Convergence to Equilibrium: Entropy Methods in Partial Differential Equations and Stochastic Analysis

MA8111: Projekt Mit Kolloquium

Alejandro Morera Alvarez

Contents

Abstract

We introduce entropy methods for evolution equations of the form

$$\partial_t u(x,t) + Au(x,t) = 0 \text{ for } x \in S,\, t > 0, \qquad u(x,0) = u_0(x) \text{ for } x \in S$$

where $S \subseteq \mathbb{R}^d$ is an open subset, X is a Banach space, $D(A) \subseteq X$ is a subset of X and $A : D(A) \to X'$ is an operator. We will use the heat equation and the linear Fokker-Planck equation as examples throughout this paper. Moreover, we discuss a special case of an entropy method due to Bakry and Emery from the viewpoint of partial differential equations and that of stochastic processes. To this end, we first present connections between evolution equations and Feller processes. These include the Hille-Yosida theorem, a generalization of Dynkin's formula, and an equation linking so-called stationary solutions of an evolution equation with stationary measures of a Feller process.

Key words: Entropy methods, Bakry-Emery, linear Fokker-Planck equation, Hille-Yosida theorem, Dynkin's formula, stationary solutions, stationary measures.

Preamble: Physics

Before we begin with the actual content of this paper, we would like to give an intuitive physical motivation for the next few pages. Consider the heat equation over $\mathbb{R}^d$:

$$\partial_t u(x,t) + \Delta u(x,t) = 0 \text{ for } x \in \mathbb{R}^d,\, t > 0, \qquad u(x,0) = u_0(x) \text{ for } x \in \mathbb{R}^d \tag{1}$$

whose solution $(x,t) \mapsto u(x,t)$ models the (internal) energy profile over the assumed-to-be thermodynamically closed system $\mathbb{R}^d$ at the point $x \in \mathbb{R}^d$ at any given time $t > 0$, for a certain an initial energy distribution u_0 at time $t = 0$. In physical systems, we expect that for large times $t > 0$, the energy profile u will tend to a solution of (1) which does not depend on time $(x,t) \mapsto u_\infty(x,t) = u_\infty(x)$, i.e. a thermodynamic equilibrium. In terms of daily life situations, this means that if a hot cup of coffee is left in a colder environment, it will give out part of its energy to its surroundings, and eventually reach a point in which its temperature is the same as that of its environment. We would like to know *how fast* does this convergence to the steady state happens.

There are several methods available to determine the rate of convergence. In this paper, we would like to address a specific one: an *entropy method* due to Bakry and Emery. An entropy method is a way to calculate an upper bound on the rate of convergence to the thermodynamic equilibrium u_∞ of an evolution equation through the use of so-called entropy functionals. Back in our example above, a way to measure the decay to thermodynamic equilibrium would be to take the $L^2(\mathbb{R}^d)$-norm of the difference between our solution $x \mapsto u(x,t)$ at time $t > 0$ and $x \mapsto u_\infty(x)$. This, however, has certain disadvantages, starting with the fact that the functions inside the norm need to be in $L^2(\mathbb{R}^d)$. Entropy methods allow us to choose from a wider class of functions to measure the decay of the solution u to equilibrium u_∞, which need not be as restrictive as a norm but at the same time still give a certain notion of distance.

Moreover, recall that the first law of thermodynamics states that energy is neither created nor destroyed. In a closed system that does no work and in which heat is not supplied, as the one modelled by (1), this means that the total energy is constant. In other words:

$$\frac{d}{dt} \int_{\mathbb{R}^d} u(\cdot,t)\, \mathrm{d}\lambda = 0\,.$$

We would like to analyze a dual interpretation of this conservation phenomenon. By this, we mean to construct from (1) a corresponding interpretation for *measures* and show that the measure corresponding to thermodynamic equilibrium preserves mass. In order to do so, it turns out that going into the realm of stochastic analysis is of great advantage. This is largely due to a result referred to as the Hille-Yosida theorem, which establishes a one-to-one correspondence between stochastic processes, semigroups, and linear operators. This result will allow us to translate the language of a deterministic problem into that of a stochastic one; more specifically: a problem involving functions to one involving measures, and vice versa.

In this paper, we will address the above phenomena, problems and questions in a more general setting:

$$\partial_t u(x,t) + Au(x,t) = 0 \text{ for } x \in S,\ t > 0, \qquad u(x,0) = u_0(x) \text{ for } x \in S \tag{2}$$

where $S \subseteq \mathbb{R}^d$ is an open subset, X is a Banach space, $D(A) \subseteq X$ is a subset of X and $A : D(A) \to X'$ is an operator.

In the first chapter, we motivate and introduce the concept of an entropy functional and entropy methods. The main purpose of doing so is to present the entropy method due to Barky and Emery, and show an example of how it is applied. In the second chapter, we address the importance of time-independent solutions to (2) and present an application. Finally, in the third chapter, we interpret our analytical problem (2) from a stochastic point of view and derive some interesting connections between the realm of functions and that of measures with regards to partial differential equations.

Without further ado, let's get to it!

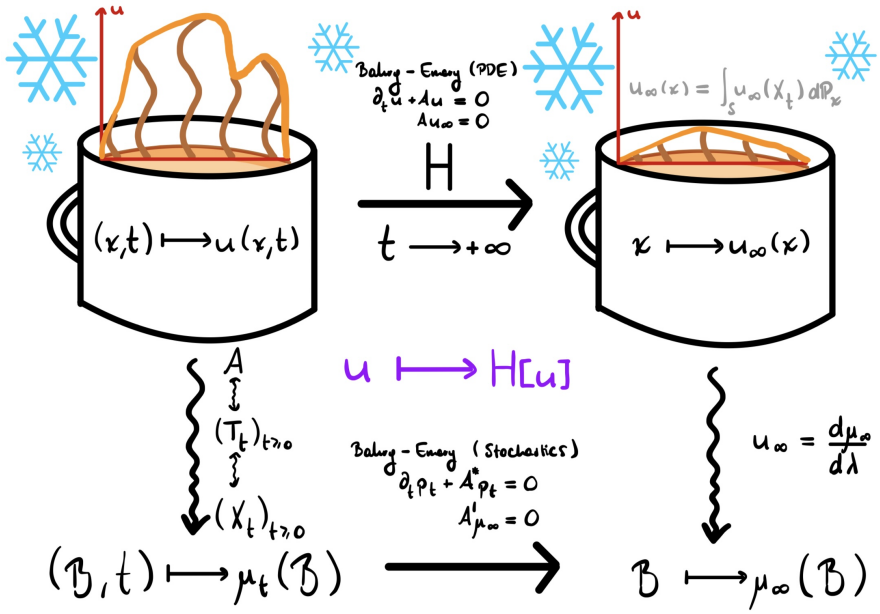

Figure 1: A research project in a cup of coffee.

Chapter 1

Entropy

In this chapter, we sketch the basic notions needed to introduce entropy methods. The first section is mainly concerned with definitions and examples in which we use what physicists call "entropy" in a mathematical setting. In the second section, we motivate and introduce Bakry's and Emery's approach to entropy methods through a first concrete example – a more thorough application is given in Chapter 2. Most of the contents in this chapter closely follow [J].

Unless otherwise stated, let $\langle \cdot, \cdot \rangle$ denote the Euclidean scalar product over $\mathbb{R}^d$ and $|| \cdot ||$ its induced norm. Norms over general function spaces X will always include their name as a subindex, i.e. $|| \cdot ||_X$.

1.1 Preliminaries

The first thing that comes into mind when one reads the word "entropy" is probably disorder and chaos in a physical system. Physicists in the 19th century even defined entropy as a measure for irreversibility of a thermodynamical process. But entropy is a concept that goes beyond physics. Disorder is present in every mathematical notion which has an inherent uncertainty with it. A classical example of such a notion is a random variable in stochastic analysis. In measure theory, these are simply measurable functions $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{A})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space and $(S, \mathcal{A})$ is some measurable space. For simplicity, we assume that S is discrete and $\mathcal{A} = \mathcal{P}(S)$, i.e. X is a discrete random variable. We further assume we know its probability density $f(x) = \mathbb{P}(X = x)$ for all $x \in S$. Then, the uncertainty carried by this random variable is captured by its **entropy**:

$$\mathcal{S} := E_{\mathbb{P}} I(X) := \int_\Omega I(f(X(\omega))) \, \mathrm{d}\mathbb{P}(\omega)$$

where $I : [0, 1] \to \mathbb{R}$. But: What is this function I? Why does the function only take values from the closed unit interval? To answer these questions, we consider the following problem statement ([C], p. 16, 18): we would like to find a so-called *information function* $I : [0, 1] \to \mathbb{R}$ taking probabilities $\mathbb{P}(A) \in [0, 1]$ as inputs, $A \in \mathcal{F}$, satisfying:

1. I decreases as $A \mapsto \mathbb{P}(A)$ increases.
2. $\mathbb{P}$-almost sure events do not contribute any information, i.e. $I(1) = 0$.
3. The information coming from two independent events is the sum of the information of these two events,

i.e. $I(\mathbb{P}(A)\mathbb{P}(B)) = I(\mathbb{P}(A)) + I(\mathbb{P}(B))$ for $A, B \in \mathcal{F}$ independent.[1]

From the functional equation for the exponential:

$$\exp(x+y) = \exp(x)\exp(y)$$

we see from the first and third condition that a candidate for such a function is given by the negative logarithm. Intuitively, the first condition says that the more information we have on an event, the less uncertainty there is. We get

$$\mathcal{S} = -\int_\Omega \log(f(X(\omega)))\,\mathrm{d}\mathbb{P}(\omega)\,. \tag{1.1}$$

Moreover, using the change of variables formula

$$\int_X h \circ g\,\mathrm{d}\mu = \int_{X'} h\,\mathrm{d}g_*\mu$$

for $h : (X', \mathcal{A}') \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ integrable, $g : (X, \mathcal{A}) \to (X', \mathcal{A}')$ measurable, and the pushforward $g_*\mu(B) := \mu(g^{-1}(B))$, we can rewrite (1.1) as

$$\begin{aligned}\mathcal{S} &= -\int_S f(x)\log f(x)\,\mathrm{d}\xi(x)\\ &= -\sum_{x\in S} f(x)\log f(x)\,.\end{aligned}$$

This interpretation of entropy goes in accordance to the observation Boltzmann made when he was studying kinetic theory (up to a minus sign). In that moment, Boltzmann wanted to prove the second law of thermodynamics,[2] which states that the entropy of an isolated system is non-decreasing. His so-called **H-theorem** states that the entropy of a particular physical system is related to the functional $f \mapsto H[f]$ given by

$$H[f] = \int_{\mathbb{R}^d\times\mathbb{R}^d} f\log f\,\mathrm{d}\lambda\,\mathrm{d}\nu \tag{1.2}$$

where ν is some measure over the Borel σ-field $\mathcal{B}(\mathbb{R}^d)$ and f is such that the integral in (1.2) is well-defined. We therefore see that these seemingly different notions of disorder in probability theory and statistical mechanics are connected through a functional $f \mapsto H[f]$. The convexity of the functional (1.1) will play a role later on, but we see that the difference between concave and convex is just a matter of taste: a minus or a plus sign. In mathematics, a convex functional $f \mapsto H[f]$ is typically used.

In the context of partial differential equations, we can also define a similar notion of "entropy", which we will call an *entropy functional*. In order to define such a mapping, we first need to build up our setting. Consider a Banach space X and an operator $A : D(A) \to X'$, where X' denotes the dual space of X, and $D(A) \subseteq X$ the domain of the operator A. Assume that for all $t > 0$ there exists a smooth solution $u(t) := u(\cdot, t) : D(A) \to \mathbb{R}$ to the **evolution equation**

$$\partial_t u + A(u) = 0, \qquad t > 0, \qquad u(0) = u_0 \tag{1.3}$$

and the stationary equation $A(u) = 0$ admits a unique, non-constant normalized solution $u_\infty \in D(A)$, referred to as the **steady state**. Note that we say a vector is **normalized**, if its entries are non-negative and the sum of its values or integral over its domain is equal to 1. We write u_∞ for the steady state of the considered partial

[1] Two events $A, B \in \mathcal{F}$ are independent, if the probability of their intersection equals the product of their probabilities.
[2] He was trying to do so from a statistical mechanical point of view.

differential equation.

Definition 1.1.1 ([J], Def. 1.1, Def. 1.2 p. 11) A **Lyapunov functional** is a mapping $H : D(A) \to \mathbb{R}$ which satisfies

$$\frac{d}{dt}H[u(t)] \leq 0$$

for all $t > 0$ and all smooth solutions u to (1.3).

A Lyapunov functional $H : D(A) \to \mathbb{R}$ is called an **entropy functional** of (1.3), if it is convex.

