Semigroup methods for large deviations of Markov processes

Kraaij, Richard

DOI
10.4233/uuid:25eca1e5-b077-40c9-a216-d74026d4e887

Publication date
2016

Document Version
Final published version

Citation (APA)
Kraaij, R. (2016). *Semigroup methods for large deviations of Markov processes*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:25eca1e5-b077-40c9-a216-d74026d4e887

SEMIGROUP METHODS FOR LARGE DEVIATIONS OF MARKOV PROCESSES

RICHARD CLEMENS KRAAIJ

COLOPHON

Richard Clemens Kraaij

Large deviations for Markov processes, Delft, 2016

ISBN: 978-94-6233-313-0

Cover art:

Lun-Yi London Tsai

Almost Vector Bundle, 2005, oil on acrylic on canvas, 75 x 120 cm

www.londontsai.com

Cover design by Jorine Eeftens.

Printed by Gildeprint Drukkerijen, Enschede.

This document was typeset in *Linux Libertine* using *classicthesis* developed by André Miede.

SEMIGROUP METHODS FOR
LARGE DEVIATIONS OF MARKOV PROCESSES

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE TECHNISCHE
UNIVERSITEIT DELFT OP GEZAG VAN DE
RECTOR MAGNIFICUS PROF. IR. K.C.A.M. LUYBEN,
VOORZITTER VAN HET COLLEGE VOOR PROMOTIES,

IN HET OPENBAAR TE VERDEDIGEN OP 9 SEPTEMBER 2016 OM 15:00 UUR

DOOR

RICHARD CLEMENS KRAAIJ

MASTER OF SCIENCE
VRIJE UNIVERSITEIT AMSTERDAM, NEDERLAND

GEBOREN TE TÖNISVORST, DUITSLAND

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. F.H.J. Redig en Prof. dr. W.Th.F. den Hollander

Samenstelling van de promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. F.H.J. Redig	Technische Universiteit Delft
Prof. dr. W.Th.F. den Hollander	Universiteit Leiden
Onafhankelijke leden:	
Prof. dr. J.M.A.M. van Neerven	Technische Universiteit Delft
Prof. dr. C. Külske	Ruhr-Universität Bochum
Prof. dr. C. Maes	Katholieke Universiteit Leuven
Prof. dr. J. Feng	University of Kansas
Prof. dr. ir. G. Jongbloed	Technische Universiteit Delft, reservelid

Dit project werd gefinancierd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), beurs nummer 600.065.130.12N109.

CONTENTS

Part I

INTRODUCTION

1 INTRODUCTION

This thesis focusses on two aspects of large deviations for Markov processes:

- Proving large deviation principles for trajectories of Markov processes,
- Using Hamiltonian dynamics to study trajectories that have minimal Lagrangian cost.

Additionally, to facilitate the understanding of Markov processes on Polish spaces in relation to functional analytic techniques:

- The study of strongly continuous semigroups on the space of bounded continuous functions with the strict topology.

In this chapter, we introduce the main ideas behind large deviation principles for Markov processes.

1.1 LARGE DEVIATIONS FOR MARKOV PROCESSES

1.1.1 *Coin tosses and large deviations*

A well known principle in the process of coin tossing is the fact that the coin lands heads about half of the cases. This averaging principle also shows up with card games, roulette, and various other games of chance.

This common knowledge can be made mathematically rigorous and is called the *law of large numbers*. Suppose we model our sequence of coin tosses by a collection of random variables

$$X_n = \begin{cases} 0 & \text{if the } n\text{-th coin lands tail,} \\ 1 & \text{if the } n\text{-th coin lands heads.} \end{cases}$$

If the coin is fair, then the law of large numbers tells us that with probability one

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to \frac{1}{2}.$$

In other words, the coin lands on its head about half of the cases. To use this principle in practice, one needs to quantify how well the law of large numbers describes the average of these n coins if n is very large, but finite. One method is to study the asymptotics of the probability that the average is deviating from 0.5. In particular, one can prove that

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \approx \alpha\right] \approx e^{-nI(\alpha)}, \tag{1.1.1}$$

where $I(\alpha) = \alpha \log 2\alpha + (1-\alpha)\log 2(1-\alpha)$, see e.g. Dembo and Zeitouni [1998]. The $\approx$ signs can be made precise, but for the purposes here, it should be interpreted in the following way: the probability of the average $\frac{1}{n}\sum_{i=1}^{n} X_i$ to be close to a decays exponentially in n with rate $I(a)$. Note that we have $I(\alpha) = 0$ if and only if $\alpha = \frac{1}{2}$, the average that we expect from the law of large numbers.

A result like (1.1.1) is called a *large deviation principle* (LDP) with *rate function I*. This principle quantifies the leading order exponentially small probability of deviations from the law of large numbers behaviour. Such large deviation principles can be proven to apply in a wide range of settings.

1.1.2 *Large deviations of the average of Brownian trajectories*

Another setting where a large deviation principle applies is for the trajectory of averages of independent copies of Brownian motion. Consider a sequence of independent standard Brownian motions B_i on $\mathbb{R}$. For any fixed time $t \geq 0$, we know that $B_i(t)$ has a normal distribution with variance t and as a consequence we have a large deviation principle

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} B_i(t) \approx \alpha\right] \approx e^{-nI_{\mathcal{N}(0,t)}(\alpha)}, \tag{1.1.2}$$

where $I_{\mathcal{N}(0,t)}(\alpha) = \frac{\alpha^2}{2t}$. The interesting feature of stochastic processes is that the distributions for different times are correlated. It can be shown that for times $t_1 < t_2$, it holds that

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} B_i(t_1) \approx \alpha_1, \frac{1}{n}\sum_{i=1}^{n} B_i(t_2) \approx \alpha_2\right] \approx e^{-nI_{t_1,t_2}(\alpha_1,\alpha_2)}, \tag{1.1.3}$$

for some function I_{t_1,t_2} that we will define below. Because the value of the $B_i(t_2)$ clearly depends on $B(t_1)$, I_{t_1,t_2} is not the sum of the rates for the averages of $B_i(t_1)$ and $B_i(t_2)$.

Considering the trajectory of averages $\{L_n(t)\}_{t\geq 0}$, where

$$L_n(t) := \frac{1}{n}\sum_{i=1}^{n} B_i(t),$$

we also have a path-space large deviation principle: *Schilder's theorem*, Schilder [1966]. This result states that for any trajectory $\gamma : [0,\infty) \to \mathbb{R}$, we have the following exponential decay of the probability

$$\mathbb{P}\left[\{L_n(t)\}_{t\geq 0} \approx \gamma\right] \approx e^{-nI_S(\gamma)},$$

where

$$I_S(\gamma) = \begin{cases} \frac{1}{2}\int_0^\infty \dot{\gamma}(s)^2 \mathrm{d}s & \text{if } \gamma \text{ is absolutely continuous} \\ \infty & \text{otherwise.} \end{cases}$$

$\dot{\gamma}(s)$ denotes the derivative of the trajectory $s \mapsto \gamma(s)$, which exists almost everywhere due to the absolute continuity of γ. Thus, having a large speed for the average gives us a fast decay of probability on the exponential scale. The law of large numbers, which states that $L_n(t) \to 0$ almost surely for all t, is reflected in I_S as the zero trajectory has 0 cost.

