$$

where we have used the fact that $t_{n+1} \leq 3t_0 2^{n+1}$. If we choose $\epsilon > 0$ smaller than $\frac{\log(1/p_0)}{36 t_0^2}$, then the sum above converges. This proves the claim in the case that μ is centered.

If μ is not centered, then (again with X and Y independent random variables distributed according to μ) $X - Y$ has a centered Gaussian distribution on E. It follows that for some $\epsilon > 0$, $\mathbb{E}\exp(\epsilon\|X-Y\|^2) < \infty$. By Fubini's theorem, there exists some $y \in E$ such that $\mathbb{E}\exp(\epsilon\|X-y\|^2) < \infty$. Since $\|X\|^2 \leq 2\|X-y\|^2 + \|y^2\|$, it follows that

$$\mathbb{E}\exp\left(\frac{\epsilon}{2}\|X\|^2\right) \leq e^{\epsilon\|y\|^2}\mathbb{E}\exp(\epsilon\|X-y\|^2) < \infty. \qquad \square$$

As a corollary of Theorem 6.1.4, we may take means of Gaussian random vectors. In particular, the following corollary states that every Gaussian measure is some shift of a centered Gaussian measure. In the future, we will mainly study centered Gaussian measures.

Corollary 6.1.5. *Let X be a Gaussian random variable on E.*

(a) X is integrable.

(b) For all $\phi \in E^*$, $E\phi(X) = \phi(\mathbb{E}X)$.

(c) $X - \mathbb{E}X$ is a centered Gaussian random variable on E.

Example 6.1.6. *Let $E = C([0,1])$, the space of continuous, real-valued functions $[0,1]$, with the norm $\|f\| = \sup_{t\in[0,1]} |f(t)|$. This is a separable Banach space; for example, the set of polynomials with rational coefficients is countable and dense. The dual space E^* is the set of finite, signed measures on $[0,1]$.*

Let X be a Brownian motion on $[0,1]$, starting at 0. We remark that the usual construction of Brownian motion gives a process that is measurable under the product σ-algebra $\mathbb{R}^{[0,1]}$. However, this coincides with $\mathcal{B}(E)$ since the closed unit ball $\{f \in E : \|f\| \leq 1\}$ may be written as the countable intersection $\bigcap_{t\in\mathbb{Q}}\{f \in E : |f(t)| \leq 1\}$.

We claim that X is a centered Gaussian random variable on E. To check that it is Gaussian, consider first a signed measure $\mu \in E^*$ of the form $\mu = \sum_{i=1}^n a_i \delta_{t_i}$. Then $\mu(X) = \sum_{i=1}^n a_i X(t_i)$, which has a Gaussian measure by the definition of Brownian motion. Note that every $\mu \in E^*$ may be written as the weak limit of measures μ_n, where each μ_n is a finite sum of Dirac measures. Moreover, recall (or prove as an exercise, using the Fourier transform) that every pointwise limit of Gaussian random variables on $\mathbb{R}$ is Gaussian. Since every $\mu_n(X)$ is Gaussian, it follows that $\mu(X)$ is also Gaussian. Hence, X is Gaussian. To see that X is centered, it suffices to observe that whenever $\mu = \sum_{i=1}^n a_i \delta_{t_i}$ then

$$\mathbb{E}\mu(X) = \sum_{i=1}^n a_i \mathbb{E}X(t_i) = 0.$$

6.1.3 The Cameron-Martin space

Let γ be a centered Gaussian measure on E. Since every linear functional $\phi \in E^*$ is a square-integrable function $E \to \mathbb{R}$, E^* naturally embeds into $L^2(\gamma)$, the Hilbert space of (equivalence classes of) square-integrable functions $E \to \mathbb{R}$. Let E^*_γ denote the closure of E^* in $L^2(\gamma)$; as a closed, subspace of a Hilbert space, E^*_γ is also a Hilbert space. We write $\langle\cdot,\cdot\rangle_{*,\gamma}$ for its inner product and $|\cdot|_{*,\gamma}$ for its norm. Note that the inclusion from E^* to E^*_γ may not be injective: for example, if γ is just a Dirac measure then all linear functionals $\phi \in E^*$ coincide as elements of $L^2(\gamma)$.

Now let X be a E-valued random variable with distribution γ. Consider the map $J : E^*_\gamma \to E$ defined by $J(\phi) = \mathbb{E}[X\phi(X)]$ (where $X\phi(X)$ is integrable because $\mathbb{E}\|X\phi(X)\| = \mathbb{E}[\|X\||\phi(X)|]$, which is finite by the Cauchy-Schwarz inequality and Fernique's theorem). The map J is linear and injective: linearity is obvious; for injectivity, note that for any $\phi \in E^*$ that is not zero as an element of $L^2(\gamma)$, $\phi(J(\phi)) = \mathbb{E}[\phi(X)^2] \neq 0$, and so $J(\phi) \neq 0$. Since E^* is dense in E^*_γ, it follows that J is injective on E^*_γ also.