Example 1.1.2 ([J], Exa. 1.1) Let us consider the heat equation over the unit torus:

$$\partial_t u(x,t) = \Delta u(x,t), x \in \mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d, t > 0, \qquad u(\cdot,t) = u_0 \text{ in } \mathbb{T}^d \tag{1.4}$$

i.e. the domain is set in such a way so that we have periodic boundary conditions. Choosing $\mathbb{T}^d$ allows us to disregard some smoothness conditions of the boundary and avoids possible problems due to unboundedness of the underlying set. Assume $u_0 > 0$ is smooth and normalized, i.e. $||u_0||_{L^1(\mathbb{T}^d)} = 1$. For $\alpha \in (1, 2]$, we can define the functionals

$$H_\alpha[u] := \frac{1}{\alpha(\alpha-1)} \int_{\mathbb{T}^d} u^\alpha \, \mathrm{d}\lambda$$

and in the limit

$$H_1[u] := \lim_{\alpha \to 1} H_\alpha[u] = \int_{\mathbb{T}^d} u(\log(u) - 1) \, \mathrm{d}\lambda \, .$$

Notice the similarity of the above functional and (1.2).

We claim these mappings H_α are entropy functionals. We show our claim for $\alpha \neq 1$. Indeed, convexity is immediate, and using integration by parts together with (1.4), we conclude that H_α is an entropy functional: for all $t > 0$

$$\frac{d}{dt}H_\alpha[u(t)] = \frac{1}{\alpha - 1} \int_{\mathbb{T}^d} u^{\alpha-1} \Delta u \, \mathrm{d}\lambda = -\int_{\mathbb{T}^d} u^{\alpha-2} ||\nabla u||^2 \, \mathrm{d}\lambda \leq 0 \, .$$

△

Remark 1.1.3 (Conservation of Mass, [F], S. 4, Exe. 2) Solutions to the heat equation preserve mass. We show this for the case $S = \mathbb{R}^d$ and assuming that the solution u and its derivatives decay rapidly, i.e. there exists a constant $C > 0$ such that for all $x \in \mathbb{R}^d$ and all times $t > 0$:

$$|u(x,t)|, \, ||\nabla u(x,t)||, \, |\Delta u(x,t)| \leq \frac{C}{(1 + ||x||)^{d+1}} \, .$$

Define

$$M(t) := \int_{\mathbb{R}^d} u \, \mathrm{d}\lambda$$

to be the mass at time $t > 0$. Let ν be the unit normal vector function to the unit ball and σ be the measure

induced by the Gram-matrix on a sphere of radius $R > 0$. We have using the dominated convergence theorem

$$\begin{aligned}\left|\frac{d}{dt}M(t)\right| = \left|\int_{\mathbb{R}^d} \partial_t u \,\mathrm{d}\lambda\right| &= \left|\int_{\mathbb{R}^d} \Delta u \,\mathrm{d}\lambda\right| \\ &= \lim_{R\to\infty}\left|\int_{B_R(0)} \Delta u \,\mathrm{d}\lambda\right| \\ &= \lim_{R\to\infty}\left|\int_{\partial B_R(0)} \langle \nabla u, \nu\rangle \,\mathrm{d}\sigma\right| \\ &\leq \lim_{R\to\infty}\int_{\partial B_R(0)} ||\nabla u|| \,\mathrm{d}\sigma \\ &\leq \lim_{R\to\infty}\int_{\partial B_R(0)} \frac{C}{(1+R^{d+1})}\,\mathrm{d}\sigma = \lim_{R\to\infty}\frac{C\cdot\sigma(\partial B_R(0))}{(1+R)^{d+1}} = 0\end{aligned}$$

because $\sigma(\partial B_R(0)) = \kappa_d R^d$ for some dimension-dependent constant $\kappa_d > 0$. We shall see that under certain assumptions, a similar conservation principle is true for measures.

Note that our change from the unit torus to the space $\mathbb{R}^d$ was mainly done in order to use the decay assumptions we made on u and its derivatives. Intuitively, what we are doing by choosing $\mathbb{R}^d$ instead of $\mathbb{T}^d$ is making the space large enough so that no mass escapes.

Definition 1.1.4 ([J], Def. 1.3, p. 14) Let H be an entropy functional and u be a smooth solution to (1.3). For $t > 0$, we call

$$P[u(t)] = -\frac{d}{dt}H[u(t)]$$

the **entropy production**. An inequality of the form $\Phi(H[u]) \leq P[u]$ is called an **entropy-entropy production inequality**.

Remark 1.1.5 By definition, entropy production is always non-negative. This goes in accordance to physical entropy due to the second law of thermodynamics.

Example 1.1.6 The entropy production of H_α from Exa. 1.1.2 for $\alpha \neq 1$ is given by the non-negative functional:

$$P[u(t)] = \int_{\mathbb{T}^d} u^{\alpha-2}||\nabla u||^2 \,\mathrm{d}\lambda$$

for all $t > 0$. △

1.2 Entropy Methods: Motivation and Strategy

The general setting we want to consider is the following: take an evolution equation of the type (1.3) and assume we want to determine large-time asymptotics, i.e. the behaviour of the solution $(x,t) \mapsto u(x,t)$ for large $t > 0$. In many cases, we could directly apply spectral methods in order to analyze this (see later discussion on Poincaré inequalities), so:

What are the benefits of using entropy functionals to analyze an evolution equation instead of spectral methods?

In this section, we first give some intuition on what entropy methods are and why one would apply them instead of other alternatives, such as spectral methods, which may lead to the same results. Then, we sketch the steps

to the specific entropy method that Bakry and Emery developed. We closely follow the ideas and discussions in ([J], Sec. 1.3).

The answer to the above stated question is twofold: spectral methods mostly allow results in special metrics, most remarkably the one induced by the L^2-norm, *and* require A to be linear (or, more precisely, they require some linearization technique in order to apply them). Both can be large restrictions for general partial differential equations. In the two examples that follow, we shall study these two limitations that spectral methods have, and how entropy methods resolve them.

Example 1.2.1 (Alternative Distances) Consider the heat equation over the torus $\mathbb{T}^d$ as in (1.4) for a square integrable, non-negative initial condition $u_0 \in L^2(\mathbb{T}^d)$. By semigroup theory, we may write the solution $u(t)$ in terms of eigenvalues $(\lambda_k)_{k\in\mathbb{N}}$ and eigenfunctions $(v_k)_{k\in\mathbb{N}}$ of the Laplace operator Δ:

$$u(t) = \sum_{k\in\mathbb{N}} \exp(-\lambda_k t)\langle u_0, v_k\rangle_{L^2(\mathbb{T}^d)} v_k \tag{1.5}$$

where $\lambda_1 = 0$, v_1 is constant, and $\langle\cdot,\cdot\rangle_{L^2(\mathbb{T}^d)}$ denotes the inner product in $L^2(\mathbb{T}^d)$ between two functions in that same space. Therefore, we see that the convergence rate in $L^2(\mathbb{T}^d)$ is governed by $\lambda_2 > 0$, i.e. we have

$$||u(t) - u_\infty||^2_{L^2(\mathbb{T}^d)} \leq ||u_0 - u_\infty||^2_{L^2(\mathbb{T}^d)} \exp(-2t\lambda_2)\,. \tag{1.6}$$

However, note that in order to get (1.6), one uses the fact that the eigendecomposition in (1.5) is tailored to measure decays with respect to the $L^2(\mathbb{T}^d)$-norm and not to other distances. In particular, deriving an inequality of the form

$$H_1[u(t)] \leq H_1[u_0]\exp(-4t/C) \tag{1.7}$$

similar to (1.6) may not be as simple using spectral methods. We therefore emphasize that the assumptions of $u_0 \in L^2(\mathbb{T}^d)$ and the fact that we used the corresponding norm as a measure of decay were crucial to obtaining the representation in (1.5) and with it (1.6). Note also that having $H_1[u_0] < \infty$ does not necessarily imply $||u_0||_{L^2(\mathbb{T}^d)} < \infty$, i.e. that $u_0 \in L^2(\mathbb{T}^d)$. This means that one can apply the functional H_1 to a wider class of functions than $L^2(\mathbb{T}^d)$. △

Example 1.2.2 (Nonlinear Equations) Consider the following operator

$$A(u) := -\nabla\cdot\left(u\nabla\frac{\Delta\sqrt{u}}{\sqrt{u}}\right)$$

in the context of (1.3). An analysis using spectral methods would require different and possibly more advanced strategies, e.g. linearization techniques. △

Using the steps presented in the box below, one can derive an inequality of the form (1.7):

1. Compute the entropy production

$$-\frac{d}{dt}H[u(t)] = \langle A(u(t)), \frac{d}{du}H[u(t)]\rangle$$

where the braces $\langle\cdot,\cdot\rangle$ denote some kind of Banach space product.

2. Derive an inequality of the form

$$cH[u] \leq \langle A(u(t)), H'[u(t)]\rangle$$

where $H'[u]$ denotes the Fréchet derivative (for general normed spaces) of H, which we assume to exist. This implies $dH/dt \leq -cH$.

3. Use Gronwall's lemma to conclude that

$$H[u(t)] \leq H[u_0] \exp(-ct) .$$

Example 1.2.3 (Entropy Method: Heat Equation, [J], Sec. 1.3) Consider the heat equation as in (1.4). We expect that $u(t)$ converges to the steady state u_∞ (as defined above) as $t \to \infty$. We therefore consider the expression:

$$H_2[u] = ||u - u_\infty||^2_{L^2(\mathbb{T}^d)} .$$

The functional H_2 is clearly convex, since it is given by a norm. Using partial integration, we see that its derivative is non-positive:

$$\frac{d}{dt} H_2[u(t)] = 2 \int_{\mathbb{T}^d} (u(t) - u_\infty) \partial_t u(t) \, \mathrm{d}\lambda = 2 \int_{\mathbb{T}^d} (u(t) - u_\infty) \Delta u(t) \, \mathrm{d}\lambda = -2 \int_{\mathbb{T}^d} ||\nabla u(t)||^2 \, \mathrm{d}\lambda \leq 0$$

meaning that entropy production is non-negative. Therefore, H_2 is an entropy functional for (1.4). Using the classical Poincaré inequality:

$$\left|\left| f - \int_S f \, d\mu \right|\right|^2_{L^2(\mathbb{T}^d)} \leq C ||\nabla f||^2_{L^2(\mathbb{T}^d)}$$

for f in the Sobolev space $H^1(S)$, we get

$$\frac{d}{dt} H_2[u(t)] \leq \frac{-4}{C} ||u(t) - u_\infty||^2_{L^2(\mathbb{T}^d)} = \frac{-4}{C} H_2[u(t)]$$

and thus, by Gronwall's lemma, we conclude

$$H_2[u(t)] \leq H_2[u_0] \exp(-4t/C) .$$

△

Notice that the exponential decay and the functional inequality were derived in two separate steps. In the next chapter, we will see a way to get these simultaneously with a specific entropy method due to Bakry and Emery.

Chapter 2

Partial Differential Equations

2.1 Stationary Solutions: A Bird's Eye View

Many evolution equations tend to have special solutions which do not depend on time. In the previous chapter, we referred to these solutions as steady states. In this section, we analyze partial differential equations of the form (1.3) with unique, non-constant, normalized steady states, i.e. we assume that the equation

$$Au_\infty = 0$$

is uniquely solvable up to the above mentioned constrains. The goal of this section is merely to introduce the importance of stationary solutions; we will prove every claim in the next chapter once we have enough tools from stochastic analysis.

Example 2.1.1 ([F], Ch. 2) Steady states u_∞ of the heat equation $\partial_t u = \Delta u$ over some open set $S \subseteq \mathbb{R}^d$ are called **harmonic functions**. These are, by definition, those functions u_∞ which satisfy the **Laplace equation**:

$$\Delta u_\infty = 0$$

over S. For instance, in the case $S = \mathbb{R}^d \setminus \{0\}$ with $d \geq 3$, radially-symmetric harmonic solutions are given by **Newton potentials**:

$$u(x) = \frac{1}{\omega_d \cdot (d-2)} \frac{1}{||x||^{d-2}}$$

where $\omega_d > 0$ is a weight factor which includes the Lebesgue measure of the unit sphere in d dimensions. Here, however, we only have local integrability, i.e. u is integrable over any compact subset of $\mathbb{R}^d$. The fact that we have an explicit form for these types of solutions, as well as their importance in the solution to the **Poisson equation**:

$$\Delta u = g$$

(for some $g : \mathbb{R}^d \to \mathbb{R}$ with certain smoothness conditions) motivate their deeper study. We shall not study this here, but refer to ([F], Ch. 2). △

But why are stationary solutions so important for general evolution equations? There are several reasons – in this paper, we would like to discuss three of them. A special case of the first one was already seen in Exa. 1.2.1.