From the path-space large deviation principle, we can recover the large deviation principles for individual times via the *contraction principle*. Thus, we are able to recover (1.1.2) from Schilder's theorem:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} B_i(t) \approx \alpha\right] \approx e^{-nJ(\alpha)},$$

where J is given by

$$J(\alpha) = \inf\left\{I_S(\gamma) \,|\, \gamma(t) = \alpha\right\}.$$

This rate function is given by a conditional version of I_S, where we are only interested in those trajectories that give the correct behaviour at time t, i.e. that end in α at time t. In this simple setting, we can explicitly find the minimizing trajectory $\gamma_{t,\alpha}$, which is given by a linear function:

$$\gamma_{t,\alpha}(s) = \begin{cases} s\frac{\alpha}{t} & \text{if } s \leq t \\ \alpha & \text{if } s \geq t. \end{cases}$$

A straightforward calculation yields

$$I_S(\gamma_{t,\alpha}) = \frac{1}{2}\int_0^t \left(s\frac{\alpha}{t}\right)^2 \mathrm{d}s = \frac{\alpha^2}{2t},$$

which equals $I_{\mathcal{N}(0,t)}(\alpha)$ as in (1.1.2). A similar optimization procedure gives us the large deviation rate function for (1.1.3):

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} B_i(t_1) \approx \alpha_1,\ \frac{1}{n}\sum_{i=1}^{n} B_i(t_2) \approx \alpha_2\right]$$
$$\approx \exp\left\{-n\left(\frac{\alpha_1^2}{2t_1} + \frac{(\alpha_2-\alpha_1)^2}{2(t_2-t_1)}\right)\right\}.$$

The two time rate function has an interesting conditional structure. The first term corresponds to the rate for the large deviations at time t_1, whereas the second term corresponds to the rate for large deviations at time t_2, given that we were at α_1 at time t_1. This conditional structure arises from the integral form of I_S. This integral form is in turn a consequence of the Markov property of Brownian motion.

These properties are instances of a general principle, and even hold for sequences of processes with mean-field interaction.

1.1.3 *Mean-field interacting models: the Curie-Weiss model*

The results of the sections above can be taken beyond the case of averages of independent random variables. A notable example with weak interactions is the *Curie-Weiss model* which is a so-called mean-field model for the behaviour of ferromagnets. It gives a microscopic description for the states of a collection of atoms of a ferromagnet, from which we can derive the behaviour of a macroscopic quantity of interest: the magnetization.

We model a magnet by n atoms each having a magnetic spin $\sigma_i \in \{-1,1\}$. We define the empirical magnetization $x_n(\sigma) := \frac{1}{n}\sum_{i=1}^n \sigma_i$ and define a probability distribution $\mu_{n,\beta}$ on the microscopic state space $\{-1,1\}^n$ by

$$\mu_{n,\beta}(\mathrm{d}\sigma) := e^{n2^{-1}\beta x_n(\sigma)^2} Z_{\beta,n}^{-1}\mathbb{P}_n(\mathrm{d}\sigma). \tag{1.1.4}$$

Here $\mathbb{P}_n$ is the product $(\frac{1}{2},\frac{1}{2})$ measure on $\{-1,1\}^n$. $\beta \geq 0$ has the interpretation of the inverse temperature $\beta = T^{-1}$ and $Z_{n,\beta}$ is a normalising constant. Note that for $\beta = 0$, i.e. infinite temperature, we have that $\mu_{n,0} = \mathbb{P}_n$, describing non-interacting spins.

We are interested in the behaviour of this magnetization $x_n(\sigma)$ for large n, as this is the macroscopic quantity that we can observe externally. Suppose that β is small. Then the measures $\mu_{n,\beta}$ are close to the product measures and we expect x_n to converge to 0, just as in the coin-flip example. For large β, however, the spins tend to have the same value, but states with many positive or negative spins are equally likely, so the law of large numbers breaks down. This is reflected in the large deviation principle as

$$\mu_{n,\beta}(x_n(\sigma) \approx \alpha) \approx e^{-nI(\alpha)},$$

where

$$I(\alpha) = \frac{1-\alpha}{2} \log \frac{1-\alpha}{2} + \frac{1+\alpha}{2} \log \frac{1+\alpha}{2} - \frac{1}{2}\beta\alpha^2 - C$$

and where C is such that the minimum of I equals zero, see e.g. Section 3.4 in Rassoul-Agha and Seppäläinen [2015]. For $\beta \leq 1$, the rate function has a unique minimizer at 0, reflecting the law of large numbers behaviour, and for $\beta > 1$ there are two distinct minimizers reflecting the concentration on microscopic configurations with a majority of positive or negative spins.

1.1.4 *Mean-field interacting processes and the McKean-Vlasov equation*

As in going from coin-flips to Schilder's theorem, also here we can add dynamics to the Curie-Weiss model to study the large deviations of the trajectory of the empirical magnetisation. To generalize, we consider n stochastic processes $\{Y_{n,i}(t)\}_{1\leq 1\leq n}$ on some subset of $\mathbb{R}^d$. In the Curie-Weiss model example, these processes represent the evolution of the individual spins. We assume that these n processes interact in such a way that the vector $(Y_{n,1}, \ldots, Y_{n,n})$ is Markovian on $(\mathbb{R}^d)^n$ and the evolution of an individual process depends only on the others via the average $x_n(t) := n^{-1} \sum Y_i(t)$. Then the evolution of $x_n(t)$ itself is also Markovian on some set $E \subseteq \mathbb{R}^d$. Therefore, the microscopic Markovian evolution for $\{Y_{n,i}\}_{1\leq n}$ induces a macroscopic Markovian evolution x_n.

Under suitable conditions, we can show that the trajectory of the mean $\{x_n(t)\}_{t\geq 0}$ converges as $n \to \infty$ to the solution of a differential equation, the so called *McKean-Vlasov equation*. This convergence is a form of the law of large numbers, just as in the case considered above for the averages of Brownian motion that converge to the 0 trajectory, but here the processes interact weakly and can be of completely different nature. The law of large numbers shows that the macroscopic evolution becomes, in the limit, deterministic and as such, simpler than the systems where n is finite.

This law of large numbers is useful to study the evolution of average quantities of very large interacting systems. Large deviation principles around the McKean-Vlasov equation are proven in various contexts. These contexts include Schilders's theorem, Schilder [1966] and the theory of random perturbations of dynamical systems by Freidlin and Wentzell [1998]. A non-exhaustive collection of papers where large deviations for trajectories of spin-flip models are proven is Comets [1987], Léonard [1995] and Dai Pra and den Hollander [1996]. In the measure valued context we have the work by Dawson and Gärtner [1987] and recently there is the work by Feng and Kurtz [2006].

Under appropriate conditions on the processes x_n on $E \subseteq \mathbb{R}^d$, we have that

$$\mathbb{P}\left[\{x_n(t)\}_{t\geq 0} \approx \{\gamma(t)\}_{t\geq 0}\right] \approx e^{-nI(\gamma)}, \tag{1.1.5}$$

for $\gamma : [0,\infty) \to E$. I takes the form

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s),\dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{AC}$ denotes the set of absolutely continuous trajectories. I_0 quantifies the large deviations for $\{x_n(0)\}_{n\geq 0}$ alone, and $\mathcal{L} : E \times \mathbb{R}^d \to [0,\infty)$ is a *Lagrangian*. This Lagrangian is convex in $\dot{\gamma}(s)$ and satisfies $\mathcal{L}(\gamma(s),\dot{\gamma}(s)) = 0$ along the solutions of the McKean-Vlasov equation. To conclude, the large deviation principle quantifies how close the trajectory $\{x_n(t)\}_{t\geq 0}$ is to the law of large numbers limit.