Let $H_\gamma \subset E$ be the range of J on E^*_γ. Since J is injective, we may define an inner product $\langle \cdot, \cdot \rangle_\gamma$ on H_γ by $\langle x, y \rangle_\gamma = \langle J^{-1}(h), J^{-1}(y) \rangle_{*,\gamma}$. Then H_γ is a Hilbert space; we call it the *Cameron-Martin space* for γ. Sometimes, either E^*_γ or H_γ is also called the *reproducing kernel Hilbert space* of γ.

The Cameron-Martin space may also be characterized as the Hilbert space that induces the $L^2(\gamma)$ norm on E^*:

Lemma 6.1.7. *H_γ is the unique Hilbert space H that is continuously included in E and satisfies*

$$|\phi|_{*,\gamma} = \sup\{\phi(h) : h \in H, |h|_H \leq 1\}$$

for all $\phi \in E^*$.

Proof. By definition, $|h|_\gamma \leq 1$ if and only if $h = J(\psi)$ for some $\psi \in E^*_\gamma$ with $\mathbb{E}[\psi(X)^2] \leq 1$. Hence,

$$\begin{aligned}
\mathbb{E}[\phi(X)^2] &= \sup\{\mathbb{E}[\psi(X)\phi(X)] : \psi \in E^*_\gamma, \mathbb{E}[\psi(X)^2] \leq 1\} \\
&= \sup\{\phi(J(\psi)) : \psi \in E^*_\gamma, \mathbb{E}[\psi(X)^2] \leq 1\} \\
&= \sup\{\phi(h) : h \in H_\gamma, |h|_\gamma \leq 1\}.
\end{aligned}$$

Moreover, the embedding of H_γ into E is continuous because

$$\|J(\phi)\|^2 \leq \mathbb{E}[\|X\||\phi(X)|]^2 \leq \mathbb{E}[\|X\|^2]|J(\phi)|^2_\gamma.$$

Hence, H_γ satisfies the conditions of the lemma.

Now suppose that H is another Hilbert space with the same properties. Then we may define a map $\tilde{J} : E^* \to H \subset E$ by

$$\langle \tilde{J}(\phi), h \rangle_H = \phi(h);$$

indeed, note that because H is continously embedded in E, the right hand side above is a bounded linear functional on H, and hence can be represented as an inner product. Moreover, the assumption of the lemma implies that for every $\phi \in E^*$,

$$|\phi|_{*,\gamma} = \sup\{\langle \tilde{J}(\phi), h \rangle_H : |h|_H \leq 1\} = |\tilde{J}(\phi)|_H.$$

That is, $\tilde{J}$ is an isometry from $E^* \subset L^2(\mu)$ to $H \subset E$. Hence,

$$\psi(\tilde{J}(\phi)) = \langle \tilde{J}(\phi), \tilde{J}(\psi)\rangle_H = \mathbb{E}[\phi(X)\psi(X)] = \psi(J(\phi))$$

for any $\psi, \phi \in E$. It follows that $\tilde{J} = J|_{E^*}$. Since H_γ is the closure of the range of $J|_{E^*}$, if we can show that the range of $\tilde{J}$ is dense in H then it will imply $H = H_\gamma$. For this last step, note that if $h \in H$ is orthogonal to the range of $\tilde{J}$ then $0 = \langle h, \tilde{J}(\phi)\rangle_H = \phi(h)$ for all $\phi \in E^*$, so $h = 0$. □

A consequence of Lemma 6.1.7 is that the Cameron-Martin space H_γ is an *intrinsic* feature of γ, in the sense that it is unchanged if we embed E into a larger space:

Corollary 6.1.8. *Let $E_1 \subset E_2$ be separable Banach spaces. Let γ_1 be a Gaussian measure on E_1 and let γ_2 be its extension to E_2 (i.e. $\gamma_2(A) = \gamma_1(A \cap E)$). Then $H_{\gamma_1} = H_{\gamma_2}$.*

Proof. Take $X \sim \gamma_2$ and $\phi_2 \in E_2^*$. Let $\phi_1 \in E_1^*$ be ϕ_2 restricted to E_1. Note that $X \in E_1$ almost surely, and hence

$$|\phi_2|^2_{*,\gamma} = \mathbb{E}[\phi_2(X)^2] = \mathbb{E}[\phi_1(X)^2] = |\phi_1|^2_{*,\gamma}.$$

Now, Lemma 6.1.7 implies that

$$|\phi_2|_{*,\gamma} = |\phi_1|_{*,\gamma} = \sup\{\phi_1(h) : |h|_{H_{\gamma_1}} \leq 1\} = \sup\{\phi_2(h) : |h|_{H_{\gamma_1}} \leq 1\}.$$