There, we saw that solutions of (1.4) converge to the steady state u_∞. In fact, this can be generalized to a wider class of partial differential equations. Intuitively speaking, this means that stationary solutions are stable equilibrium points of the evolution equations considered in this paper, i.e. those with a unique, normalized steady state. This is the functional-analytic viewpoint of the definition ([K], Def. 2.1), where the "distance" would be the so-called **relative entropy** $H[u|u_\infty] := H[u] - H[u_\infty]$.[1] The rate at which convergence takes place will be addressed in the next section.

The second reason comes from a special case of Dynkin's formula (cf. Ch. 3, Sec. 2). Loosely speaking, it turns out that for linear operators A in the partial differential equation (1.3), the value of the stationary solution u_∞ at a point $x \in S$ is its own integral evaluated at a specific stochastic process. This generalizes the mean value property of harmonic functions

$$u_\infty(x) = \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} u \, \mathrm{d}\lambda$$

for $x \in S$, but in a stochastic setting and for a broader class of evolution equations.

The third reason is the most relevant for this paper. This is because it connects stationary solutions to so-called stationary measures. Denoting the latter by μ_∞, we shall see that these are solutions to the adjoint steady state equation, i.e. $A'\mu_\infty = 0$. Here, A' denotes the Banach space adjoint of some operator A acting on a *measure space* in the sense that

$$\int_S Au \, \mathrm{d}\mu = \int_S u \, \mathrm{d}A'\mu$$

for functions $u : S \to \mathbb{R}$ and Borel measures $\mu : \mathcal{B}(S) \to [0, \infty)$. This duality is motivated by Riesz' representation theorem, which states that there is a one-to-one correspondence between measures and certain bounded linear functionals over continuous functions. A thorough treatment of this viewpoint will come in Ch. 3, Sec. 3.3.

2.2 Bakry-Emery: Partial Differential Equations

In this section, we illustrate how the exponential decay and the functional inequality from the steps presented in the previous chapter can be proven *simultaneously*. This entropy method is due to Bakry and Emery. As we shall see, their main idea is to estimate the second derivative of the given entropy functional. We closely follow the ideas and discussions in [J], Sec. 2.1.

Example 2.2.1 (Linear Fokker-Planck) Consider the linear partial differential equation

$$\partial_t u = \nabla \cdot (\nabla u + u \nabla V) =: A^* u, \qquad t > 0, \qquad u(0) = u_0 \text{ in } \mathbb{R}^d \tag{2.1}$$

for some smooth, convex potential V with $\exp(-V) \in L^1(\mathbb{R}^d)$, where A^* denotes the Banach space adjoint of some operator A acting on a *function space*. This means, for instance, that if we are working in $L^p(\mathbb{R}^d)$ for some $p \in (1, \infty)$, then the adjoint A^* acts on $L^q(\mathbb{R}^d)$, where $q^{-1} := 1 - p^{-1}$ is the Hölder conjugate to p. Formally, let us simply see this as the mapping A^* that satisfies

$$\int_{\mathbb{R}^d} vAu \, \mathrm{d}\lambda = \int_{\mathbb{R}^d} uA^* v \, \mathrm{d}\lambda$$

for u in a primal space and v in its dual (e.g. $L^p(\mathbb{R}^d)$ and $(L^p(\mathbb{R}^d))' \cong L^q(\mathbb{R}^d)$, typically under certain smoothness conditions of the functions). The reason why we (can) identify this specific operator with the adjoint A^* in the

[1] Relative entropy $H[u|u_\infty]$ is not a metric, but we can still interpret it as a kind of distance from the steady state to $u(t)$.

general case will become clear later on. Note that we make a formal difference between A^* and A', although they are closely related to each other, as we shall also see.

We assume u_0 is non-negative and has unit mass. Then, the solution u is also non-negative and the equation conserves mass, similar to what we showed in Rmk. 1.1.3. Moreover, the unique steady state u_∞ can be calculated from

$$0 = A^* u_\infty = \nabla \cdot (\nabla u_\infty + u_\infty \nabla V) \implies 0 = \nabla u_\infty + u_\infty \nabla V = u_\infty \nabla (\log u_\infty + V)\,.$$

If we assume $u_\infty > 0$, it follows that $\log u_\infty + V$ is constant. This implies:

$$u_\infty(x) = Z \exp(-V(x)), \qquad Z = \frac{1}{\int \exp(-V(x))\,\mathrm{d}\lambda}$$

where Z is a constant which normalizes the mass of the steady state u_∞. △

Now, let $\phi : [0,\infty) \to \mathbb{R}$ be a smooth, strictly convex function with $\phi(1) = \phi'(1) = 0$ and $1/\phi''$ concave. Define an entropy functional by

$$H_\phi[u] := \int \phi(u/u_\infty)\,\mathrm{d}\lambda\,.$$

Notice that this entropy functional generalizes (1.2) with $\phi(x) = x(\log(x) - 1)$.

We introduce some notation we will exclusively use in the following important proof.

Notation 2.2.2 For a symmetric matrix $A \in \mathbb{R}^{d\times d}$, we write $A \geq 0$ whenever A is positive semi-definite. Moreover, recall that any such matrix A canonically induces a symmetric, bilinear form $\langle \cdot, \cdot \rangle_A$:

$$\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto \langle x, y \rangle_A := x^T A y$$

The proof of the following result is an example of an application of the entropy method due to Barky and Emery.

Theorem 2.2.3 (Exponential Decay of Fokker-Planck, [J], Thm. 2.1) Let $\phi \in \mathcal{C}^4([0,\infty))$ be as above, assume $H_\phi[u_0] < \infty$ and let V be a potential satisfying the Bakry-Emery condition: there exists some $\lambda > 0$ with

$$\nabla^2 V - \lambda \cdot \mathrm{id} \geq 0\,. \tag{2.2}$$

Then, any normalized, smooth solution to (2.1) converges exponentially fast to the steady state in L^1, i.e. for all $t > 0$, we have

$$||u(t) - u_\infty||_{L^1} \leq \exp(-t\lambda) K_\phi H_\phi[u_0]^{1/2}$$

Proof: Step 1: Set $\rho = u/u_\infty$ and notice $\partial_t u = \nabla \cdot (u_\infty \nabla \rho)$. We then get that the entropy production is given by

$$P[u(t)] = -\frac{d}{dt} H_\phi[u(t)] = -\int \phi'(\rho) \partial_t u\,\mathrm{d}\lambda = \int \phi''(\rho) ||\nabla \rho||^2 u_\infty\,\mathrm{d}\lambda \geq 0$$

where we used integration by parts, Gauss' theorem, and the convexity of ϕ.

Step 2: We now estimate the second time-derivative of the entropy functional H_ϕ

$$\frac{d^2}{dt^2} H_\phi[u] = -\int (\phi'''(\rho)\partial_t u ||\nabla \rho||^2 + 2\phi''(\rho) \nabla \rho \cdot \partial_t \nabla \rho u_\infty)\,\mathrm{d}\lambda\,.$$

We split the integral additively into two I_1 and I_2, respectively. Using integration by parts as well as (2.1), we

can rewrite I_1 as

$$I_1 = \int (\phi''''(\rho)||\nabla\rho||^4 + 2\phi'''(\rho)\langle\nabla\rho, \nabla\rho\rangle_{\nabla^2\rho})u_\infty \,\mathrm{d}\lambda\,.$$

Additionally, after performing algebraic manipulations, differential calculus, and (2.1), we get

$$I_2 = -2\int \phi''(\rho)\big(\nabla\cdot(\nabla^2\rho\nabla\rho) - ||\nabla^2\rho||^2 - \langle\nabla\rho, \nabla V\rangle_{\nabla^2\rho} - \langle\nabla\rho, \nabla\rho\rangle_{\nabla^2 V}\big)u_\infty \,\mathrm{d}\lambda$$

Using integration by parts in the first term and the Bakry-Emery condition (2.2) in the last term,

$$\begin{aligned} I_2 \geq 2\int \Big(\phi'''(\rho)\langle\nabla\rho, \nabla\rho\rangle_{\nabla^2\rho}u_\infty + \phi''(\rho)\langle\nabla\rho, \nabla u_\infty + u_\infty\nabla V\rangle_{\nabla^2\rho} + \phi''(\rho)||\nabla^2\rho||^2 u_\infty\Big)\,\mathrm{d}\lambda \\ + 2\lambda\int \phi''(\rho)||\nabla\rho||^2 u_\infty \, d\lambda \end{aligned}$$

where, since u_∞ is the unique zero of A', the second term vanishes, and the last integral is equal to the negative of the first time derivative of H_ϕ. Adding the last expression together with I_1 yields

$$\begin{aligned} \frac{d^2}{dt^2}H_\phi[u] \geq 2\int \phi''(\rho)||\nabla^2\rho + \phi'''(\rho)/\phi''(\rho)\nabla\rho\otimes\nabla\rho||^2 u_\infty \,\mathrm{d}\lambda \\ + \int(\phi'''(\rho) - 2\phi'''(\rho)^2/\phi''(\rho))||\nabla\rho||^4 u_\infty \,\mathrm{d}\lambda - 2\lambda\frac{d}{dt}H_\phi[u] \end{aligned}$$

and note that $\phi'''' - 2(\phi''')^2/\phi'' = -(\phi'')^2(1/\phi'')''$, which is non-negative, since $1/\phi''$ is concave by assumption. We conclude for $t > 0$:

$$\frac{d^2}{dt^2}H_\phi[u] \geq -2\lambda H_\phi[u]\,. \tag{2.3}$$

Step 3: We show the exponential decay of the entropy functional H_ϕ by integrating both sides of (2.3) over $t \in (s, \infty)$

$$\frac{d}{dt}H_\phi[u(s)] - \lim_{t\to\infty}\frac{d}{dt}H_\phi[u(t)] \leq -2\lambda(H_\phi[u(s)] - \lim_{t\to\infty}H_\phi[u(t)])\,.$$

Note that if we show that both limits vanish, we are done using Gronwall's lemma:

$$H_\phi[u(s)] \leq H_\phi[u_0]\exp(-2\lambda s) \tag{2.4}$$

Indeed, note that (2.3) is also a statement about the derivative of the entropy production, to which we can also apply Gronwall's lemma to yield

$$P_\phi[u(t)] \leq P_\phi[u_0]\exp(-2\lambda t)$$

which implies that $\lim_{t\to\infty} P_\phi[u(t)] = 0$, showing that the first limit vanishes. Now, assuming we can interchange limits and entropy production, we conclude $\lim_{t\to\infty} u(t) = u_\infty$, since P_ϕ vanishes precisely at that functional. Hence, using $0 = \phi(1) = \phi(\lim_{t\to\infty} u/u_\infty)$, we get that the second limit vanishes as well.

Step 4: The exponential decay in L^1 is given by the Csiszár-Kullback-Pinsker inequality together with (2.4).

□

This approach appears to have predictive steps, but coming up with the inequalities and knowing when to apply integration by parts as well as showing why we can do so is a non-trivial task. For instance, notice that the smoothness condition on ϕ is required *because* we did integration by parts. It is not obvious from the statement of the theorem that we will require such a smoothness degree: e.g. one could be temped to think that $\phi \in \mathcal{C}^2([0,\infty))$

would suffice, because in the end, we are trying to exploit the *convexity* of the entropy functional, which is related to the second derivative.

Moreover, it is important to note that the latter proof-type can be made more systematic with an approach called **method of systematic integration by parts**, which we shall not discuss here, but refer to ([J], Ch. 3). It is also worth mentioning that Barky and Emery instead considered a stochastic approach when they proposed their method. We will discuss this partially in the next chapter.

Setting $s = 0$ and $u_0 = u$ in the above result, we get the following corollary.