As in the example that considered the averages of independent Brownian motions, the large deviation principle for the trajectories with a rate function in Lagrangian form gives a way to study the rate function of the large deviation principle of $\{x_n(t)\}$ for fixed $t \geq 0$. By the contraction principle, we obtain

$$\mathbb{P}\left[x_n(t) \approx a\right] \approx e^{-nI_t(a)}, \tag{1.1.6}$$

where

$$I_t(a) = \inf_{\substack{\gamma\in\mathcal{AC} \\ \gamma(t)=a}} \left\{ I_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s),\dot{\gamma}(s))\mathrm{d}s \right\}. \tag{1.1.7}$$

In contrast to the case in which we studied the behaviour of averages of Brownian motion, it is in general not possible to obtain an explicit representation for I_t. However, the representation of I_t can be interpreted as an action minimization problem in classical mechanics. Techniques from classical mechanics can thus be used to obtain information on the rate function I_t which would be very difficult to obtain from the law of $x_n(t)$ itself. We find that extremals γ of (1.1.7) solve the second order *Euler-Lagrange equations*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{L}_v(\gamma(t),\dot\gamma(t)) = \mathcal{L}_x(\gamma(t),\dot\gamma(t)), \qquad \mathcal{L}_x(\gamma(0),\dot\gamma(0)) = DI_0(\gamma(0)).$$

Here $\mathcal{L}_x, \mathcal{L}_v$ denote the derivative of $\mathcal{L}$ with respect to the first and second coordinate. DI_0 denotes the gradient of I_0. Following the theory of classical mechanics, we can switch to the easier first order *Hamilton equations* by doubling the dimension of the problem. We define the *Hamiltonian*

$$H(x,p) = \sup_{v\in\mathbb{R}^d} \langle p, v\rangle - \mathcal{L}(x,v) \tag{1.1.8}$$

and the momentum $p(t) = \mathcal{L}_v(\gamma(t),\dot\gamma(t))$. Rewriting the Euler-Lagrange equation, we find that $(x(t),p(t))$ satisfies the *Hamilton equations*:

$$\begin{bmatrix}\dot x\\ \dot p\end{bmatrix} = \begin{bmatrix}H_p(x,p)\\ -H_x(x,p)\end{bmatrix}, \qquad p(0) = DI_0(\gamma(0)). \tag{1.1.9}$$

Similar to the notation for $\mathcal{L}$, H_x and H_p denote the derivatives of H with respect to the first and second coordinate. The Hamilton equations can be seen as an extension of the McKean-Vlasov equation. Suppose $x(t)$ solves the McKean-Vlasov equation, so $\mathcal{L}(x(t),\dot x(t)) = 0$. Then, as $\mathcal{L}$ is non-negative, it follows by the convexity of $\mathcal{L}$ in the second coordinate that $p(t) := \mathcal{L}_v(x(t),\dot x(t)) = 0$. In other words, the McKean-Vlasov equation equals

$$\dot x(t) = H_p(x(t),0). \tag{1.1.10}$$

The evolution of $p(t)$ satisfies $\dot p(t) = -H_x(x(t),p(t)) = -H_x(x(t),0) = 0$ as $H(x,0) = 0$ for all $x \in E$. So by considering the large deviations for the trajectories of $x_n(t)$, we do not only find the McKean-Vlasov equation in a natural way, but obtain a formalism that describes all optimal trajectories in the sense of (1.1.7).

1.2 USING HAMILTONIAN DYNAMICS TO STUDY OPTIMAL TRAJECTORIES

We use the extension of the McKean-Vlasov equation by the Hamilton equations for two applications : for Gibbs-non-Gibbs transitions and for the study of the entropy along the McKean-Vlasov equation.

1.2.1 *Gibbs-non-Gibbs transitions*

We revisit the Curie-Weiss model where we considered the distribution

$$\mu_{n,\beta}(\mathrm{d}\sigma) = e^{n2^{-1}\beta x_n(\sigma)^2} Z_{\beta,n}^{-1} \mathbb{P}_n(\mathrm{d}\sigma),$$

on $\{-1,1\}^n$ and where $x_n(\sigma) = \frac{1}{n}\sum_{i=1}^n \sigma_i$.

A quantity that is of interest in addition to the limiting behaviour of $x_n(\sigma)$ as n goes to infinity, is the limiting distribution of a single spin, given that the average of all other spins converges.

In general, for a sequence of permutation invariant measures $\nu_n \in \mathcal{P}(\{-1,1\}^n)$, we consider

$$\gamma_n^\nu(\mathrm{d}\sigma_1 \,|\, \alpha_n) := \nu_n\left(\mathrm{d}\sigma_1 \,\middle|\, \frac{1}{n-1}\sum_{i=2}^n \sigma_i\right)$$

given any configuration $(\sigma_2, \ldots, \sigma_n)$ such that $\frac{1}{n-1}\sum_{i=2}^n \sigma_i = \alpha_n$.

We say that a magnetisation $\alpha \in [-1,1]$ is *good* for the sequence ν_n if there is some neighbourhood $\mathcal{N}$ of α such that for all $\hat{\alpha} \in \mathcal{N}$ and all sequences $\alpha_n \to \hat{\alpha}$, we have that the weak limit $\lim_n \gamma_n^\nu(\cdot \,|\, \alpha_n)$ exists and is independent of the chosen sequence α_n. If so, we denote this limit by $\gamma(\cdot \,|\, \hat{\alpha})$.

We call a magnetization α *bad*, if it is not good. Finally, we say that the sequence ν_n is *sequentially Gibbs* if all magnetizations are good.

It is straightforward to verify that the sequence $\mu_{n,\beta}$ of the Curie-Weiss model is sequentially Gibbs. However, it has been shown that the Gibbs property can be lost under the evolution of a Markov process, see Külske and Le Ny [2007], Ermolaev and Külske [2010], Fernández et al. [2013]. If the sequence of Markov processes satisfies a large deviation principle for the trajectories, it was shown in Ermolaev and Külske [2010], den Hollander et al. [2015] that a bad magnetization α corresponds to the non-uniqueness of optimal trajectories for

$$I_t(\alpha) = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(t)=\alpha}} \left\{ I_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s \right\},$$

which in turn is equivalent to non-differentiability of I_t at α. Using the first order Hamilton equations, it becomes possible to obtain concrete information on the existence of multiple optimal solutions, and as a consequence information on the occurrence of bad magnetizations.

1.2.2 *Exponential decay of entropy along the McKean-Vlasov equation*

As a second application, we consider the decay of entropy along solutions of the Hamilton equations. In the general context of a Markov processes $X(t)$ with some stationary measure μ, it is well known that the relative entropy $\nu \mapsto S(\nu \,|\, \mu) = \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathrm{d}\mu$ is decreasing along the distribution of the Markov process. To be precise, if $\rho(t)$ is the law of $X(t)$, then $t \mapsto S(\rho(t) \mid \mu)$ is decreasing.