By the uniqueness part of Lemma 6.1.7 applied to E_2, it follows that $H_{\gamma_2} = H_{\gamma_1}$. □

Before proceeding to some examples, we will mention without proof one fundamental property of Cameron-Martin spaces (which, historically, was the property that Cameron and Martin were originally interested in). Given a measure γ on E and some $h \in E$, we write γ_h for the shifted measure $\gamma_h(A) = \gamma(A - h)$. Cameron and Martin were studying the problem of when γ_h has a density with respect to γ. For finite-dimensional Gaussian measures this is easy, and it depends only on the support of γ: assuming for convenience that γ is centered, γ_h has a density with respect to γ if and only if h belongs to the support of γ.

Theorem 6.1.9 (Cameron-Martin formula)**.** *Let γ be a Gaussian measure on a separable Banach space E. If $h \in E \setminus H_\gamma$ then γ_h and γ are mutually singular. If $h \in H_\gamma$ then γ_h and γ are mutually absolutely continuous, and*

$$\frac{d\gamma_h}{d\gamma} = \exp\left(g(x) - \frac{1}{2}|h|^2_\gamma\right),$$

where $g = J^{-1}(h) \in X^*_\gamma$.

Example 6.1.10. *Let X be a Brownian motion on $[0,1]$ and let γ be its distribution. We already saw that X is a centered Gaussian random variable, and that if $\mu \in E^*$ is a finite, signed measure then*

$$|\mu|^2_{*,\gamma} = \int_0^1 \int_0^1 s \wedge t \, d\mu(s) \, d\mu(t).$$

In order to compute the Cameron-Martin space of γ, we rewrite this quantity:

$$\begin{aligned}\int_0^1 \int_0^1 s \wedge t \, d\mu(s) \, d\mu(t). &= \int_0^1 \int_0^1 \int_0^1 1_{\{r \le s\}} 1_{\{r \le t\}} \, dr \, d\mu(s) \, d\mu(t) \\ &= \int_0^1 \mu([r,1])^2 \, dr.\end{aligned}$$

Now let $g : [0,1] \to \mathbb{R}$ be a smooth function with $g(0) = 0$. Integrating by parts,

$$\int_0^1 g'(r) \mu([r,1]) \, dr = \int_0^1 g(r) d\mu(r).$$

Since $C^\infty([0,1])$ is dense in $L^2([0,1])$,

$$\begin{aligned}|\mu|_{*,\gamma} &= \left(\int_0^1 \mu([r,1])^2 \, dr \right)^{1/2} \\ &= \sup \left\{ \int_0^1 \mu([r,1]) g'(r) \, dr : g \in C^\infty([0,1]), g(0) = 0, \int_0^1 g'(r)^2 \, dr \le 1 \right\} \\ &= \sup \left\{ \int_0^1 g(r) \, d\mu(r) : g \in C^\infty([0,1]), g(0) = 0, \int_0^1 g'(r)^2 \, dr \le 1 \right\}.\end{aligned}$$

Therefore, let $H^0_\gamma \subset C([0,1])$ be the set of smooth g such that $g(0) = 0$ and $\int_0^1 g'(r)^2 \, dr < \infty$, with the pre-Hilbert norm

$$|g|^2_\gamma = \int_0^1 g'(r)^2 \, dr.$$

According to Lemma 6.1.7 and the computation above, H_γ is the completion of H^0_γ with respect to $|\cdot|_\gamma$. Equivalently, H_γ consists of all measurable functions g such that there exists $h \in L^2([0,1])$ with $g(x) = \int_0^x h(t) \, dt$.

Example 6.1.11. *Let $\{X_n : n \ge 1\}$ be a sequence of independent real-valued Gaussian variables, with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = \sigma_n^2$. Depending on the properties of the sequence σ_n, the sequence $X = (X_1, X_2, \dots)$ may be viewed as a Gaussian random variable in one of various different spaces. The details are left as an exercise.*

- *If there is some $M \in \mathbb{R}$ such that $\sum_{n \geq 1} \exp(-M/\sigma_n^2) < \infty$ then X defines an ℓ^∞-valued random variable, where ℓ^∞ is the set of bounded sequences of real numbers, with the norm $\|x\|_\infty = \sup_n |x_n|$. Note that this is not a separable Banach space; there is no difficulty in extending the definition of Gaussian measures to non-separable Banach spaces.*
- *If for every $\epsilon > 0$, $\sum_{n \geq 1} \exp(-\epsilon/\sigma_n^2) < \infty$ then X defines a c_0-valued random variable, where c_0 is the set of sequences of real numbers that converge to zero, with the norm $\|x\|_\infty = \sup_n |x_n|$.*
- *If $\sum_{n \geq 1} \sigma_n^p < \infty$ for some $1 \leq p < \infty$ then X defines an ℓ^p-valued random variable, where ℓ^p is the set of sequences x with*