Corollary 2.2.4 (Convex Sobolev Inequality, [J], Cor. 2.1, p. 24) Under the assumptions of Thm. 2.2.3, we have

$$\int \phi\left(\frac{u}{u_\infty}\right) u_\infty \, \mathrm{d}\lambda \leq \frac{1}{2\lambda} \int \phi''\left(\frac{u}{u_\infty}\right) \left|\left| \nabla \frac{u}{u_\infty} \right|\right|^2 u_\infty \, \mathrm{d}\lambda$$

Example 2.2.5 ([J], Exa. 2.2, p. 25) Let ρ be an integrable function with unit mass with respect to $d\nu = u_\infty d\lambda$ and $\sqrt{\rho} \in H^1(\nu)$. Choosing $\phi = s^2 - 1$ yields

$$\int (\rho^2 - 1) u_\infty \, \mathrm{d}\lambda \leq \frac{1}{\lambda} \int ||\nabla \rho||^2 u_\infty \, \mathrm{d}\lambda$$

or, using that both ρu_∞ and u_∞ have unit mass with respect to the Lebesgue measure, we have using notation from probability theory

$$E_\nu \rho^2 - (E_\nu \rho)^2 \leq \frac{1}{\lambda} \int ||\nabla \rho||^2 u_\infty \, \mathrm{d}\lambda$$

$$E_\nu \rho := \int \rho \, \mathrm{d}\nu$$

The above can be written more compactly using the variance

$$\mathrm{Var}_\nu(\rho) \leq \frac{2}{\lambda} \mathcal{E}(\rho) \tag{2.5}$$

where $\mathrm{Var}_\nu(\rho)$ denotes the **variance** of ρ

$$\mathrm{Var}_\nu(\rho) := E_\nu \rho^2 - (E_\nu \rho)^2$$

and $\mathcal{E}$ denotes the so-called **Dirichlet integral** of ρ

$$\mathcal{E}(\rho) := \frac{1}{2} \int ||\nabla \rho||^2 \, \mathrm{d}\nu$$

Inequalities such as (2.5) are referred to as **Poincaré inequality**. We shall treat such inequalities in Chapter 3. △

Chapter 3

Stochastic Processes

We would like to have alternative descriptions of evolution equations in terms of other mathematical objects. For instance, observe that the solution $(x,t) \mapsto u(x,t)$ to a linear evolution equation of the form (1.3) induces a semigroup $T_t : u_0 \mapsto T_t u_0$ for all $t \geq 0$ in the sense that the family of maps $(T_t)_{t\geq 0}$ records the evolution of the initial condition:

$$T_0 u_0(x) = u_0(x) = u(x,0) \text{ and } T_t T_s u_0(x) = T_{t+s} u_0(x) = u(x,t+s)$$

for all $x \in S$ and $t, s \geq 0$. Moreover, notice that the operator A can be interpreted as the time derivative of the above semigroup using (1.3):

$$Au(x,t) = \partial_t u(x,t) = \frac{d}{dt} T_t u_0(x)$$

In the first section of this chapter, we formalize these ideas and show a one-to-one correspondence between the objects T_t, A, and a specific type of stochastic process, called Feller process. The correspondence between these objects is known as the Hille-Yosida theorem. This will allow us to move back and forth between different perspectives and derive one of our main results, which is a duality relation between stationary solutions u_∞ of (1.3) and so-called stationary measures μ_∞ in stochastic analysis. At the end of this chapter, we will see that it was through these stochastic lenses that Bakry and Emery derived their entropy method described in the previous chapter.

3.1 Hille-Yosida Theorem for Feller Processes

Let S be a locally compact space, referred to as **state space**, endowed with some σ-field $\mathcal{B}$. One often has $S = \mathbb{R}^d$ endowed with the corresponding Borel σ-field with respect to the standard topology, or $S = \{1, ..., d\}$ for some $d \geq 1$ endowed with its power set as a σ-field. Moreover, let $\mathcal{C}_0(S)$ be the space of all vanishing continuous functions over S:

$$\mathcal{C}_0(S) = \{f \in \mathcal{C}(S) \mid \forall \epsilon > 0 \exists\, K \subseteq S \text{ compact } : \sup_{x \in S \setminus K} |f(x)| < \epsilon\}$$

endowed with the supremum norm. Let Ω be the set of all càdlàg functions $\omega : [0,\infty) \to S$. The word "càdlàg" is an acronym that comes from French: *continue à driote avec limite à gauche*, i.e. right-continuous with left-limits. We refer to Ω as **path space** and $\omega \in \Omega$ as **paths**. We endow Ω with a σ-field $\mathcal{F}$ and a probability measure $\mathbb{P}$ over $\mathcal{F}$ in order to have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $t \geq 0$, we denote by $X_t : \Omega \to S$ the canonical projection

of such paths into state space $X_t(\omega) = \omega(t)$. For a family of such maps $(X_t)_{t\geq 0}$, let $(\mathcal{F}_t)_{t\geq 0}$ be its associated right-continuous initial filtration, i.e. $\mathcal{F}_t$ is the smallest sub σ-field which makes all $(X_s)_{s\in[0,t]}$ measurable, and for $t \geq 0$ it holds:

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$$

Moreover, we assume the filtration exhausts the underlying σ-field in the following sense: $\mathcal{F} = \sigma\big(\bigcup_{t\geq 0} \mathcal{F}_t\big)$.

Notation 3.1.1 As we will be working with probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$, we adopt the notation

$$E_x f := \int f \, \mathrm{d}\mathbb{P}_x \tag{3.1}$$

for some integrable function $f \in L^1(\mathbb{P}_x)$, where $\mathbb{P}_x$ denotes some probability measure over path space Ω which depends on $x \in S$. Moreover, we write:

$$E_{X_t} f := \left(\int f \, \mathrm{d}\mathbb{P}_x \right)\Big|_{x=X_t} .$$

The notation used in (3.1) raises an interesting point of view which will play an important role later on. Notice that an integral can be interpreted as the action of a measure μ to an integrable function f. Whenever this viewpoint is of interest, we shall write

$$\langle \mu, f \rangle := \int f \, \mathrm{d}\mu \, .$$

There is no risk of confusion between the above bilinear form and the Euclidean scalar product, since it will be clear from context to which one we are referring to. Furthermore, in the case these two objects *do* coincide in a certain sense, as will happen later on, we will get the same result either way.

Moreover, for a random variable $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$, we shall write $X \sim \mu$, whenever the pushforward measure of $\mathbb{P}$ under X is given by μ, i.e. $\mathcal{S} \ni B \mapsto \mu(B) = X_*\mathbb{P}(B) := \mathbb{P}(X^{-1}(B))$. For $S = \mathbb{R}$, we write $X \sim \mathcal{N}(m, \sigma^2)$ whenever X is normally distributed with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$.

Definition 3.1.2 ([B], Def. 1.1) A **Feller process** is a family of maps $(X_t)_{t\geq 0}$ with $X_t : \Omega \to S$ defined as $X_t(\omega) := \omega(t)$ together with a family of probability measures $(\mathbb{P}_x)_{x\in S}$ on Ω and a filtration $(\mathcal{F}_t)_{t\geq 0}$ as we had above, which satisfy the following axioms:

(IC) $\forall x \in S : \mathbb{P}_x(X_0 = x) = 1$.

(MP) $\forall s, t \geq 0 : \; E_x(f(X_{s+t}) \mid \mathcal{F}_t) = E_{X_t} f(X_s)$.

(FP) $\forall f \in \mathcal{C}_0(S), t \geq 0 : x \mapsto E_x f(X_t)$ is in $\mathcal{C}_0(S)$.

We can define a family of operators $(T_t)_{t\geq 0}$ over $\mathcal{C}_0(S)$ by

$$T_t f(x) := E_x f(X_t) \tag{3.2}$$

Notice that without the so-called *Feller property* (FP) there is no reason why $T_t f$ should lie again in $\mathcal{C}_0(S)$. This is quite a strong condition, but in fact, many stochastic processes have this property. Thanks to the above combination of properties, one can show that Feller processes have the so-called **strong Markov property**:

$$E_x[(Y \circ \theta_\tau)(\omega) | \mathcal{F}_\tau] = E_{X_\tau} Y(\omega) \qquad \mathbb{P}\text{-a.s.}$$

for all bounded, measurable functions $Y : \Omega \to \mathbb{R}$ and $\mathbb{P}$-a.s. finite stopping times $\tau : \Omega \to [0, \infty]$. Here $\theta_t : \Omega \to \Omega$, $\omega \mapsto \omega(\cdot + t)$ denotes the time-shift.

Processes which only satisfy the first two properties only have the so-called **Markov property**, which is the same as the strong one but with deterministic times $t \geq 0$ instead of stopping times τ. Such processes are called **Markov processes**. The *initial condition* (IC) tells us that the process $(X_t)_{t \geq 0}$ is started at the point $x \in S$ $\mathbb{P}$-a.s. Together with the *Markov property* (MP), this implies that the family of operators $(T_t)_{t \geq 0}$ in (3.2) forms a particular type of semigroup.

Example 3.1.3 One often only writes the family of maps $(X_t)_{t \geq 0}$ to refer to the Feller process. For instance, given some non-zero drift direction $b \in \mathbb{R}^d$, we have that

$$X_t = x + tb$$

for $t \geq 0$, is a Feller process. The family $(X_t)_{t \geq 0}$ is commonly referred to as a **drift**.

A more complicated example is given by **Brownian motion** $(B_t)_{t \geq 0}$. It is characterized by having stationary, independent normally-distributed increments, i.e. given $0 \leq t_0 < t_1 < \cdots < t_m$, we have that $B_{t_i} - B_{t_{i-i}} \sim \mathcal{N}(0, t_i - t_{i-1})$ and paths $t \mapsto \omega(t) = B_t(\omega)$ are continuous $\mathbb{P}$-a.s. If $x = 0$ in the initial condition, we say we have **standard Brownian motion**. Later on, we will discuss an alternative way to describe this process with help of the convolution and, maybe most surprisingly, with the heat equation. △

Definition 3.1.4 ([P], Def. 2.1) A **probability semigroup** is a family of bounded, linear operators $(T_t)_{t \geq 0}$ from $\mathcal{C}_0(S)$ to itself, which satisfy the following axioms:

(S1) $T_0 = \mathrm{id}$.

(S2) $\forall f \in \mathcal{C}_0(S) : T_t f \xrightarrow{t \searrow 0} f$ with respect to $|| \cdot ||_\infty$.

(S3) $T_{t+s} = T_t T_s$.

(S4) $f \geq 0 \implies T_t f \geq 0$.

(S5) There exists a sequence $(f_n)_{n \in \mathbb{N}} \subseteq \mathcal{C}_0(S)$ which is bounded in norm, i.e. $\sup ||f_n|| < \infty$, such that for all $x \in S$ we have $f_n(x) \to 1$ and $T_t f_n(x) \to 1$ as $n \to \infty$.

Remark 3.1.5 Given a Feller process $(X_t)_{t \geq 0}$, one can show that the family $(T_t)_{t \geq 0}$ as defined in (3.2) is a probability semigroup. It is a non-trivial task to show that the converse is also true, i.e. that for every probability semigroup $(T_t)_{t \geq 0}$ there exists a corresponding Feller process $(X_t)_{t \geq 0}$ satisfying (3.2). We refer to the literature for a proof of this result (cf. [B], Thm. 1.16) and provide a sketch in Figure 3.2.

The property (S4) states that the operators $(T_t)_{t \geq 0}$ are positive. This allows us to apply a special version of Riesz' representation theorem in order to show the above one-to-one correspondence. This version of Riesz' theorem states that there is a bijection between measures and positive linear, bounded functionals over the continuous functions given by

$$\mu \mapsto \left(f \mapsto \int_S f \, \mathrm{d}\mu \right)$$

cf. [Ru], Thm. 2.14. On the other hand, (S5) is of purely probabilistic nature. Indeed, the intuition behind introducing (S5) is that the function $f = 1$ belongs to $\mathcal{C}_0(S)$ if and only if S is compact. In the locally compact case, the sequence $(f_n)_{n \in \mathbb{N}} \subseteq \mathcal{C}_0(S)$ given in (S5) approximates the behaviour that $f = 1 \notin \mathcal{C}_0(S)$ would have if

the state space S were compact, i.e. that for all $t \geq 0$

$$T_t 1 = 1\,.$$

But why do we want this behaviour? This will be motivated in Exa. 3.1.7.

Example 3.1.6 Instances of probability semigroups include

$$T_t f(x) = f(x + tb)$$

for the drift, where $x \in \mathbb{R}^d$ and $t \geq 0$. Moreover, by the definition used in Exa. 3.1.3, we see that for a Brownian motion we have for $x \in \mathbb{R}^d$ and $t \geq 0$:

$$T_t f(x) = (g_t * f)(x)$$

where g_t is the density of a normal distribution with parameters $x \in \mathbb{R}^d$ and $t > 0$:

$$g_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(\frac{||x||^2}{2t}\right)$$

which we shall refer to as a **Gaussian kernel**. The constant mass of the family of Gaussian kernels $(g_t)_{t>0}$ concentrates around 0 as $t \searrow 0$. More generally, writing $g_t(x, y) := g_t(x - y)$, we have that as $t \searrow 0$, the mass concentrates about the point $x \in \mathbb{R}^d$ in the following sense:

$$\lim_{t \searrow 0} g_t(x, y)\, d\lambda(y) = \mathrm{d}\delta_x(y)\,.$$

Intuitively, what the above states is that in the limit $t \searrow 0$, integration with respect to $g_t(x, y)\, d\lambda(y)$ acts like a Dirac distribution concentrated at the point $x \in \mathbb{R}$.