Now suppose this Markov process X has a generator A, then, at least formally, $\{\rho(t)\}_{t \geq 0}$ solves the *Kolmogorov forward equation* $\dot{\rho}(t) = A^* \rho(t)$. In this setting, we say that $S(\cdot \mid \mu)$ is a *Lyapunov function* for the Kolmogorov forward equation. To connect this framework to the McKean-Vlasov equation and large deviations, we consider large deviations of the measure valued trajectories of the average of n independent copies $X^1, X^2, \ldots, X^n$ of X:

$$\rho_n(t) := \frac{1}{n} \sum_{i \leq n} \delta_{X^i(t)}.$$

As n goes to infinity, the trajectories $\{\rho_n(t)\}_{t \geq 0}$ converge almost surely to the solution of the Kolmogorov forward equation, which thus coincides with the McKean-Vlasov equation in this setting. This means, that at least intuitively, we are back in the setting of the previous sections. Also the relative entropy can be interpreted in this framework, namely, the relative entropy is the large deviation rate function of $\{\rho_n(0)\}_{n \geq 0}$, if $X(0)$ is distributed according to the stationary measure μ.

This basic principle can be explored further for systems that have mean-field interaction. We return to the setting where $\{x_n(t)\}_{t \geq 0}$ are Markov processes on some subset $E \subseteq \mathbb{R}^d$ that satisfy a large deviation principle for the trajectories. Suppose that I_0 is the rate function of $x_n(0)$ in the case that $x_n(0)$ is distributed according to the stationary distribution of the process with n particles. Then, it follows that the rate function I_t at time t equals I_0 and, in particular, we find that $I_0(x(t)) \leq I_0(x(0))$ for any solution of the McKean-Vlasov equation.

In the non-interacting case, where $X(t)$ is either a diffusion process or a jump process, it is well known that the (modified) logarithmic Sobolev inequality implies that the relative entropy decays exponentially along the solutions of the Kolmogorov forward equation, see for example Bobkov and Tetali [2006] and Bakry et al. [2014]. Studying the Hamiltonian function H in the mean-field setting reveals a similar structure for the decay of the rate function I_0 of the stationary measures along the solution of the McKean-Vlasov equation.

1.3 INTERACTING LATTICE SPIN SYSTEMS

More sophisticated models in the study of interacting spin systems are lattice systems, where the interactions are not mean-field, but, for example, nearest neighbour. We consider the lattice $\mathbb{Z}^d$ and on each site $i \in \mathbb{Z}^d$ there is a spin $\sigma_i \in \{-1, 1\}$. Also in this case we are interested in the average magnetic spin, but because of spatial nature of our system, our limiting procedure is more involved in comparison to the mean-field Curie-Weiss model.

We define a shift operator $\theta_i : \{-1,1\}^{\mathbb{Z}^d} \to \{-1,1\}^{\mathbb{Z}^d}$ by $(\theta_i \sigma)_j = \sigma_{i+j}$ and define volumes $\Lambda_n = [-n, n]^d \cap \mathbb{Z}^d$. Finally, we define the empirical measure

$$L_n(\sigma) := \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} \delta_{\theta_i \sigma} \qquad \in \mathcal{P}(\{-1,1\}^{\mathbb{Z}^d}). \tag{1.3.1}$$

If σ has a translation invariant (ergodic) distribution μ, it follows by the ergodic theorem that $L_n(\sigma) \to \mu$ almost surely with respect to μ.

As above, we can ask for large deviations around this limiting theorem. If μ is a product measure, we find

$$\mu\left(L_n(\sigma) \approx \nu\right) \approx e^{-|\Lambda_n| s(\nu \mid \mu)},$$

where s is the relative entropy density

$$s(\nu \mid \mu) = \lim_{n \to \infty} \frac{1}{|\Lambda_n|} S(\nu_n \mid \mu_n)$$

and where μ_n, ν_n are the restrictions to $\{-1,1\}^{\Lambda_n}$ and S is the relative entropy.

This large deviation principle also holds if we replace μ by a Gibbs measure, see Georgii [2011]. As in the mean-field setting, it has been shown that the Gibbs property of a measure can be lost under the evolution of Markovian dynamics. Additionally, it is expected that the emergence of bad configurations in this context corresponds to non-uniqueness of optimal trajectories of the path-space large deviation principle, see van Enter et al. [2010].

Even though we will not touch upon this particular conjecture, we provide a first step by proving the path-space large deviation principle for the trajectories of empirical measures.

1.4 FUNCTIONAL ANALYTIC THEORY IN RELATION TO PROBABILITY AND MEASURE THEORY

1.4.1 *Semigroup theory in the study of Markov processes*

At the core of proving weak convergence or large deviation results for a sequence of (Feller) Markov processes X_n on a Polish space E following the methods in Ethier and Kurtz [1986] and Feng and Kurtz [2006] lies the use of functional analytic semigroup theory. This is based on a scheme of reduction steps that reduces the convergence, or large deviation question, on the Skorokhod space to that of the finite dimensional distributions. Because the processes are Markovian, the study of the finite dimensional distributions reduces in turn to the study of the processes at two times. For the weak convergence question, it suffices to study the sequence of transition operators $\{S_n(t)\}_{t\geq 0}$, where $S_n(t) : C_b(E) \to C_b(E)$ is defined by

$$S_n(t)f(x) = \mathbb{E}\left[f(X_n(t)) \,|\, X_n(0) = x\right].$$

By the tower property for conditional expectations, one sees that $S(t)S(r) = S(t+r)$, i.e. S is a semigroup. For the large deviation question, it is not the conditional expectation that is of importance, but the family of conditional log-Laplace transforms

$$V_n(t)f(x) = \frac{1}{n}\log S_n(t)e^{nf}(x),$$

which also form a semigroup. Both these semigroups are defined on a possibly infinite dimensional function space. The behaviour of sequences of such semigroups, however, is easily introduced by considered semigroups on $\mathbb{R}$.

Consider a continuous semigroup $\{z(t)\}_{t\geq 0}$ on $\mathbb{R}$, i.e. $z(t) \in \mathbb{R}$, $z(t)z(s) = z(t+s)$ and $z(0) = 1$ and $t \mapsto z(t)$ is continuous. It follows that $z(t)$ must be of the form $z(t) = e^{at}$ for some $a \in \mathbb{R}$. Note that $a = \frac{\mathrm{d}}{\mathrm{d}t}z(t)|_{t=0}$.

Now suppose that we have a collection of semigroups $z_{a(n)}$ of the form $z_{a(n)}(t) = e^{ta(n)}$. If we have $a(n) \to a$, then for any $T > 0$

$$\lim_{n\to\infty}\sup_{t\leq T}|z_{a(n)}(t) - z_a(t)| = 0.$$

In the infinite dimensional setting, we study the convergence of semigroups by the same principle. We will focus below only on the linear semigroups $\{S_n(t)\}_{t\geq 0}$, as this theory is more developed than that of the non-linear semigroups $\{V_n(t)\}_{t\geq 0}$. We define the generators A_n of $\{S_n(t)\}_{t\geq 0}$ by

$$A_n f := \frac{\mathrm{d}}{\mathrm{d}t} S_n(t) f|_{t=0} = \lim_{t\downarrow 0} \frac{S_n(t)f - f}{t}.$$

Note that $A_n f$ is not defined for all functions, but only for a subset of $C_b(E_n)$ that depends on the topology in which we take the limit. We expect these generators to play a crucial role in the determination of the limiting behaviour of the semigroups $S_n(t)$. In particular, in analogy to the one-dimensional example above, we expect that if an operator A that is the generator of a semigroup $\{S(t)\}_{t\geq 0}$ exists, the convergence $A_n f \to Af$ for sufficiently many f implies that $S_n(t)f \to S(t)f$ uniformly for t in compact intervals.