$$\|x\|_p^p = \sum_{n \geq 1} |x_n|^p < \infty.$$

In each of the cases above, the Cameron-Martin norm is given by

$$|x|_\gamma^2 = \sum_{n \geq 1} \frac{x_n^2}{\sigma_n^2}$$

and the Cameron-Martin space is the set of $x \in E$ (for one of the various choices of E) such that $|x|_\gamma^2 < \infty$. Note that in each of the cases, any sequence with $|x|_\gamma < \infty$ belongs to E. To check that this is the Cameron-Martin norm, note that for any linear function of the form $\phi(x) = \sum_{n \geq 1} x_n a_n$,

$$|\phi(x)|_{*,\gamma}^2 = \mathbb{E}[\phi(X)^2] = \sum_{n \geq 1} \sigma_n^2 a_n^2 = \sup \left\{ \sum_{n \geq 1} a_n x_n : \sum_{n \geq 1} \frac{x_n^2}{\sigma_n^2} \leq 1 \right\}.$$

In c_0 and ℓ^p for $p < \infty$, all linear functionals have the form above. In ℓ^∞, not all do; however, one can check that linear functionals of the form above generate $\mathcal{B}(E)$, and that it is enough to check the condition of Lemma 6.1.7 on a set of linear functionals that generate $\mathcal{B}(E)$.

6.1.4 Gaussian Hilbert spaces

In the case that γ is a centered Gaussian measure on a Hilbert space E, several nice simplifications occur. In particular, E^* is canonically isomorphic to E, and so we have the inclusions

$$H_\gamma \subset E \subset E_\gamma^*.$$

Moreover, there is a nice notion of covariance on Gaussian Hilbert spaces: let $K : E \to E$ be defined by

$$\langle Kx, y\rangle_E = \langle x, y\rangle_{*,\gamma} = \mathbb{E}[\langle X, x\rangle_E \langle X, y\rangle_E],$$

where $X \sim \gamma$ and we use $\langle \cdot, \cdot \rangle_E$ to denote the inner product under which E is a Hilbert space. By the Cauchy-Schwarz inequality,

$$\mathbb{E}[\langle X, x\rangle_E \langle X, y\rangle_E] \le |x|_E |y|_E \mathbb{E}[|X|_E^2],$$

and Fernique's theorem implies that $\mathbb{E}[|X|_E^2] < \infty$. It follows then that K is a bounded operator. Note that K is also symmetric (and hence self-adjoint) and non-negative.

Recall that a compact operator is one that maps bounded sets to compact sets, or equivalently one that maps weakly converging sequences to strongly converging sequences. If $x_n \in E$ is a sequence that weakly converges to zero, then

$$|\sqrt{K} x_n|_E^2 = \langle K x_n, x_n\rangle_E \le \mathbb{E}[\langle X, x_n\rangle_E^2],$$

which converges to zero by the dominated convergence theorem, since the integrand is bounded by $|X|_E^2 \max_n |x_n|_E^2$. Then $\sqrt{K}$ is compact, and so K is also. Now, every compact operator on a Hilbert space may be decomposed in terms of its eigenvectors and eigenvalues: there exists a complete orthonormal sequence $\{e_n\}$ and a real sequence $\{\lambda_n\}$ such that

$$Kx = \sum_{n\ge 1} \lambda_n e_n \langle e_n, x\rangle_E.$$

Moreover, zero is the only possibly accumulation point of the sequence $\{\lambda_n\}$. Note also that since $X = \sum_{n\ge 1} e_n \langle X, e_n\rangle_E$,

$$\mathbb{E}[|X|_E^2] = \sum_{n\ge 1} \mathbb{E}[\langle X, e_n\rangle_E^2] = \sum_{n\ge 1} \langle K e_n, e_n\rangle_E = \sum_{n\ge 1} \lambda_n.$$

Hence, K is a *trace-class operator*, meaning an operator such that

$$\sum_{n\ge 1} \langle K e_n, e_n\rangle_E < \infty$$

for some (or equivalently, every) orthonormal basis e_n. Finally, one can characterize H_γ, E, and E_γ^* in terms of the orthonormal basis $\{e_n\}$ of K's eigenvectors:

$$
\begin{aligned}
E &= \Big\{ \sum_{n\geq 1} a_n e_n : \sum_{n\geq 1} a_n^2 < \infty \Big\} && \text{with norm } \Big| \sum_{n\geq 1} a_n e_n \Big|_E^2 = \sum_{n\geq 1} a_n^2 \\
H_\gamma &= \Big\{ \sum_{n\geq 1} a_n e_n : \sum_{n\geq 1} \frac{a_n^2}{\lambda_n} < \infty \Big\} && \text{with norm } \Big| \sum_{n\geq 1} a_n e_n \Big|_\gamma^2 = \sum_{n\geq 1} \frac{a_n^2}{\lambda_n} \\
E_\gamma^* &= \Big\{ \sum_{n\geq 1} a_n e_n : \sum_{n\geq 1} \lambda_n a_n^2 < \infty \Big\} && \text{with norm } \Big| \sum_{n\geq 1} a_n e_n \Big|_{*,\gamma}^2 = \sum_{n\geq 1} \lambda_n a_n^2.
\end{aligned}
$$