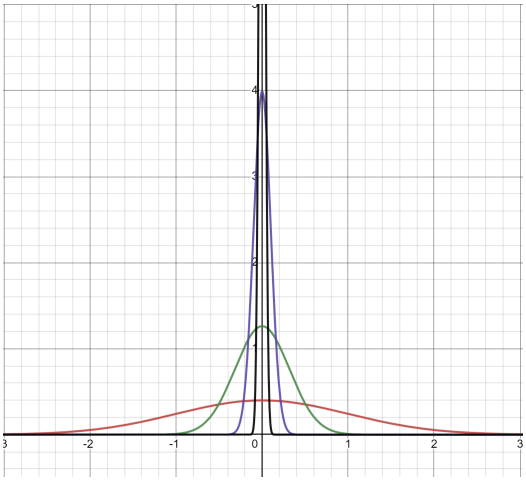

Figure 3.1: Concentration of Gaussian mass: The smaller $t > 0$, the narrower the distribution around $x = 0$.

△

The last element in our triad of objects in stochastic processes is the so-called (infinitesimal probability) generator A. Before we get into its actual definition, let us briefly motivate why we even are missing an object with a deterministic example. After this example, we will start seeing some similarities between the objects in ordinary differential equations and stochastic analysis.

Example 3.1.7 Recall from real analysis that any matrix $A \in \mathbb{R}^{d\times d}$ induces an initial value problem

$$\begin{cases} \dot{x} = Ax \\ x(0) = x_0 \end{cases}$$

where $x : [0,\infty) \to \mathbb{R}^d$ and $x_0 \in \mathbb{R}^d$ is some initial condition. The above problem has the unique solution $x(t) = \exp(tA)x_0$, where

$$t \mapsto \exp(tA) = \sum_{k\in\mathbb{N}} \frac{(tA)^k}{k!}.$$

The existence and uniqueness of the solution induces a well-defined map:

$$T_t : \mathbb{R}^d \to \mathbb{R}^d,\ x_0 \mapsto T_t x_0 := \exp(tA)x_0$$

which satisfies the properties (S1)-(S3): (S2) is clear because $\mathbb{R}^d$ is finite-dimensional. The properties (S1) and (S3) are seen as follows:

$$T_0 x_0 = \exp(0\cdot A)x_0 = \mathrm{id}\cdot x_0 = x_0 \qquad T_{t+s}x_0 = \exp((t+s)A)x_0 = \exp(tA)\exp(sA)x_0 = T_t T_s x_0.$$

Moreover, under the conditions $A_{i,j} \geq 0$ for $i \neq j$ and $\sum_{j=1}^d A_{ij} = 0$, we even get (S4) and (S5). In fact, we see that (S4) follows by definition of A. Moreover, notice that here $S = \{1,...,d\}$, and is therefore compact. Thus, in order to prove that (S5) holds, it suffices to show that $T_t 1 = 1$, where the 1 is meant as $(1,...,1)^T \in \mathbb{R}^d$, i.e. pointwise. Indeed, we see for all $i \in \{1,...,d\}$:

$$\sum_{j=1}^d (A^2)_{ij} = \sum_{j=1}^d \left(\sum_{k=1}^d A_{ik}A_{kj}\right)_{ij} = \sum_{k=1}^d A_{ik}\sum_{j=1}^d A_{kj} = 0.$$

Inductively, the same holds for A^n, $n \geq 3$. Therefore, for all $i = 1,...,d$

$$\sum_{j=1}^d (\exp(tA))_{ij} = 1$$

so that $\exp(tA)1 = 1$. In addition, we have the following differential properties:

$$\frac{d}{dt}T_t x_0 = A\exp(tA)x_0 = \exp(tA)Ax_0$$

and in particular:

$$Ax_0 = \lim_{t\searrow 0} \frac{T_t x_0 - x_0}{t}$$

where the limit is taken with respect to any norm (all norms over a d-dimensional space are equivalent for $d \in \mathbb{N}_{\geq 1}$). △

In short, in terms of what we have seen so far, we have for the above ordinary differential equation that $X_t := x_0$ is our (deterministic) Feller process, $(T_t)_{t\geq 0}$ defined as $T_t x_0 := \exp(tA)x_0$ is the corresponding probability semigroup, and... what is $A \in \mathbb{R}^{d\times d}$? The idea of a generator is to mimic and generalize the matrix A. However, one needs to be careful, because although generators are indeed linear operators, like in the example above, they are generally unbounded.

Definition 3.1.8 ([P], Def. 1.1) Given a probability semigroup $(T_t)_{t\geq 0}$, we define its (infinitesimal probability) **generator** A as the linear operator

$$A : D(A) \to \mathcal{C}_0(S), f \mapsto Af := \lim_{t\searrow 0} \frac{T_t f - f}{t} \tag{3.3}$$

where $D(A) \subseteq \mathcal{C}_0(S)$ is the set where the limit in (3.3) exists in the topology generated by $||\cdot||_\infty$ and such that the limit (3.3) lies in $\mathcal{C}_0(S)$.

Therefore, by definition, any probability semigroup induces a generator. The following theorem tells us when the converse is true: given a linear operator, under which conditions is it induced from a probability semigroup as in (3.3). The following result is a statement about a one-to-one correspondence between families of bounded, linear operators over $\mathcal{C}_0(S)$ and their corresponding (right-sided) derivatives. Together with the remark after Exa. 3.2, we get three perspectives of one same object: stochastic (Feller process), semigroup-theoretic (probability semigroup), and functional-analytic (generator) – cf. diagram at the end of the section.

Theorem 3.1.9 (Hille-Yosida, [P], Thm. 3.1; [B], Thm. 2.14) A linear operator A is the generator of a probability semigroup $(T_t)_{t\geq 0}$ if and only if the following conditions are met:

(L1) A is a closed operator, i.e. $(f_n) \subseteq D(A)$ with $f_n \xrightarrow{||\cdot||_\infty} f \in D(A)$ and $Af_n \xrightarrow{||\cdot||_\infty} g$ implies $g = Af$.

(L2) $D(A)$ is dense in $\mathcal{C}_0(S)$.

(L3) $(0, \infty) \subseteq \rho(A)$.

(L4) $\forall \lambda > 0 : ||(\lambda \cdot \mathrm{id} - A)^{-1}|| \leq 1/\lambda$.

(L5) $\forall \lambda > 0 \exists (f_n)_{n\in\mathbb{N}} \subseteq D(A) : ((\mathrm{id} - \lambda A)f_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}_0(S)$, $\sup_{n\in\mathbb{N}} ||f_n|| < \infty$, and for all $x \in S$:

$$f_n(x) \to 1 \qquad Af_n(x) \to 0\,.$$

Similar to (S5), the property (L5) provides a proxy for the function $f = 1 \notin \mathcal{C}_0(S)$ for the case that S is not compact. In fact, in the case that S is compact, one can write (L5) as

$$A1 = 0$$

which mimics the properties of a matrix whose rows sum up to zero, like the one we had in Exa. 3.1.7. Moreover, we have the **Kolmogorov equation**:

$$Af = \frac{d^+}{dt} T_t f \Big|_{t=0} \tag{3.4}$$

for $f \in D(A)$, where $\frac{d^+}{dt}$ denotes the right-sided derivative in (3.3). By the **exponential formula** (cf. [P], Thm. 8.3), we can also recover the probability semigroup:

$$T_t f = \lim \left(\mathrm{id} - \frac{t}{n} A\right)^{-n} f \tag{3.5}$$

for $f \in \mathcal{C}_0(S)$, generalizing the finite-dimensional case. It is important to note that (3.5) is another way of writing the exponential of a bounded operator as we had it in Exa. 3.1.7. However, conversely, given an unbounded operator, we cannot write its exponential as a series, since that would require $||A|| < \infty$.

The proof of the exponential formula makes clever use of the so-called **resolvent** of A,

$$R_\lambda(A) = (\lambda \cdot \mathrm{id} - A)^{-1} = \int_{(0,\infty)} \exp(-t\lambda) T_s f \, \mathrm{d}\lambda(s)$$

which can be though of as the Laplace transform of the mapping $s \mapsto T_s f$. We will discuss this in a longer version of this paper.

We summarize what we have seen so far in the following:

Corollary 3.1.10 Feller processes, probability semigroups, and infinitesimal probability generators are in a one-to-one correspondence.

Example 3.1.11 The drift has the associated generator

$$Af = \langle b, \nabla f \rangle .$$

This is easy to see:

$$Af = \lim_{t \searrow 0} \frac{f(x+tb) - f(x)}{t} = \lim_{t \searrow 0} \frac{f(x) + \langle tb, \nabla f \rangle + \mathcal{O}(t^2) - f(x)}{t} = \langle b, \nabla f \rangle .$$

With a little more work, one can show that the generator for Brownian motion is given by the (weighted) Laplace operator $\frac{1}{2}\Delta : D(\Delta) \to \mathcal{C}_0(S)$, which is unbounded. In addition, we see that its corresponding Kolmogorov equation (3.4) is the heat equation.

More generally, evolution equations of the form (1.3) induce generators. For instance, the linear Fokker-Planck equation (2.1) induces the generator $Au = \nabla \cdot (\nabla u + u \nabla V)$. △

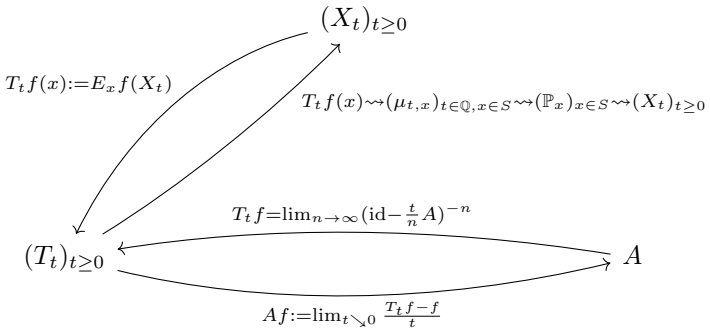

Fig. 3.2: One-to-one correspondence between Feller processes, probability semigroups and generators.

Remark 3.1.12 (On Fig. 3.2) The idea for the direction from probability semigroups to Feller processes is to use Riesz' representation theorem to induce probability measures $\mu_{t,x}$ over rational times and all of $x \in S$ as $T_t f(x) = \langle \mu_{t,x}, f \rangle$, and then to use a martingale result to lift them to a Feller process.

3.2 Dynkin's Formula and the Mean Value Property

Before we state the main result for this section, we need to recall a concept from probability theory: martingales. Rick Durrett describes them as "the fortune at time n of a player who is betting on a fair game" ([D], p. 205). By this, he means that these stochastic processes have constant expected value, as we shall prove.

Definition 3.2.1 ([G], Def. 5.1, Continuous Time) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_t)_{t\geq 0} \subseteq \mathcal{F}$ be a right-continuous filtration. A **martingale** with respect to $(\mathcal{F}_t)_{t\geq 0}$ is a family of maps $(M_t)_{t\geq 0}$ such that for all $t \geq 0$:

(M1) $M_t \in \mathcal{F}_t$.

(M2) $M_t \in L^1(\mathbb{P})$.

(M3) $\forall s \in [0,t] : E(M_t \mid \mathcal{F}_s) = M_s$.

Lemma 3.2.2 Martingales have constant expected value.

Proof: The conditional expectation of a random variable M with respect to any sub σ-field $\mathcal{F}$ is $\mathcal{F}$-measurable, since any sub σ-field $\mathcal{F}$ contains the ground set Ω. More explicitly, recall the definition of the conditional expected value of M with respect to $\mathcal{F}$:

$$\int_A E(M|\mathcal{F})\,\mathrm{d}\mathbb{P} = \int_A M\,\mathrm{d}\mathbb{P}$$

for all $A \in \mathcal{F}$. Setting $A = \Omega$ yields $E[E(M|\mathcal{F})] = EM$. Now, for all $s \in [0,t]$, we have using (M3):

$$EM_t = E[E(M_t|\mathcal{F}_s)] = EM_s\,.$$

□

The following result allows us to generate a martingale from a Feller process. It is an important statement in itself; here, we will only use it as an intermediate result to get to the so-called mean value property of a stationary solution u_∞ of an evolution equation. An example of an operator to keep in mind while reading the following result is that of Brownian motion $(B_t)_{t\geq 0}$, which has the Gaussian kernel $p_t(x,y) := g_t(x-y)$ as a convolution kernel, together with the weighted Laplace operator $A = \frac{1}{2}\Delta$ as a generator.