In the discussion above, the topologies on $C_b(E)$ that are considered are intentionally left undefined. The approach described above works very well in the setting that E is a compact space and the topology on $C_b(E)$ is the supremum norm topology. In this setting, the semigroups $\{S_n(t)\}_{t\geq 0}$ are strongly continuous for the norm, i.e. we have that for every $t \geq 0$ the maps $S_n(t) : (C_b(E), \|\cdot\|) \to (C_b(E), \|\cdot\|)$ are continuous, and additionally, we have that $t \mapsto S_n(t)f$ is norm continuous for all f and n. Thus, we can use the theory of strongly continuous semigroups on Banach spaces, and the semigroup convergence result for linear semigroups is known as the Trotter-Kato theorem, see Engel and Nagel [2000] or Ethier and Kurtz [1986].

The work by Feng and Kurtz [2006] shows that this approach can also be applied to the non-linear semigroups $V_n(t)$ and this approach naturally leads us to the Hamiltonian H that has featured the discussion in the earlier sections of the introduction. Calculating the generator H_n of the semigroup $\{V_n(t)\}_{t\geq 0}$, we formally find by the chain rule that

$$H_n f := \frac{\mathrm{d}}{\mathrm{d}t} V_n(t) f|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{n} \log S_n(t) e^f|_{t=0} = \frac{1}{n} e^{-nf} A_n e^{nf}.$$

Thus, if an operator H, such that $H_n f \to Hf$ for sufficiently many f, exists and if H generates a semigroup $\{V(t)\}_{t\geq 0}$, then by the Crandall-Liggett theorem we find $V_n(t)f \to V(t)f$.

Various techniques to show that H determines a limiting semigroup $\{V(t)\}_{t\geq 0}$ have been introduced in Feng and Kurtz [2006] and we will use a number of these techniques in Chapter 3 when we consider the large deviation behaviour of mean-field interacting spin systems.

In mean-field examples with state-space $E \subseteq \mathbb{R}^d$, the operator H is often of the form $Hf(x) = H(x, \nabla f(x))$ for some Hamiltonian function $H : E \times \mathbb{R}^d \to \mathbb{R}$. It is exactly this function that appeared before in equation (1.1.8). In fact, using this approach one finds the function H first from the limiting procedure $H_n f \to Hf$, after which $\mathcal{L}$ is defined as the Legendre transform of H.

Using $\mathcal{L}$ the semigroup $\{V(t)\}_{t\geq 0}$ can be rewritten using variational methods as

$$V(t)f(x) = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(0)=x}} f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s,$$

which can be used to prove that the large deviation principle holds for the trajectories with a rate function in Lagrangian form.

1.4.2 *Semigroups for processes on a Polish space*

For Feller processes on Polish non-compact E, the semigroups corresponding to Markov processes are usually not strongly continuous for the norm, an issue that already appears for the semigroup of conditional expectations of a process like standard Brownian motion on $\mathbb{R}$.

For processes on $\mathbb{R}$, or locally compact spaces in general, we can salvage the Banach space approach by considering the space $(C_0(E), \|\cdot\|)$, the space of functions that vanish at infinity. For non-locally compact E, however, it is not possible to recover the Banach space approach. Various other approaches to prove results like the Trotter-Kato theorem have been introduced.

For example, results have been obtained by considering a notion of convergence for sequences called *buc(bounded and uniformly on compacts)* convergence, i.e. $f_n \to f$ for (buc) if $\sup_n \|f_n\| < \infty$ and $\sup_{x\in K} |f_n(x) - f(x)| \to 0$ for all compact sets $K \subseteq E$. Stated in this form, (buc) convergence is not a topological notion, so many of the functional analytic techniques are not available.

An alternative modern approach to studying weak convergence of Markov processes on Polish spaces is via the martingale problem, see for example Ethier and Kurtz [1986] or Stroock and Varadhan [1979]. This approach salvages the idea of a generator by noting that for f in the domain of A_n the process

$$f(X_n(t)) - f(X_n(0)) - \int_0^t A_n f(X_n(s))\mathrm{d}s$$

is a martingale, which is essentially a probabilistic way of saying that $\frac{\mathrm{d}}{\mathrm{d}t} S_n(t)f = A_n S_n(t)f$.

Even though the idea of the martingale problem has been very effective, the connection to functional analysis that has been useful in the compact setting, has been lost.

1.4.3 *A suitable locally convex topology for the space of bounded continuous functions*

The basic underlying reason for this disconnect is found by considering the continuous dual space of $(C_b(E), \|\cdot\|)$. The continuous dual space is the space of all continuous linear maps of $C_b(E)$ to $\mathbb{R}$ and is usually denoted by $(C_b(E), \|\cdot\|)'$. If X is compact, the Riesz representation theorem tells us that the dual space equals the space of regular Borel measures. This is also the case if E is locally compact and we consider $(C_0(E), \|\cdot\|)'$. For non-compact spaces E, however $(C_b(E), \|\cdot\|)'$ is strictly larger than the space of regular Borel measures.

It is exactly the identification of the continuous dual space with the space of regular Borel measures that makes functional analysis so effective to study probability measures, and which in turn is the reason why this strong connection fails if we consider $(C_b(E), \|\cdot\|)$ if E is non-compact.

The leading principle, thus, should be to find a locally convex topology on $C_b(E)$ so that the dual coincides with the space of regular Borel measures. A topology that has this property is the strict topology β, see Sentilles [1972]. β has more desirable properties as it is separable, satisfies the Stone-Weierstrass theorem and the Arzela-Ascoli theorem. In this thesis, we will show that we also have the closed graph, inverse-, and open mapping theorems between two spaces of this type.

Additionally, a part of this thesis is devoted to studying $(C_b(E), \beta)$ and semigroup theory on locally convex spaces like $(C_b(E), \beta)$. As a result, we find that the solution to a well posed martingale problem always gives a strongly continuous semigroup for the strict topology, reconnecting the probabilistic theory to the functional analytic one.

1.5 OUTLINE OF THE THESIS

The thesis is divided into three parts:

(I) An introductory part, including this introduction and Chapter 2 introducing the important mathematical concepts.

(II) Large deviations of Markov processes and the applications thereof, including Chapters 3 to 7.

(III) Functional analytic methods related to the study of Markov processes on non-compact Polish spaces, including Chapters 8 to 10.

As mentioned above, in *Chapter 2*, we start with a mathematical introduction of the various probabilistic and functional analytic concepts

We proceed with *Chapter 3*, where we prove the path-space large deviation principle for mean-field dynamics in a finite dimensional setting. The proof relies on the verification of the uniqueness of viscosity solutions to a class of Hamilton-Jacobi equations.