As another interpretation, the ℓ^2 case of Example 6.1.11 is essentially the only construction of a Gaussian Hilbert space.

6.1.5 The white noise representation

Theorem 6.1.12. *Let γ be a Gaussian measure on the separable Banach space E. Let $e_1, e_2, \ldots$ be an orthonormal basis for H_γ that belongs to JE^*, and let $\xi_1, \xi_2, \ldots$ be independent standard Gaussian variables on $\mathbb{R}$. Then $\sum_{n\geq 1} \xi_i e_i$ converges almost surely to a random variable with distribution γ.*

Before proceeding to the proof of Theorem 6.1.12, we need to introduce a few tools. The first is Prohorov's theorem, which will help us prove convergence in distribution:

Definition 6.1.13 (Tightness)**.** *A sequence $\{\mu_n\}$ of probability measures is* tight *if for every $\epsilon > 0$, there exists a compact set K such that for every n $\mu_n(K) \geq 1-\epsilon$. A sequence of random variables is tight if their distributions are tight.*

Theorem 6.1.14 (Prohorov)**.** *A sequence of probability measures is tight if and only if it is weakly relatively compact (i.e. if every subsequence has a further subsequence that converges in distribution).*

In order to apply Prohorov's theorem, we first observe that every probability measure on a separable Banach space can be exhausted by compact sets. Equivalently, every singleton set $\{\mu\}$ is tight. This property is also sometimes called *inner regular*: μ is inner regular if $\{\mu\}$ is tight.

Lemma 6.1.15. *For every probability measure μ on a separable Banach space E, and for every $\epsilon > 0$, there exists a compact set K with $\mu(K) \geq 1-\epsilon$.*

Proof. Let $\{x_n\}$ be a countable, dense subset of E. Define

$$F_{m,k} = \bigcup_{i=1}^{m} \bar{B}(x_i, 1/k),$$

where $\bar{B}(x, r)$ is the closed ball of radius r centered at x. Since $F_{m,k} \to E$ as $m \to \infty$, there exists some m_k such that $\mu(F_{m_k,k}) \geq 1 - 2^{-k}\epsilon$. Note that $F_{m_k,k}$ is closed.

Let $K = \bigcap_{k=1}^{\infty} F_{m_k,k}$. Then $\mu(K) \geq 1 - \sum_{k\geq 1} 2^{-k}\epsilon = 1 - \epsilon$. Moreover, K is compact because it is closed and totally bounded. □

The second tool that we will need for the proof of Theorem 6.1.12 is a so-called *maximal inequality.* This will help us to upgrade the convergence in distribution (provided by Prohorov's theorem) into almost sure convergence.

Lemma 6.1.16. *Let $\{X_n\}$ be a sequence of independent random variables on E with symmetric distributions (i.e. X_n and $-X_n$ have the same distribution). Let $S_n = \sum_{i=1}^{n} X_n$. Then for all $t \geq 0$,*

$$\Pr(\max_{i\leq n} \|S_i\| \geq t) \leq 2\Pr(\|S_n\| \geq t).$$

Proof. Let N be the smallest i that $\|S_i\| \geq t$. Then

$$\Pr(\max_{i\leq n} \|S_i\| \geq t) = \sum_{i=1}^{n} \Pr(N = i).$$

Conditioned on $N = i \leq n$, the probability that $\|S_n\| \geq t$ is at least $\frac{1}{2}$, since the random variables $\|S_N + \sum_{i=N+1}^{n} X_i\|$ and $\|S_N - \sum_{i=N+1}^{n} X_i\|$ have the same distribution conditioned on N, and at least one of them is at least t. Hence, $\Pr(N = i) \leq 2\Pr(N = i, \|S_n\| \geq t)$, which implies the claim. □

Proof of Theorem 6.1.12. Since the claim of the theorem depends only on the distribution of the sequence $\{\xi_n\}$, it suffices to prove the theorem on a specific probability space. In particular, we take the probability space (E, γ). Take $\phi_n = J^{-1}(e_n) \in E^*$, which then form an orthonormal basis for E_γ^* which is dual to the basis $\{e_n\}$ in the sense that $\phi_m(e_n) = \delta_{mn}$. Let $\xi_n : E \to \mathbb{R}$ be given by $\xi_n(x) = \phi_n(x)$; since the ϕ_n are orthonormal in E_γ^*, the ξ_n are independent real-valued Gaussian variables. Let $S_n(x) = \sum_{i=1}^{n} e_i\xi_i(x)$ and set $S(x) = x$. Note that S is an E-valued random variable with distribution γ; our goal will be to show that $S_n \to S$ almost surely.