Theorem 3.2.3 (Generalized Dynkin's Formula) Let $(X_t)_{t\geq 0}$ be a Feller process with corresponding probability measures $(\mathbb{P}_x)_{x\in S}$, let A its corresponding generator, with $T_t f(x) = (p_t * f)(x)$ as its probability semigroup, where the convolution is given with respect to some measure μ over $\mathcal{B}(\mathbb{R}^d)$, i.e.

$$T_t f(x) = \int_{\mathbb{R}^d} f(y) p_t(x,y)\,\mathrm{d}\mu(y)$$

and p_t satisfies the evolution equation

$$\partial_t p_t(x,y) = A p_t(x,y) \tag{3.6}$$

for all $t > 0$, $x \in S$. In addition, we assume that p_t has the following limiting behaviour

$$\lim_{t\searrow 0} p_t(x,y)\,d\mu(y) = \mathrm{d}\delta_x(y)\,.$$

Moreover, assume $f : \mathbb{R}^d \to \mathbb{R}$ is a suitably smooth function which fulfills the integrability conditions:

$$f(X_t),\ \int_{[0,t]} |Af(X_s)|\,\mathrm{d}\lambda(s) \in L^1(\mathbb{P}_x)$$

and that the generator A is symmetric with respect to μ in the following sense:

$$\int_{\mathbb{R}^d} p_t(x,y) A f(y)\,\mathrm{d}\mu(y) = \int_{\mathbb{R}^d} f(y) A p_t(x,y)\,\mathrm{d}\mu(y) \tag{3.7}$$

for all $x \in \mathbb{R}^d$ and $t \geq 0$.

Then,

$$M_t := f(X_t) - \int_{[0,t]} Af(X_s)\,\mathrm{d}\lambda(s) \tag{3.8}$$

is a martingale with respect to the filtration induced by the Feller process $(X_t)_{t\geq 0}$.

Remark 3.2.4 Notice we do not require that $f \in D(A)$; this would restrict us a lot because of the requirement that all these functions necessarily vanish at infinity (cf. definition of $\mathcal{C}_0(S)$). For instance, take $A = \frac{1}{2}\Delta$ with domain $D(A) \subseteq \mathcal{C}_0(\mathbb{R})$. Then, $f = x^2$ is in $\mathcal{C}^\infty(\mathbb{R})$; in particular, we can take two derivatives of f. We have $\frac{1}{2}\Delta f = 1$, which does not vanish at infinity, which means $\frac{1}{2}\Delta f \notin \mathcal{C}_0(\mathbb{R})$. In other words, f does not lie in $D(A)$. However, since the conditions from Thm. 3.2.3 are fulfilled,[1] we can still apply the above formula (3.8) to get a martingale:

$$M_t = B_t^2 - t$$

for all $t \geq 0$, where $(B_t)_{t\geq 0}$ denotes Brownian motion.

Proof: The following generalizes the setting in ([B], T8.1). The (several) additional assumptions required stem from the general nature of the operator A. Define $(M_t)_{t\geq 0}$ as in (3.8). The property (M1) is clear and (M2) is satisfied by assumption. To show (M3), we take some $s \in [0,t]$ and recall that we denote by E_x the expected value with respect to the measure $\mathbb{P}_x$ for $x \in S$, and $(\mathcal{F}_t)_{t\geq 0}$ is the corresponding filtration to the Feller process $(X_t)_{t\geq 0}$. We have

$$\begin{aligned} E(M_t|\mathcal{F}_s) &= E\left[f(X_t) - \int_{[0,s]} Af(X_u)\,\mathrm{d}\lambda(u) - \int_{[s,t]} Af(X_u)\,\mathrm{d}\lambda(u)\middle|\mathcal{F}_s\right] \\ &= E_{X_s} f(X_{t-s}) - \int_{[0,s]} Af(X_u)\,d\lambda(u) - \int_{[0,t-s]} E_{X_s} Af(X_u)\,d\lambda(u) \end{aligned}$$

where we have used Fubini's theorem and the Markov property. Using that p_t satisfies (3.6) together with the symmetry condition (3.7), we get that the last integrand can be rewritten as

$$E_{X_s} Af(X_u) = \int_{\mathbb{R}^d} p_u(X_s,y) Af(y)\,\mathrm{d}\mu(y) = \int_{\mathbb{R}^d} f(y) Ap_u(X_s,y)\,\mathrm{d}\mu(y) = \int_{\mathbb{R}^d} \partial_u p_u(X_s,y)\,\mathrm{d}\mu(y)\,.$$

Using that $\lim_{\epsilon\searrow 0} p_\epsilon(x,y)\,d\mu(y) = \mathrm{d}\delta_x(y)$, we get:

$$\begin{aligned} \int_{[0,t-s]} E_x Af(X_u)|_{x=X_s}\,\mathrm{d}\lambda(u) &= \lim_{\epsilon\searrow 0} \int_{\mathbb{R}^d} \int_{[\epsilon,t-s]} \partial_u p_u(X_s,y)\,\mathrm{d}\lambda(u)\,\mathrm{d}\mu(y) \\ &= \int_{\mathbb{R}^d} p_{t-s}(X_s,y) f(y)\,\mathrm{d}\mu(y) - \lim_{\epsilon\searrow 0} \int_{\mathbb{R}^d} p_\epsilon(X_s,y) f(y)\,\mathrm{d}\mu(y) \\ &= E_{X_s} f(X_{t-s}) - f(X_s). \end{aligned}$$

[1] The density g_t for normal random variables satisfies the heat equation $\partial_t g_t(x) = \frac{1}{2}\Delta g_t(x)$ and the integrability conditions are clearly satisfied.

We conclude

$$\begin{aligned} E_x(M_t|\mathcal{F}_s) &= E_{X_t} f(X_{t-s}) - \int_{[0,s]} Af(X_u)\,\mathrm{d}\lambda(u) - E_{X_t} f(X_{t-s}) + f(X_s) \\ &= f(X_s) - \int_{[0,s]} Af(X_u)\,\mathrm{d}\lambda(u) \\ &= M_s\,. \end{aligned}$$

□

The power of stationary solutions with regards of the above result is made clear in the following corollary.

Corollary 3.2.5 (Martingale Mean Value Property) Consider the setting of Thm. 3.2.3, and let $u_\infty \in \ker A$ be a stationary solution to (1.3) such that $u_\infty(X_t) \in L^1(\mathbb{P}_x)$ for all $x \in S$ and $t \geq 0$. Then,

$$(u_\infty(X_t))_{t\geq 0}$$

is a martingale. In particular, stationary solutions can be written as the expected value of u_∞ evaluated at a process X_t started at $x \in S$:

$$u_\infty(x) = \int_S u_\infty(X_t)\,\mathrm{d}\mathbb{P}_x$$

for any $t \geq 0$.

Proof: This follows from Thm. 3.2.3 and Lemma 3.2.2.

□

Remark 3.2.6 The above corollary in the case $A = \frac{1}{2}\Delta$ can be seen as a probabilistic and generalized version of the mean value property for harmonic functions: let u_∞ be harmonic over the set S. Then:

$$u_\infty(x) = \int_S u_\infty(B_t)\,\mathrm{d}\mathbb{P}_x$$

for all $t \geq 0$, where $(\mathbb{P}_x)_{x\in S}$ denotes the probability measures corresponding to Brownian motion started at $x \in S$ (cf. Def. 3.1.2). Intuitively, one can see this special case as fixing a point in state space $x \in S$, and then testing out the desired harmonic function evaluated at Brownian motion started at that point in space, in order to get the value of the function at that point.

The above special case should not be confused with the probabilistic formula for the solution of the Dirichlet problem of the Laplace equation:

$$u_\infty(x) = \int_S \phi(B_{H_{\partial S}})\mathbb{1}_{H_{\partial S}<\infty}\,\mathrm{d}\mathbb{P}_x$$

where $\phi : \partial S \to \mathbb{R}$ is the given boundary condition, and $H_{\partial S}$ is the hitting time of Brownian motion started at $x \in S$ to the boundary.

3.3 Stationary and Reversible Measures

Let $(X_t)_{t\geq 0}$ be a Feller process, and $(T_t)_{t\geq 0}$ and A be the corresponding probability semigroup and generator, respectively. For every $t \geq 0$, we define a measure μ_t over a locally compact space S by

$$B \mapsto \mu_t(B) := \int_S \mathbb{P}_x(X_t \in B)\, \mathrm{d}\mu(x)\,. \tag{3.9}$$

Recall that in the partial differential equations considered in the previous chapters, we have assumed a *normalized* initial condition $||u_0||_X = 1$, which is preserved throughout the evolution in time. A similar phenomenon will happen in this section for the evolution of the measure counterpart $(\mu_t)_{t\geq 0}$ which is induced by the Feller process and therefore, by the Hille-Yosida theorem, also by a linear evolution equation with operator A.

Definition 3.3.1 ([B], Def. 7.1) A non-zero measure $\mu_\infty \in \mathcal{M}(S)$ is said to be **invariant** or **stationary** under the Feller process $(X_t)_{t\geq 0}$, if

$$\forall t \geq 0 : \ \mu_t = \mu_\infty\,.$$

We write $\mathcal{I}$ for the set of all stationary measures of a given Feller process and $\mathcal{I}_1 \subseteq \mathcal{I}$ for the set of stationary probability measures.

It is often useful to rewrite stationarity in terms of the corresponding probability semigroup or generator.

Theorem 3.3.2 (Stationarity, [B] Thm. 7.2) The following statements are equivalent:

1. The measure $\mu \in \mathcal{M}(S)$ is invariant under $(X_t)_{t\geq 0}$.
2. For all $f \in \mathcal{C}_0(S)$ and $t \geq 0$:
$$\int T_t f \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu\,.$$
3. For all $f \in D(A)$:
$$\int Af \,\mathrm{d}\mu = 0\,.$$

Proof: 1) $\iff$ 2): This is a matter of rewriting the definitions.

2) $\implies$ 3): Assume $f \in D(A)$. Then, using the dominated convergence theorem:

$$\int Af \, d\mu = \lim_{t\to 0} \frac{1}{t} \int (T_t f - f)\,\mathrm{d}\mu = 0\,.$$

3) $\implies$ 2): Since $D(A)$ is dense in $\mathcal{C}_0(S)$, it suffices to show the claim for some element $f \in D(A)$. We have again using the dominated convergence theorem, the Kolmogorov equation and Fubini's theorem:

$$\frac{d}{dt}\int T_t f \,\mathrm{d}\mu = \int \frac{d}{dt} T_t f \,\mathrm{d}\mu = \int T_t A f \,\mathrm{d}\mu = T_t\Big(\int Af \,\mathrm{d}\mu\Big) = 0\,.$$

□

Remark 3.3.3 The third condition in Thm. 3.3.2 can be interpreted from the point of view of the measure instead of functions in the domain of A. This perspective gives us the third reason we were hinting towards in Chapter 2 as to why stationary solutions are of interest. Indeed, viewing the integral as the action of an element

of the dual space, e.g. a measure space acting on $f \in \mathcal{C}_0(S)$:

$$\int f \, \mathrm{d}\mu = \langle \mu, f \rangle$$

we can characterize stationary measures μ_∞ by the condition:

$$A' \mu_\infty = 0 \tag{3.10}$$

where the apostrophe denotes the Banach space adjoint acting on a *measure* space. This follows because

$$0 = \int Af \ \mathrm{d}\mu_\infty = \langle \mu_\infty, Af \rangle = \langle A' \mu_\infty, f \rangle$$

for all $f \in D(A)$. Since $D(A)$ is dense in $\mathcal{C}_0(S)$ by (L1), we get (3.10). Moreover, assuming μ_∞ is absolutely continuous with respect to λ, i.e. $\mu_\infty \ll \lambda$, by the Radon-Nykodym theorem, we get some integrable function $u_\infty \in L^1(S, \lambda)$ with

$$d\mu_\infty = u_\infty d\lambda \,. \tag{3.11}$$

Therefore, again since $D(A)$ is dense in $\mathcal{C}_0(S)$ by (L1), we get using the fundamental lemma of calculus of variations:

$$A^* u_\infty = 0$$

where A^* is now the Banach space adjoint of A acting on a *function* space (cf. Ch. 2). We will soon see an example of such adjoints.

Now, let us consider the converse situation: given a normalized stationary solution u_∞ to the evolution equation (1.3), we can define a probability measure μ_∞ by (3.11). Then, essentially following the above steps in reverse, we see that μ_∞ is a stationary measure. Overall, we get a one-to-one correspondence between stationary solutions u_∞ to evolution equations (1.3) and stationary measures μ_∞ of the corresponding Feller process.

$$\begin{array}{ccc} A & \xleftarrow{\text{Hille-Yosida}} & (X_t)_{t\geq 0} \\ \downarrow & & \downarrow \\ u_\infty & \xleftarrow{u_\infty = \frac{\mathrm{d}\mu_\infty}{\mathrm{d}\lambda}} & \mu_\infty \end{array}$$

The following examples show how to use the third characterization for stationarity given in Thm. 3.3.2.