We proceed with two chapters on the applications of the mean-field results. In *Chapter 4*, we study the behaviour of the entropy under the evolution of the McKean-Vlasov equation. We give a sufficient condition for exponential decay of this entropy. Additionally, we give conditions for the convexity of the entropy along entropic geodesics. In *Chapter 5*, we use ideas from Hamiltonian mechanics and optimal control theory to study the optimal trajectories for (1.1.7). We obtain rigorous and context-independent methods to decide whether optimal trajectories arriving at a fixed point are unique, information that is of importance in the study of mean-field Gibbs-non-Gibbs transitions.

We proceed with two chapters on the path-space large deviations of measure valued trajectories of Markov processes. *Chapter 6* studies the large deviation behaviour of the trajectories of the empirical density of n independent copies of a Feller process. In this setting, it is generally unclear how to take the derivative with respect to time of the law of the process. In analogy to the setting of diffusion processes on a manifold, we introduce a method to find a suitable class of test functions, so that the dual space can be used as a space of ‘speeds’. In *Chapter 7*, we study the large deviations of trajectories of the empirical measure, i.e. (1.3.1) taking averages over shifts, of lattice interacting systems. We prove the large deviation principle, but without a Lagrangian representation of the rate function. We do however conjecture, that the methods developed in Chapter 6 give the correct form.

In the final three chapters, we turn to the functional analytic aspects of semigroup theory and the strict topology. In *Chapter 8*, we consider strongly continuous semigroups on locally convex spaces that include $(C_b(E), \beta)$ where E is Polish. We prove a Hille-Yosida theorem and generalize various classical results from the Banach space setting to the class of locally convex spaces under consideration. In *Chapter 9*, we reconnect the martingale problem approach with the functional analytic approach to semigroups. In the final *Chapter 10*, we prove that $(C_b(E), \beta)$ satisfies the conclusions of the Banach-Dieudonné theorem. As a consequence, we obtain the closed graph, inverse-, and open mapping theorems between $(C_b(E), \beta)$ and $(C_b(F), \beta)$ for separable metric spaces.

2

MATHEMATICAL INTRODUCTION

Before introducing the definitions of the topics that will be discussed in this thesis, we first introduce some basic notation. We denote $\mathbb{R}^+ = [0, \infty)$. (E, d) will denote a complete separable metric space. Often, we will consider Polish spaces E, spaces such that there exists a metric d so that (E, d) is a complete separable metric space. On E we consider the following objects:

- The Borel σ-algebra $\mathcal{B}(E)$.
- The space of Radon measures $\mathcal{M}(E)$.
- The space of probability measures $\mathcal{P}(E)$.
- The space of measurable functions $M(E)$.
- The space of bounded measurable functions $M_b(E)$.
- The space of continuous and bounded functions $C_b(E)$.
- If E is locally compact, the space of continuous functions that vanish at infinity $C_0(E)$.
- The Skorokhod space $D_E(\mathbb{R}^+)$ of trajectories $\gamma : \mathbb{R}^+ \to E$, that are right continuous and have left limits.

For any set $A \subseteq E$, we denote by $\overline{A}$, A° the closure and the interior of A. We denote by A^c the complement of A in E.

We say that $(\Omega, \mathcal{F})$ is *measurable space* if Ω is some arbitrary set, and $\mathcal{F}$ is a σ-algebra on Ω. We say that $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space* if $(\Omega, \mathcal{F})$ is a measurable space and if $\mathbb{P}$ is a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$.

For a measure $\mathbb{P}$ on a measurable space $(\Omega_1, \mathcal{F}_1)$ and a measurable map $\pi : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$, we write $\pi_\#\mathbb{P}$ for push-forward measure of $\mathbb{P}$ on $\mathcal{F}_2$:

$$\pi_\#\mathbb{P}(A) = \mathbb{P}(\pi^{-1}(A)) \qquad \forall A \in \mathcal{F}_2.$$

For any complete separable metric space (E, d), we say that $X : \Omega \to E$ is an E valued *random variable* if X is measurable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{B}(E))$.

If we talk about a collection of E valued random variables $\{X_\alpha\}$, we will implicitly assume the existence of a common probability space $(\Omega, \mathcal{F})$ on which they are defined.

For a collection of random variables $\{X_i\}_{i \in I}$ taking values in E, we write $\sigma\{X_i \,|\, i \in I\}$ for the σ-algebra generated by the random variables $\{X_i\}_{i \in I}$.

In the next few sections, we define and motivate definitions in a number of areas. Motived by the transition semigroup of a Markov process, we start by introducing the general theory of strongly continuous semigroups and their generators in Section 2.1. We consider both linear and non-linear semigroups as we will encounter non-linear semigroups in the study of the large deviation behaviour of Markov processes. The main goal is to understand the conditions under which an operator generates a semigroup. As a tool for this question in the context of the space of continuous functions, we introduce the theory of viscosity solutions in Section 2.2. We proceed with some basic definitions for the study of time-homogeneous Markov processes in Section 2.3, where we will see that linear semigroups play a prominent role. Large deviation theory follows thereafter in Section 2.4 and we show that strongly continuous non-linear semigroups naturally appear in the study of large deviations for Markov processes. We conclude in Section 2.5 with an introduction to locally convex spaces. In particular, we will use this theory to introduce a locally convex space which is suited for the study of Markov transition semigroups for a Markov process defined on a non-compact Polish space.

2.1 STRONGLY CONTINUOUS SEMIGROUPS

Let $(X, \|\cdot\|)$ be a Banach space. Consider a family of continuous operators $\{T(t)\}_{t \geq 0}$ mapping X into X. To avoid confusion, note that we have not assumed the operators to be linear.

Definition 2.1.1 (Strongly continuous semigroup)**.** We say that $\{T(t)\}_{t \geq 0}$ is a *semigroup* if $T(0) = \mathbb{1}$ and $T(t)T(s) = T(t+s)$ for $s, t \geq 0$. We say that $\{T(t)\}_{t \geq 0}$ is *strongly continuous semigroup* if $t \mapsto T(t)x$ is continuous for every $x \in X$. Finally, we say that the semigroup is *contractive* if for all $t \geq 0$, we have $\|T(t)\| \leq 1$.

Before introducing the generator of T, we set some notation for non-continuous operators on X. A non-continuous operator $A = (A, \mathcal{D}(A))$ is given by a domain $\mathcal{D}(A) \subseteq X$ and a map $A : \mathcal{D}(A) \to X$. Also, we will write A for the graph of the map: $A = \{(x, Ax) \,|\, x \in \mathcal{D}(A)\}$. Finally, in some cases we even allow for multi-valued operators.

We say that $(A, \mathcal{D}(A))$ is *closed* if $\{(x, Ax) \,|\, x \in \mathcal{D}(A)\}$ is closed in the product space $X \times X$ with the product topology. We say that $\mathcal{D}$ is a *core* for $(A, \mathcal{D}(A))$, if the closure of $\{(x, Ax) \,|\, x \in \mathcal{D}\}$ in the product space contains $\{(x, Ax) \,|\, x \in \mathcal{D}(A)\}$.

To avoid confusion, we will denote linear semigroups by either $S(t)$ or $T(t)$ and their linear generators by $(A, \mathcal{D}(A))$. Non-linear semigroups will be denoted by $V(t)$ and their generators by $(H, \mathcal{D}(H))$.

2.1.1 *Generators of linear semigroups*

Now consider a strongly continuous semigroup of linear operators $\{T(t)\}_{t\geq 0}$.