For any $m \le n$, $\phi_m(S_n) = \phi_m(x) = \phi_m(S)$. On the other hand, if $m > n$ then $\phi_m(S_n) = 0$. It follows that $\phi_m(S - S_n)$ and $\phi_m(S)$ are independent for all m (since one of them is always trivial). Hence, $\phi(S - S_n)$ and $\phi(S)$ are independent for every ϕ in E^*_γ, which is the closed linear span of $\{\phi_m\}$. Since E^*_γ generates $\mathcal{B}(E)$ up to γ-null sets, it follows that $S - S_n$ and S_n are independent; let $R_n = S - S_n$.

Next, we show that R_n is a tight sequence of random variables. For $\epsilon > 0$, choose a compact set $K \subset E$ such that $\gamma(K) \ge 1 - \epsilon/2$. Then

$$\gamma(K) = \Pr(S \in K) = \Pr(R_n + S_n \in K) \ge 1 - \epsilon/2.$$

By Fubini's theorem (over the independent random variables R_n and S_n), there exists some $x \in E$ such that

$$\Pr(x + R_n \in K) \ge 1 - \epsilon/2.$$

Since R_n is symmetric, we have also that $\Pr(x - R_n \in K) \ge 1 - \epsilon/2$. Hence, with probability at least $1 - \epsilon$, both $x + R_n$ and $x - R_n$ belong to K. On this event, $R_n \in K - K = \{y - z : y \in K, z \in K\}$, which is a compact set. To summarize, $\Pr(R_n \in K - K) \ge 1 - \epsilon$, which implies that $\{R_n\}$ is tight.

By Prohorov's theorem, there is a random variable R such that some subsequence of R_n converges to R in distribution. Since $\phi_m(R_n) = \phi_m(S) - \phi_m(S_n) \to 0$ as $n \to \infty$, it follows that $\phi_m(R) = 0$ for every m (recalling that every $\phi_m \in E^*$ is a continuous function on E). Since the ϕ_m generate $\mathcal{B}(E)$ up to γ-null sets, every possible subsequential limit R of R_n is the constant random variable 0. Hence, $S - S_n$ converges to zero in distribution, and therefore also in probability. That is, $\Pr(\|S - S_n\| > \epsilon) \to 0$ for every $\epsilon > 0$.

Finally, we upgrade the convergence of $S - S_n$ to be almost sure. By Lemma 6.1.16, for every $m < n$,

$$\begin{aligned}\Pr(\max_{m<\ell\le n} \|S_\ell - S_m\| \ge t) &\le 2\Pr(\|S_n - S_m\| \ge t) \\ &\le 2\Pr(\|S_m - S\| \ge t/2) + 2\Pr(\|S_n - S\| \ge t/2).\end{aligned}$$

Taking $n \to \infty$,

$$\Pr(\sup_{\ell>m} \|S_\ell - S_m\| \ge t) \le 2\Pr(\|S_m - S\| \ge t/2).$$

Since the right hand side converges to zero as $m \to \infty$, it follows that S_ℓ has a limit almost surely (and it can only be zero, since that is the limit in distribution). $\square$

6.1.6 A compact Markov triple

Using the orthonormal system of coordinates provided by the Cameron-Martin space H_γ, we will construct compact Markov triples on Gaussian measures. Let $\{\phi_n\}$ be an orthonormal (in $L^2(\gamma)$) basis of E^*, and take $e_n = J(\phi_n)$. Take $\mathcal{A}_0$ to be the set of functions of the form $f(x) = g(\phi_1(x), \dots, \phi_k(x)$, where $g : \mathbb{R}^k \to \mathbb{R}$ is a smooth, bounded function such that all of its derivatives vanish super-polynomially fast near ∞. Define $\Gamma : \mathcal{A}_0 \times \mathcal{A}_0 \to \mathcal{A}_0$ by

$$\Gamma(f_1, f_2)(x) = \langle \nabla g_1(\phi_1(x), \dots, \phi_k(x), \nabla g_2(\phi_1(x), \dots, \phi_k(x)) \rangle,$$

where $f_i(x) = g_i(\phi_1(x), \dots, \phi_k(x))$. Note that (since $\phi_1, \dots, \phi_k$ are orthonormal in $L_2(\gamma)$), $(\phi_1(x), \dots, \phi_k(x))$ has a standard Gaussan distribution on $\mathbb{R}^k$. Integrating by parts, we have

$$Lf(x) = \Delta g(\phi_1(x), \dots, \phi_k(x)) - \sum_{i=1}^k \phi_i(x) \partial_i g(\phi_1(x), \dots, \phi_k(x)).$$