Example 3.3.4 The Lebesgue measure is invariant for Brownian motion: Gauss' theorem of integration yields

$$\int \Delta f \, d\lambda = \lim_{R\to\infty} \oint_{\partial B_R(0)} \langle \nabla f, \nu \rangle \, dS = 0$$

for $f \in D\big(\frac{1}{2}\Delta\big) \subseteq \mathcal{C}_0(\mathbb{R}^d)$. △

Example 3.3.5 (Parts from [B], H. 5.3) Consider a bounded generator $A \in \mathcal{L}(\ell^2)$ for the Hilbert space $\ell^2 := \ell^2(\mathbb{N})$. We can write the set of stationary probability measures as

$$\mathcal{I}_1 = \ker(A^*) \cap \partial B_1^{||\cdot||_{\ell^1}}(0)$$

constraining ourselves to sequences with non-negative entries (i.e. we do not deal with signed measures with mass 1). This setting becomes even more interesting if we interpret each of the ℓ^2-spaces involved. The primal space, i.e. the one being acted upon by A, can be considered as a subset of a function space $D(A) \subseteq \{f : \mathbb{N} \to \mathbb{R} \mid$

$||f||_{\ell^2} < \infty\}$, which in this case is simply equal to ℓ^2 because the operator A is bounded. In the finite-dimensional case, we would write elements in this space as *column* vectors. Now, although elements in $\mathcal{I}_1 \subseteq (\ell^2)' \cong \ell^2$ are also sequences (using the Fréchet-Riesz representation theorem for Hilbert spaces), we can also interpret them as probability measures or, more specifically, as *densities* of a measure with respect to the counting measure. In particular, this interpretation says that the measure induced by this "invariant density" is absolutely continuous with respect to the counting measure. Nonetheless, whenever we refer to an invariant measure of a bounded operator over ℓ^2, we will mean the density. For instance, let $\mu_\infty \in \mathcal{I}_1$ be such a measure. Then,

$$A^*\mu_\infty = 0, \qquad \sum_{k\in\mathbb{N}} (\mu_\infty)_k = 1\,.$$

In the finite-dimensional case, we would write the measures as *row* vectors.

The name *stationary* measure comes from the fact that for a non-negative initial condition $(u_0^{(i)})_{i\in\mathbb{N}} \in \ell^2$ of unit mass (similar to the initial conditions we considered in Chapter 2), we have, denoting by $\mathbb{P}_i$ the probability measure of the Feller process with initial value $i \in \mathbb{N}$:

$$\lim_{t\to\infty} \mathbb{P}_i(X_t = j) = \lim_{t\to\infty} (T_t u_0^{(i)})_j = (\mu_\infty)_j\,.$$

Observe that this limit is independent of the initial state $i \in \mathbb{N}$. This intuitively means that the transients die out in the limit, and all that we are left with is the stationary measure. Notice the similarity of this behaviour to that of stationary solutions u_∞ of a partial differential equation from the last chapter.

Furthermore, similar to partial differential equations, where we need not necessarily have a normalized stationary solution, note that $\mathcal{I}_1$ could also be empty. For instance, the generator of the Poisson process over $\mathbb{N}$:

$$A = \lambda \begin{pmatrix} -1 & 1 & 0 & 0 & \dots \\ 1/2 & -1 & 1/2 & 0 & \dots \\ 0 & 1/2 & -1 & 1/2 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

where $\lambda > 0$ has $\mu = 0$ as its only element in $\ker(A^*)$, and therefore $\mathcal{I}_1 = \emptyset$. △

Remark 3.3.6 The notation we introduced $\langle \mu, f\rangle$ is compatible with the case of an ℓ^p action product $\langle x, y\rangle = \sum_{n\in\mathbb{N}} x_n y_n$ for $x \in \ell^p$ and $y \in \ell^q$ where $p \geq 1$ and $p^{-1} + q^{-1} = 1$. This is the reason we mentioned the *column* and *row* vector perspective.

Example 3.3.7 ([B], H. 11.3) The finite-dimensional operator

$$A = \begin{pmatrix} -1 & 1/2 & 1/2 \\ 1/3 & -1 & 2/3 \\ 1/2 & 1/2 & -1 \end{pmatrix}$$

has a unique stationary probability measure:

$$\mu_\infty = \frac{1}{27}(8, 9, 10)\,.$$

One can simply see this by checking that $\mu_\infty A = 0$. The following questions arise:

Does there always exist a stationary measure in the finite-dimensional case?
If so: is it unique when we require it to be a probability measure?

The answer to both questions is yes, and the proof is given by the Frobenius-Perron theorem (cf. [R], Thm. 4.2), which is given in the language of the corresponding semigroup $(T_t)_{t\geq 0}$. In fact, this theorem states that there exists a unique unit-mass left eigenvector $\mu_\infty \in (\mathbb{R}^n)^*$ to T_1 with eigenvalue 1:

$$\mu_\infty = \mu_\infty T_1 \,.$$

Equivalently, by the Kolmogorov equation (3.4):

$$0 = \mu_\infty A$$

assuming that the matrix describing T_1 is irreducible, i.e. that for every pair of indices $i, j \in \{1, ..., d\}$ there exists some integer n such that T_1^n allows the possibility for a transition from i to j, i.e. $(T_1^n)_{ij} > 0$. Notice that existence of a left eigenvector is simple from the definition of a generator:

$$\ker(A^*) \neq \{0\} \iff 0 \in \sigma(A^*) = \overline{\sigma(A)} = \sigma(A) \iff 0 \in \ker(A)$$

and since $\sum_{j=1}^d (A)_{ij} = 0$, the claim follows. Uniqueness is more involved and will not be addressed here. △

Remark 3.3.8 In the context of partial differential equations, the above questions translate to:

Does the operator A always have a non-trivial kernel?
If so: is this element unique if we also require normalization by 1?

The answer to both questions is no. To see why it is so for the first one, consider the identity operator $A = \mathrm{id}$. Then, $\ker(A) = \{0\}$, i.e. A has a trivial kernel. For the second question, we consider $A = \Delta$ over a bounded, open nonempty interval $S = (a, b) \subseteq (0, \infty)$. Clearly $\langle ax + b \mid a, b \in \mathbb{R}\rangle \subseteq \ker(A) \subseteq D(A)$, but then, we have for instance that

$$u_\infty^{(1)} = \frac{1}{b-a} \neq \frac{2}{b^2 - a^2} x = u_\infty^{(2)}$$

are two different, normalized elements of $\ker(\Delta)$.

We briefly discuss reversible measures from the stochastic point of view. We shall proceed with their treatment in a longer version of this paper.

Notation 3.3.9 For $f, g \in L^2(S, \mathcal{A}, \mu)$, we denote by $(f, g)_\mu$ the (real) scalar product in $L^2(S, \mathcal{A}, \mu)$, i.e.

$$(f, g)_\mu := \int_S fg \, d\mu \,.$$

Whenever it is clear which measure we are using, we shall omit it in the notation. Note that we distinguish this bilinear form from the action bracket we had in (3.1.1) by using $(\cdot, \cdot)$ instead of $\langle \cdot, \cdot \rangle$.

Definition 3.3.10 ([B], Def. 7.3) A measure $\mu \in \mathcal{M}_1(S)$ is called **reversible** under the Markov process $(\mathbb{P}_x)_{x\in S}$, if the corresponding generator A is self-adjoint in $L^2(\mu)$:

$$(f, Ag)_\mu = (Af, g)_\mu \,.$$

Lemma 3.3.11 (Reversibility) The following statements are equivalent:

1. The measure $\mu \in \mathcal{M}_1(S)$ is reversible.
2. The family $(T_t)_{t\geq 0}$ is self-adjoint in $L^2(\mu)$.

Proof: 2) $\implies$ 1): We have

$$(f, Ag) = \lim_{t\to 0} \frac{1}{t}(f, T_t g - g) = \lim_{t\to 0} \frac{1}{t}(T_t f - f, g) = (Af, g)\,.$$

1) $\implies$ 2): Using the fundamental theorem of calculus and the Kolmogorov equation, we get

$$\begin{aligned}(f, T_t g) - (T_t f, g) = \int_{[0,t]} \frac{d}{ds}(T_s g, T_{t-s} f)\, ds &= \int_{[0,t]} (\frac{d}{ds} T_s g, T_{t-s} f) - (T_s g, \frac{d}{ds} T_{t-s} f)\, ds \\ &= \int_{[0,t]} (AT_s g, T_{t-s} f) - (T_s g, AT_{t-s} f)\, ds = 0\,.\end{aligned}$$

□

Proposition 3.3.12 ([B], L. 7.4) If a measure is reversible, it is also stationary.

Proof: Let $f \in D(A)$ and $(g_n)_{n\in\mathbb{N}} \subseteq D(A)$ be a sequence which converges pointwise to 1 and $\sup_{n\in\mathbb{N}} ||g_n|| < \infty$ given by (S5). Then, using the dominated convergence theorem,

$$\int_S Af\, d\mu = (Af, \lim_{n\to\infty} g_n) = \lim_{n\to\infty} (Af, g_n) = \lim_{n\to\infty} (f, Ag_n) = (f, \lim_{n\to\infty} Ag_n) = 0\,.$$

□

3.4 Poincaré Inequalities

We now consider special types of inequalities which describe how solutions decay to equilibrium.

Definition 3.4.1 ([Ba], p. 177) A **Markov triple** (S, μ, Γ) consists of a state space S, a finite invariant, reversible measure μ over Ω, and a **carré du champ operator** Γ corresponding to A:

$$\mathcal{A} \ni (f, g) \mapsto \Gamma(f, g) = \frac{1}{2}(A(fg) - fA(g) - gA(f))$$

where $\mathcal{A}$ is an algebra of bounded measurable functions.

We call

$$D(\mathcal{E}) \ni f \mapsto \mathcal{E}(f) = \int_S \Gamma(f)\, d\mu$$

the corresponding **Dirichlet form** to Γ.

Definition 3.4.2 ([Ba], Def. 4.2.1) A Markov triple (S, μ, Γ) is said to satisfy a **Poincaré inequality** with constant $C > 0$, if for all $f \in D(\mathcal{E})$, we have

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

We refer to the smallest $C > 0$ as the **Poincaré constant**.

Such inequalities control the convergence to equilibrium of the CSG $(T_t)_{t\geq 0}$ towards the invariant measure.

Example 3.4.3 (Ornstein-Uhlenbeck: Poincaré Inequality, [Ba], Sec. 4.1) Consider a one-dimensional special case of the Ornstein-Uhlenbeck generator A

$$Af = f'' - xf'$$

with corresponding carré du champ operator $\Gamma(f) := \Gamma(f,f) = f'^2$. Using partial integration twice, one can show that the standard Gaussian probability measure μ is reversible with respect to A:

$$\int f(g'' - xg')\exp(-x^2/2)\,d\lambda = -\int f'g'\exp(-x^2/2)\,d\lambda = \int g(f'' - xf')\exp(-x^2/2)\,d\lambda$$

for $f, g \in L^2(\mu)$. Therefore, the Dirichlet form is given by $\mathcal{E}(f) = \int f'^2\,d\mu$. Using results from ([Ba], Sec. 2.7.1) the Hermite polynomimals $(h_k)_{k\in\mathbb{N}}$ form a orthonormal basis in $L^2(\mu)$, so we can write

$$f = \sum_{k\in\mathbb{N}} (f, h_k)_\mu h_k\,.$$

We write $a_k = (f, h_k)_\mu$. Using reversibility of μ with respect to $(T_t)_{t\geq 0}$, we get for the coefficients of $T_t f$

$$(T_t f, h_k) = \int T_t f h_k\,d\mu = \int f T_t h_k\,d\mu = \exp(-kt)\int f h_k\,d\mu$$

since $Ah_k = -kh_k$ whence $T_t h_k = \exp(-kt)h_k$. Thus,

$$T_t f = \sum_{k\in\mathbb{N}} \exp(-kt) a_k h_k\,.$$

Hence, $T_t f$ converges in $L^2(\mu)$ towards $(f, h_0) = E_\mu f$ as $t \to \infty$. Equivalently, the corresponding Markov process X_t converges to the invariant measure μ. We get the following inequality:

$$||T_t(f - E_\mu f)||^2_{L^2(\mu)} = \sum_{k\geq 1} \exp(-2kt)a_k^2 \leq \exp(-2t)\sum_{k\geq 1} a_k^2 = \exp(-2t)||f - E_\mu f||^2_{L^2(\mu)}\,.$$

We first assume that $E_\mu f = 0$ and then recenter and use the fact that $\mathcal{E}(f + c) = \mathcal{E}(f)$. Using the Taylor expansion at $t = 0$ and using the exponential formulae $T_t f = f + tAf + o(t)$ and $\exp(-2t) = 1 - 2t + o(t)$, we get

$$E_\mu f^2 - 2t\mathcal{E}(f) + o(t) \leq E_\mu f^2(1 - 2t + o(t))$$

yielding after recentering and invariance under additive constants in the argument of $\mathcal{E}$:

$$\mathrm{Var}_\mu(f) \leq \mathcal{E}(f)\,. \tag{3.12}$$

△

Using similar but multidimensional arguments, one gets the following result:

Theorem 3.4.4 ([Ba], Prop. 4.1.1) Let μ be the standard Gaussian measure for the Borel sets of $\mathbb{R}^d$. Then,

for all $f \in D(\mathcal{E})$, we get

$$\int f^2 \, d\mu - \left(\int f \, d\mu \right)^2 = \mathrm{Var}_\mu(f) \leq \mathcal{E}(f) = \int ||\nabla f||^2 \, d\mu \, .$$

Remark 3.4.5 Poincaré inequalities not only describe an upper bound in the decay to equilibrium, but also a gap in the spectrum of the operator A. For instance, using the above $\mathcal{E}$ and assuming $-Af = \lambda f$ for some $\lambda \neq 0$ as well as some Poincaré constant $C > 0$, we get with (3.12):

$$\int f^2 \, d\mu \leq C\lambda \int f^2 \, d\mu$$

where we used that $E_\mu f = 0$ by invariance and $\lambda \neq 0$. Therefore, $C\lambda \geq 1$ and thus, by self-adjointness and definiteness of A,

$$\sigma(A) \subseteq \{0\} \cup [1/C, \infty) \, .$$

3.5 Bakry-Emery: Stochastic Analysis

Let S be a locally compact space. In this section, we assume that the considered Feller process $(X_t)_{t \geq 0}$ admits a unique stationary measure $\nu \in \mathcal{M}(S)$. The change of viewpoint from partial differential equations to stochastics relies on probability semigroups admitting densities $p_t = p_t(x, y)$ with respect to some reference measure μ:

$$T_t f(x) = \int_S p_t(x, y) f(y) \, d\mu(y)$$

similar to what we had in Thm. 3.2.3. An example is given by $p_t(x, y) = g_t(x - y)$ with $\mu = \lambda$, where g_t denotes the Gaussian density mentioned in the previous section. Using the Kolmogorov equation (3.4), we see

$$\int_S \partial_t p_t(x, y) f(y) \, d\mu(y) = \frac{d}{dt} T_t f(x) = A_x T_t f(x) = \int_S f(y) A_x p_t(x, y) \, d\mu(y) \tag{3.13}$$

in other words,

$$\partial_t p_t(x, y) = A_x p_t(x, y), \qquad p_0(x, y) \, d\mu(y) = \mathrm{d}\delta_x(y)$$

for all $y \in S$, where A_x is the generator A acting on the first component of p_t (cf. partial differential equation condition for p_t in Thm. 3.2.3). We can formally obtain the dual equation by interchanging the derivative in (3.13) using suitable smooth and vanishing functions f,

$$\int_S \partial_t p_t(x, y) f(y) \, d\mu(y) = \int_S p_t(x, y) A_y f(y) \, d\mu(y) = \int_S f(y) A'_y p_t(x, y) \, d\mu(y)$$

thereby getting

$$\partial_t p_t(x, y) = A'_y p_t(x, y)$$

for all $x, y \in S$ and $t \geq 0$.

Example 3.5.1 (Brownian Motion) As a consequence of Gauss' theorem of integration, the Laplace operator is self-adjoint with respect to the scalar product in $L^2(\lambda)$ and therefore it does not play a role whether we consider the primal or dual equation. △

Example 3.5.2 (Dual Operator: Ornstein-Uhlenbeck, [J], Exa. 2.4, p. 28) Consider the generator A defined by

$$Af = \Delta f - \langle \nabla V, \nabla f \rangle$$

for suitable $f : \mathbb{R}^d \to \mathbb{R}$. Under certain regularity assumptions, the associated Cauchy-problem has a unique solution. Let ν be an invariant measure and assume it is absolutely continuous with respect to the Lebesgue measure, so that the Radon-Nykodym derivative $u_\infty = \mathrm{d}\nu / \mathrm{d}\lambda$ exists. Then, by Thm. 3.3.2, we have

$$0 = \int_{\mathbb{R}^d} Af \, d\nu = \int_{\mathbb{R}^d} (A^* u_\infty) f \, d\lambda$$

where $A^* u = \nabla \cdot (\nabla u + u \nabla V)$ is the adjoint with respect to the $L^2(\mathbb{R}^d)$ scalar product over smooth-enough admissible functions which vanish at infinity. Note that A^* is the same operator considered in Exa. (2.1), which explains the notation we have been carrying throughout for the latter.

Assuming that the set of admissible functions f is dense in $L^2(\mathbb{R}^d)$, e.g. assuming that (L1) is fulfilled, we can conclude by the fundamental lemma of calculus of variations (as we did in Rmk. 3.3.3) that

$$A^* u_\infty = 0 \, .$$

Therefore, u_∞ corresponds to the steady state of the linear Fokker-Planck equations (2.1). By Exa. 2.2.1, we even know that $u_\infty(x) = Z \exp(-V(x))$, where $Z > 0$ is a constant which ensures that u_∞ has unit mass. Hence, under all the above assumptions, we conclude that the unique invariant measure for T_t is given by $\mu_\infty := \nu$ with

$$\mathrm{d}\mu_\infty = u_\infty \, \mathrm{d}\lambda \, .$$

△

Going back to Chapter 2, we note that Bakry and Emery stated the convexity condition from Thm. 2.2.3 in (2.2) of the potential using the carré du champ operator and the **gamma-deux operator** Γ_2:

$$\Gamma_2(f, g) := \frac{1}{2} (A\Gamma(f, g) - \Gamma(Af, g) - \Gamma(f, Ag)) \, .$$

For instance, in the above example, we have

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

and

$$\Gamma_2(f) := \Gamma_2(f, f) = ||\nabla^2 f||^2 + \langle \nabla f, \nabla f \rangle_{\nabla^2 f} \tag{3.14}$$

and therefore, if V is strictly convex with constant $\lambda > 0$, we have $\Gamma_2(f) \geq \lambda \Gamma(f)$. One can show another convex Sobolev inequality under the abstract Bakry-Emery condition:

Theorem 3.5.3 ([J], Thm. 2.2) Let $\phi \in \mathcal{C}^2([0, \infty))$ be convex with $1/\phi''$ concave. Assume there exists some $\lambda > 0$ such that for all suitable non-negative functions f

$$\Gamma_2(f) \geq \lambda \Gamma(f) \tag{3.15}$$

holds. Then,

$$\int \phi(f)\,d\mu_\infty - \phi\Big(\int f\,d\mu_\infty\Big) \leq \frac{1}{2\lambda}\int \phi''(f)\Gamma(f)\,d\mu_\infty\,.$$

Proof: The calculations are similar to those performed in the proof of Thm. 2.2.3, so we just state the important points that lead to the result. We consider the entropy functional

$$H_\phi[f] = \int \phi(f)\,d\mu_\infty\,.$$

By assumption $\phi'' \geq 0$ and teh carré du champ operator is positive, i.e. $\Gamma(f) \geq 0$ for non-negative f with vanishing support at infinity. Therefore,

$$\frac{d}{dt}H_\phi[T_t f]|_{t=0} = -\int \phi''(f)\Gamma(f)\,d\mu_\infty \leq 0\,.$$

Moreover, using the chain rule of the carré du champ operator $\Gamma(\phi'(f)) = \phi''(f)^2\Gamma(f)$ as well as (3.15), we conclude after lengthy calculations

$$\begin{aligned}\frac{d^2}{dt^2}H_\phi[T_t f]|_{t=0} &\geq 2\int \phi''(f)^{-1}\Gamma_2(\phi'(f))\,d\mu_\infty \geq 2\lambda\int \phi''(f)^{-1}\Gamma(\phi'(f))\,d\mu_\infty \\ &= 2\lambda\int \phi''(f)\Gamma(f)\,d\mu_\infty = -2\lambda\frac{d}{dt}H_\phi[T_t f]|_{t=0}\,.\end{aligned}$$

Arguing as in Thm. 2.2.3, we conclude the result.

□

Example 3.5.4 Depending on the given problem statement, one then chooses which is the best approach to take: stochastic, analytic or semigroup-theoretic. For instance, consider

$$\partial_t u = \langle b, u\rangle + \frac{1}{2}\langle u, u\rangle_A \tag{3.16}$$

for some smooth mappings $x \mapsto b(x)$ and $x \mapsto A(x) \geq 0$ (i.e. pointwise positive semidefinite) with smooth square root $\sqrt{A}$, where $(\sqrt{A})^2 = A$. Then, we have that

$$(x,t) \mapsto u(x,t) = T_t f(x) = E_x f(x + tb + \sqrt{A}B_t)$$

solves (3.16), where B_t denotes standard Brownian motion. This works quite nicely with the semigroup-approach. Nevertheless, consider the discrete time simple random walk described by

$$\mathbb{P}(X_{n+1} = y | X_n = x) = \frac{1}{2}\mathbb{1}_{|x-y|=1} \tag{3.17}$$

for $n \in \mathbb{N}$, and for $t \in \mathbb{R}\setminus\mathbb{N}$, we set the process constant. Then, one can show that the corresponding semigroup is given by

$$T_t f(x) = E_x f(X_t) = \sum_{y=-\lfloor t\rfloor}^{\lfloor t\rfloor} 2^{-\lfloor t\rfloor}\binom{\lfloor t\rfloor}{(\lfloor t\rfloor + y)/2} f(x+y)\mathbb{1}_{\lfloor t\rfloor \equiv y \bmod 2}\,.$$

for $t \geq 0$. The complexity of the above expression makes the stochastic viewpoint more favourable in this case, since it even provides more information at a first glance. One can therefore conclude that staying in the stochastic perspective (3.17) might be more convenient than solving the corresponding Kolmogorov equation directly using

the above probability semigroup, let alone calculating the underlying generator A. △

Summary and Conclusion

Entropy methods allow us to get an upper bound in the rate of convergence of solutions to evolution equations towards stationary solutions with respect to a certain "distances" called entropy functionals. Advantages against spectral methods were used as a motivation to introduce entropy methods. These include, for instance, being able to use other distance notions for the measure of decay of the solution to the stationary state and not requiring a linear(ized) evolution equation. The Bakry-Emery approach is a special case of an entropy method in which the exponential decay and functional inequality steps are done in one step. We showed a general Sobolev inequality using this approach (cf. Thm. 2.2.3), and in the proof, we saw the importance of smoothness and convexity of the entropy functional as an assumption.

In addition, the importance of stationary solutions was motivated in three ways: they were viewed as equilibrium points in a function space with respect to corresponding entropy functionals; as self-referencing objects using the integral operator and involving their corresponding Feller process; and as predual objects of stationary measures. We analyzed the latter one thoroughly when we considered the Bakry-Emery approach from two viewpoints: analytic and stochastic. Indeed, here, we first looked at the partial differential equation side and then, motivated by *Riesz' representation theorem*, we considered its measure counterpart. We have seen that under the assumptions of absolute continuity, in order to use the *Radon-Nykodym theorem*; and dense domain $D(A)$, in order to use the *fundamental lemma of calculus of variations*, both viewpoints are equivalent in the sense that the relation

$$\mathrm{d}\mu_\infty = u_\infty \, \mathrm{d}\lambda$$

– where both $Au_\infty = 0$ and $A'\mu_\infty = 0$ – yields a one-to-one correspondence between stationary solutions and stationary measures, under the assumption of a unique, normalized, non-constant solution to $Au = 0$. Finally, as seen in Exa. 3.5.4, depending on the given problem statement, one then chooses which is the best approach to take: stochastic, analytic, or semigroup-theoretic.

References

- [B] N. Berger: Markov Processes, Lecture Notes 2023.
- [Ba] D. Bakry, I. Gentil, M. Ledoux: Analysis and Geometry of Markov Diffusion Operators, Springer 2013.
- [C] T. Carter: An Introduction to Information Theory and Entropy, 2014.
- [D] R. Durrett: Probability: Theory and Examples, Version 5, January 2019.
- [F] G. Friesecke: Partial Differential Equations, Lecture Notes 2022.
- [G] N. Gantert: Probability Theory, Lecture Notes 2022.
- [J] A. Jüngel: Entropy Methods for Diffusive Partial Differential Equations, Springer, 2016.
- [K] C. Kuehn: Dynamical Systems, A Coincise Graduate Course, Lecture Notes 2022.
- [L] T. M. Liggett: Continuous Time Markov Processes, An Introduction, AMS, 2010.
- [P] A. Pazy: Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer 1983.
- [R] M. Rasmussen: Random Dynamical Systems, Lecture Notes 2023.
- [Ru] W. Rudin, Real and Complex Analysis, third ed., McGraw-Hill Book Co., New York, 1987.