Definition 2.1.2 (The generator of a linear semigroup)**.** Let $\{T(t)\}_{t\geq 0}$ be a strongly continuous semigroup of linear operators on X. Denote by

$$\mathcal{D}(A) := \left\{ x \in X \,\middle|\, \lim_{t\downarrow 0} \frac{T(t)x - x}{t} \text{ exists} \right\}.$$

The *generator* $(A, \mathcal{D}(A))$ of $\{T(t)\}_{t\geq 0}$ is a map $A : \mathcal{D}(A) \subseteq X \to X$ which maps $x \in \mathcal{D}(A)$ to $Ax = \lim_{t\downarrow 0} t^{-1}(T(t)x - x)$.

The generator $(A, \mathcal{D}(A))$ of a strongly continuous linear semigroup on a Banach space satisfies the following well known properties, see for example [Engel and Nagel, 2000, Lemma II.1.3].

Lemma 2.1.3. *The generator $(A, \mathcal{D}(A))$ of a strongly continuous semigroup $\{T(t)\}_{t\geq 0}$ of linear operators satisfies*

(a) $\mathcal{D}(A)$ is closed and dense in X.

(b) For $x \in \mathcal{D}(A)$, we have $T(t)x \in \mathcal{D}(A)$ for every $t \geq 0$ and $\frac{\mathrm{d}}{\mathrm{d}t}T(t)x = T(t)Ax = AT(t)x$.

(c) For $x \in X$ and $t \geq 0$, we have $\int_0^t T(s)x\mathrm{d}s \in \mathcal{D}(A)$.

(d) For $t \geq 0$, we have

$$\begin{aligned} T(t)x - x &= A\int_0^t T(s)x\mathrm{d}s && \text{if } x \in X \\ &= \int_0^t T(s)Ax\mathrm{d}s && \text{if } x \in \mathcal{D}(A). \end{aligned}$$

This leads us to the following question. Given a linear operator $(A, \mathcal{D}(A))$, is there a strongly continuous semigroup such that A is its generator? For Markov processes, the question extends to, given an operator$(A, \mathcal{D}(A))$, is it possible to construct a Markov process such that its transition semigroup has A as its generator. The functional analytic question is answered in general by the Hille-Yosida theorem. The result is stated in terms of the resolvent of A.

Definition 2.1.4 (The resolvent)**.** For a linear operator $(A, \mathcal{D}(A))$ on a Banach space X, denote by $\sigma(A) := \{\alpha \in \mathbb{C} \,|\, \alpha - A \text{ is bijective}\}$ the *spectrum* of A. We denote by $\rho(A) = \mathbb{C} \setminus \sigma(A)$ the *resolvent set* of A and by $\mathcal{R}(\alpha, A) = (\alpha - A)^{-1}$ the (functional analytic) *resolvent* of A.

Note that we write functional analytic resolvent. This is also the resolvent that we will use in Chapter 8. In the other sections, we will use the probabilistic resolvent, that has a slightly changed definition. It will be defined below.

Lemma 2.1.5. *Suppose that $\{T(t)\}_{t\geq 0}$ is a linear strongly continuous semigroup on the Banach space X that satisfies*

$$\|T(t)\| \leq Me^{\omega t}$$

for some $M \geq 1$ and $\omega \in \mathbb{R}$. Denote by $(A, \mathcal{D}(A))$ its generator. Then we have

(a) $\{\alpha \in \mathbb{C} \,|\, \operatorname{Re} \alpha > \omega\} \subseteq \rho(A)$*,*

(b) for $\alpha > \omega$, we have the following integral representation

$$\mathcal{R}(\alpha, A)x = \int_0^\infty e^{-\alpha t} T(t)x \mathrm{d}t,$$

(c) For $\alpha > \omega$ and $n \geq 1$

$$\|\mathcal{R}(\alpha, A)^n\| \leq \frac{M}{(\alpha - \omega)^n}.$$

These properties of the operator $(A, \mathcal{D}(A))$ are in fact sufficient for the generation of a linear strongly continuous contraction semigroup.

Theorem 2.1.6 (Hille-Yosida)**.** *For a linear operator $(A, \mathcal{D}(A))$ on a Banach space X, the following are equivalent.*

(a) $(A, \mathcal{D}(A))$ generates a strongly continuous semigroup $\{T(t)\}_{t\geq 0}$ of linear operators that satisfy $\|T(t)\| \leq Me^{\omega t}$ for some $M \geq 1$ and $\omega \in \mathbb{R}$.

(b) $(A, \mathcal{D}(A))$ is closed, densely defined, and for all $\alpha > \omega$, we have $\alpha \in \rho(A)$. Additionally, there exist $M \geq 1$ and $\omega \in \mathbb{R}$ such that for all $n \geq 1$, we have

$$\|R(\alpha, A)^n\| \leq \frac{M}{(\alpha - \omega)^n}.$$

In case these conditions are satisfied, we have the Yosida-approximation formula

$$T(t)x = \lim_{n\to\infty} \left(\frac{n}{t}\mathcal{R}\left(\frac{n}{t}, A\right)\right)^n x \tag{2.1.1}$$

uniformly for t in compact intervals.

We will revisit this theorem in Chapter 8 for a special class of locally convex spaces.

The Yosida-Approximation formula can be understood as follows. A continuous function $b : \mathbb{R}^+ \to \mathbb{R}$ that satisfies $b(0) = 1$ and $b(t)b(s) = b(t+s)$ is necessarily of the form $b(t) = e^{at}$ for some $a \in \mathbb{R}$. In other words, a is the generator of the semigroup $\{b(t)\}_{t\geq 0}$. Using a, we have multiple ways of constructing $b(t)$. One of these methods is

$$b(t) = e^{at} = \lim_{n\to\infty} \left(1 - \frac{t}{n}a\right)^{-n} = \lim_{n\to\infty} \left(\frac{n}{t}\left(\frac{n}{t} - a\right)^{-1}\right)^n.$$

This formula corresponds to (2.1.1), where a is replaced by A. Different formulas for approximating the exponential function yield different approximation schemes for semigroups. This particular scheme is useful as it involves iterates of the continuous resolvent. If one uses, for example, the approximation $e^{at} = \sum_k (at)^k k!^{-1}$ one needs the powers of the possibly non-continuous operators A^k instead.

Note that the conditions simplify if we are interested only in contraction semigroups. A contraction semigroup $\{T(t)\}_{t\geq 0}$, satisfies $\|T(t)\| \leq 1$. Hence, the conditions for generating a contraction semigroup in Theorem 2.1.6 simplify to: $(A, \mathcal{D}(A))$ *is closed, densely defined, and for every* $\alpha > 0$ *we have* $\alpha \in \rho(A)$ *and* $\|\alpha\mathcal{R}(\alpha, A)\| \leq 1$.

Because we will mainly consider contraction semigroups in this thesis, as these are the ones that turn up as transition semigroups of Markov operators, we focus our attention on the Hille-Yosida theorem for the contraction case. First we define the probabilistic resolvent.