Moreover, the associated semigroup P_t acts on $f \in \mathcal{A}_0$ in the same way that the Ornstein-Uhlenbeck semigroup acts on functions $\mathbb{R}^k \to \mathbb{R}$:

$$\begin{aligned}(P_t f)(x) &= \mathbb{E} f(e^{-t} x + \sqrt{1 - e^{-2t}} X) \\ &= \int_{\mathbb{R}^k} g(e^{-t}(\phi_1(x), \dots, \phi_k(x)) + \sqrt{1 - e^{-2t}} y) \, d\gamma_n(y)\end{aligned}$$

where $X \sim \gamma$ and γ_n is the standard Gaussian measure on $\mathbb{R}^k$. (The easiest way to see that this agrees with e^{tL} is to work backwards: starting from the above formula for P_t, we have already computed $\frac{dP_t}{dt}$ and it agrees with the formula for L above.)

Most of the properties of Definition 1.9.5 are trivial to check. We comment briefly on the last two: for $f \in \mathcal{A}_0$ with $f(x) = g(\phi_1(x), \dots, \phi_k(x))$, note that

$$\Delta g - \sum_{i=1}^k x_i \partial_i g$$

is still smooth, bounded, and has partial derivatives that vanish polynomially fast at ∞; this shows that $L\mathcal{A}_0 \subset \mathcal{A}_0$. On the other hand, the formula above for $P_t f$ implies that $P_t f$ has the form $P_t f(x) = g_t(\phi_1(x), \dots, \phi_k(x))$, and that $\partial_i g_t = e^{-t} P_t \partial_i g$. Moreover, the formula for P_t shows that if $\partial_i g$ decays super-polynomially at infinity then $P_t \partial_i g$ does also. The same argument applies to repeated derivatives of g_t, and hence $P_t f \in \mathcal{A}_0$.

Now we have a compact Markov triple on (E, γ), but in order to get useful inequalities we need to establish a curvature-dimension condition. As before, this just reduces to the finite-dimensional problem. As we computed already, the standard Gaussian measure on $\mathbb{R}^k$ satisfies $\mathrm{CD}(1, \infty)$ for any k. Since $\mathcal{A}_0$ consists only of "finite-dimensional" functions, the same inequality applies for all $f \in \mathcal{A}_0$; we conclude that (E, γ, Γ) satisfies $\mathrm{CD}(\rho, \infty)$.

In order to interpret the implications of the $\mathrm{CD}(\rho, \infty)$ condition, note that in our setting, there are multiple possible notions of Lipschitz functions:

(a) f is L-Lipschitz with respect to $\|\cdot\|$ if $|f(x) - f(y)| \le L\|x - y\|$ for every $x, y \in E$;

(b) f is L-Lipschitz with respect to $|\cdot|_\gamma$ if $|f(x) - f(x + h)| \le L|h|_\gamma$ for every $x \in E$ and $h \in H_\gamma$; or

(c) f is L-Lipschitz with respect to Γ if $f \in \mathcal{D}(E)$ and $\Gamma(f) \le 1$ γ-almost surely.

Of course, the third of these is the one that we have been using throughout these notes, and it is the one that we must use in interpreting consequences of the $\mathrm{CD}(1, \infty)$ property. However, these three notions are related:

Proposition 6.1.17. *If f is L-Lipschitz with respect to $\|\cdot\|$ then it is CL-Lipschitz with respect to $|\cdot|_\gamma$, where $C = \sup\{\frac{\|h\|}{|h|_\gamma} : h \in H_\gamma, h \neq 0\} < \infty$. If f is L-Lipschitz with respect to $|\cdot|_\gamma$ then it is L-Lipschitz with respect to Γ.*

Proof. The first statement follows from the fact that $\|h\| \le C|h|_\gamma$; moreover, we have already seen that C is finite: it is equivalent to the fact that H_γ is continuously embedded in E.

To prove the second statement, suppose that f is 1-Lipschitz with respect to $|\cdot|_\gamma$ (it suffices to check the case $L = 1$). By the white noise decomposition of Theorem 6.1.12, the random variable $X = \sum_{n \ge 1} \xi_n e_n$ has distribution γ, where ξ_n are independent real-valued standard Gaussians. In order to show that $f \in \mathcal{D}(\mathcal{E})$, we need to produce an approximating sequence in $\mathcal{A}_0$. Therefore, define $f_n : E \to \mathbb{R}$ by

$$f_n(x) = \mathbb{E} f\left(\sum_{i=1}^n \phi_i(x) e_i + \sum_{i=n+1}^\infty \xi_i e_i\right).$$

One can check directly that f_n is the conditional expectation of f with respect to the σ-algebra generated by $\phi_1, \dots, \phi_n$. In particular, $f_n \to f$ in $L^2(\gamma)$, because the set of all ϕ_i spans $L^2(\gamma)$.

Next, note that $f_n(x)$ depends only on $\phi_1(x), \dots, \phi_n(x)$. Therefore, let $g_n : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$g_n(y) = f_n\left(\sum_{i=1}^n y_i e_i\right);$$

this way, $f_n(x) = g_n(\phi_1(x), \dots, \phi_n(x))$. Then g_n is 1-Lipschitz with respect to the Euclidean metric on $\mathbb{R}^n$, since

$$\begin{aligned}
|g_n(y) - g_n(y')| &\le \mathbb{E}\left| f\left(\sum_{i=1}^n y_i e_i + \sum_{i=n+1}^\infty \xi_i e_i\right) - f\left(\sum_{i=1}^n y_i' e_i + \sum_{i=n+1}^\infty \xi_i e_i\right)\right| \\
&\le \left|\sum_{i=1}^n y_i e_i + \sum_{i=n+1}^\infty \xi_i e_i - \sum_{i=1}^n y_i' e_i + \sum_{i=n+1}^\infty \xi_i e_i\right|_\gamma \\
&= |y - y'|.
\end{aligned}$$

By the usual analysis on $\mathbb{R}^n$ (e.g. convolving with smooth mollifiers), g_n may be approximated in $L^2(\gamma_n)$ by smooth, 1-Lipschitz functions $g_{n,m}$ in such a way that $|\nabla g_{n,m}|$ converges in $L^1(\gamma_n)$ as $m \to \infty$. We call the limit $|\nabla g_n|$, and note that it may also be expressed as

$$|\nabla g_n|(x) = \lim_{\epsilon \to 0} \sup \left\{ \frac{|g_n(x) - g_n(y)|}{|x - y|} : y \ne x, |y - x| \le \epsilon \right\}. \tag{6.2}$$

It follows that $f_n \in \mathcal{D}(\mathcal{E})$ and

$$\Gamma(f_n)(x) = |\nabla g_n|^2(\phi_1(x), \dots, \phi_n(x)) \le 1.$$

Finally, we address the convergence of $\Gamma(f_n)$ as $n \to \infty$. To establish this convergence, we claim that $\Gamma(f_n)$ is a sub-martingale that is bounded by 1. From this, it will follow that $\Gamma(f_n)$ converges in $L^1(\gamma)$ to a limit that is also bounded by 1. Then $f \in \mathcal{D}(\mathcal{E})$, and $\Gamma(f)$ is equal (by definition) to the limit of $\Gamma(f_n)$.

It remains to show that $\Gamma(f_n)$ is a sub-martingale (we already know that it is bounded by 1). By Jensen's inequality, it suffices to show that $\sqrt{\Gamma(f_n)}$ is a sub-martingale, which is equivalent to showing that for every n and every $x \in \mathbb{R}^n$,

$$|\nabla g_n(x)| \le \mathbb{E}[|\nabla g_{n+1}|(x, \xi)],$$

where ξ is a standard normal variable. To check this inequality, note that $g_n(x) = \mathbb{E}[g_{n+1}(x, \xi)]$, and so by Jensen's inequality

$$\frac{|g_n(x) - g_n(y)|}{|x - y|} \le \mathbb{E}\frac{|g_{n+1}(x, \xi) - g_{n+1}(y, \xi)|}{|x - y|}.$$

Going back to (6.2), we have

$$\begin{aligned}
|\nabla g_n(x)| &\leq \sup\left\{\mathbb{E}\frac{|g_{n+1}(x,\xi) - g_{n+1}(y,\xi)|}{|x-y|} : y \neq x, |y-x| \leq \epsilon\right\} \\
&\leq \mathbb{E}\sup\left\{\frac{|g_{n+1}(x,\xi) - g_{n+1}(y')|}{|x-y'|} : y' \neq (x,\xi), |y' - (x,\xi)| \leq \epsilon\right\} \\
&= \mathbb{E}|\nabla g_{n+1}(x,\xi)|. \qquad \square
\end{aligned}$$

Bibliography

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Springer, 2014.

[Bog98] V. I. Bogachev. *Gaussian measures*. American Mathematical Society, 1998.

[DPZ14] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 2014.

[Lee06] J. Lee. *Riemannian manifolds: an introduction to curvature*. Springer, 2006.

[LNP15] M. Ledoux, I. Nourdin, and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geometric and Functional Analysis*, 25(1):256–306, 2015.

[MRN53] E. Marczewski and C. Ryll-Nardzewski. Remarks on the compactness and non direct products of measures. *Fundamenta Mathematicae*, 40(1):165–170, 1953.