Definition 2.1.7. Let $\{T(t)\}_{t\geq 0}$ be a strongly continuous linear contraction semigroup $\{T(t)\}_{t\geq 0}$ with generator $(A, \mathcal{D}(A))$ on a Banach space X. For $\lambda > 0$, define the (probabilistic) *resolvent* $R(\lambda, A)$ by $R(\lambda, A) = (\mathbb{1} - \lambda A)^{-1}$, which is also given by

$$R(\lambda, A)x := \int_0^\infty \frac{1}{\lambda} e^{-\lambda^{-1}t} T(t)x \mathrm{d}t.$$

Note that $R(\lambda, A) = \lambda^{-1}\mathcal{R}(\lambda^{-1}, A)$. Also, note that $R(\lambda, A)x$ is given by the semigroup $T(t)$ evaluated at an exponential random time with expectation λ, which in this case explains the necessity of the condition $\|R(\lambda, A)\| \leq 1$ in the Hille-Yosida theorem. In fact, the approximation formula now reads

$$T(t)x = \lim_n R\left(\frac{t}{n}, A\right)^n x$$

which in a sense is merely a law of large numbers in disguise as n exponential random variables with mean t/n converge almost surely to t as n goes to infinity . This insight, combined with appropriate concentration inequalities is the basis for the extension of the Hille-Yosida theorem to a special class of locally convex spaces in Chapter 8.

2.1.2 *Generation of non-linear contractive semigroups*

Similar generation questions can be asked for non-linear contraction semigroups and their generators. Given some non-linear operator $(A, \mathcal{D}(A))$ can we construct a semigroup $\{T(t)\}_{t\geq 0}$ such that

$$\lim_{t\downarrow 0} \frac{T(t)x - x}{t} = Ax?$$

For non-linear operators this question turns out not to be the optimal one, and instead we turn to our attention towards the Yosida-Approximation characterisation of the generator in (2.1.1).

To verify conditions like in the Hille-Yosida theorem for non-linear operators, we need to verify two main conditions, for all $\lambda > 0$, the resolvent $R(\lambda, A) : X \to X$ exists, and additionally, $\|R(\lambda, A)x - R(\lambda, A)y\| \leq \|x - y\|$. We introduce two definitions that cover these two issues.

Definition 2.1.8 (Dissipative operator)**.** We say that an operator $(A, \mathcal{D}(A))$ is *dissipative* if for all $\lambda > 0$, we have

$$\|(x - \lambda Ax) - (y - \lambda Ay)\| \geq \|x - y\|$$

for all $x, y \in \mathcal{D}(A)$.

Definition 2.1.9 (Range condition)**.** We say that an operator $(A, \mathcal{D}(A))$ satisfies the *range condition* if for all $\lambda > 0$ the range of $(\mathbb{1} - \lambda A)$ is dense in X.

It can be shown that the closure $(\overline{A}, \mathcal{D}(\overline{A}))$ of a dissipative operator $(A, \mathcal{D}(A))$ is itself dissipative and satisfies $\operatorname{rg} \mathbb{1} - \lambda\overline{A} = \overline{\operatorname{rg} \mathbb{1} - \lambda A}$. Hence, if a non-closed operator $(A, \mathcal{D}(A))$ is dissipative and satisfies the range condition, its closure $\overline{A}$ has the property that $\operatorname{rg} \mathbb{1} - \lambda\overline{A} = X$ for all $\lambda > 0$. On the other hand, the map $\mathbb{1} - \lambda\overline{A}$ is injective by the dissipativity of $\overline{A}$. Hence, we can invert the maps and define the contraction mappings $R(\lambda, \overline{A}) : X \to \mathcal{D}(\overline{A})$.

In the linear case, we obtain the Lumer-Phillips result as a consequence of the Hille-Yosida theorem. The result below also holds for non-linear operators and is called the Crandall-Liggett theorem.

Theorem 2.1.10 (Lumer-Phillips, Crandall-Liggett)**.** *For a densely defined, dissipative operator $(A, \mathcal{D}(A))$ on a Banach space X, the following are equivalent.*

(a) The closure $\overline{A}$ of A generates a contraction semigroup in the sense that

$$T(t)x = \lim_n R\left(\frac{t}{n}, \overline{A}\right)^n x$$

uniformly for t in compact intervals.

(b) The range condition holds: $\operatorname{rg}(\mathbb{1} - \lambda A)$ *is dense in X for some(hence all) $\lambda > 0$.*

Note that there exists an extension of the Crandall-Liggett theorem to the case where we consider the space $X = C_b(E)$ equipped with a notion of convergence that is weaker than the norm topology, see Feng and Kurtz [2006]. The verification of the dissipativity of an operator is often not very hard. For operators on function spaces, this can often be checked via the positive maximum principle.

Definition 2.1.11 (The positive maximum principle)**.** Let E be a Polish space. Let $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ be some operator. We say that A satisfies the *positive maximum principle* if for any two functions $f, g \in \mathcal{D}(A)$, we have the following:

(a) If x_0 is such that $f(x_0) - g(x_0) = \sup_{x \in E} \{f(x) - g(x)\}$, then $Af(x_0) - Ag(x_0) \leq 0$.

(b) If x_0 is such that $f(x_0) - g(x_0) = \inf_{x \in E} \{f(x) - g(x)\}$, then $Af(x_0) - Ag(x_0) \geq 0$.

Lemma 2.1.12. *If an operator $(A, \mathcal{D}(A))$ satisfies the positive maximum principle, then it is dissipative.*

On the other hand, checking the range condition for a non-linear operator might prove to be very hard. For function spaces, however, the theory of viscosity solutions offers a way out, see Section 2.2.

2.1.3 *Approximation of semigroups*

A second natural question to be answered for abstract semigroups is the one of approximation. Given a sequence of strongly continuous semigroups $\{T_n(t)\}_{t\geq 0}$, the goal is to find conditions which imply that the sequence converges strongly and uniformly on compact intervals to some limiting semigroup $\{T(t)\}_{t\geq 0}$. For linear semigroups, this is the content of the Trotter-Kato approximation theorem, which we will again extend to a class of locally convex spaces in Chapter 8. We state only a special case of the approximation theorem.

Theorem 2.1.13 (Trotter-Kato). *Let $\{T_n(t)\}_{n\geq 1, t\geq 0}$ be a family of strongly continuous linear contraction semigroups on a Banach space X. Then (a) implies (b).*

(a) There exists a densely defined linear operator $(A, \mathcal{D}(A))$ such that $A_n x \to Ax$ for all x in a core for $(A, \mathcal{D}(A))$ and such that the range condition holds for $(A, \mathcal{D}(A))$.

(b) The closure of $(A, \mathcal{D}(A))$ generates a strongly continuous linear semigroup $\{T(t)\}_{t\geq 0}$ and we have

$$\lim_{n\to\infty} \sup_{t\leq T} \|T_n(t)x - T(t)x\| = 0$$

for all $T \geq 0$ and $x \in X$.

As for the Hille-Yosida theorem, this result can be extended to non-linear semigroups. More importantly, the result can be extended to convergence of semigroups on different spaces.

This importance of this result in the field of probability is easily seen by the functional central limit theorem, or Donsker's theorem, which states that a suitably rescaled continuous time random walk converges to Brownian motion. Considering the semigroup analogue of this statement, this means that the transition semigroup $\{S_n(t)\}_{t\geq 0}$ of rescaled random walk on $C_0(\frac{1}{n}\mathbb{Z})$ converges to the transition semigroup of Brownian motion, that acts on $C_0(\mathbb{R})$. See Trotter [1958] where Trotter motivates a result similar to Theorem 2.1.14.

We formalise this intuitive picture for general Banach spaces. Let $\{X_n\}_{n\geq 1}$ be a sequence of Banach spaces. For every n let $\eta_n : X \to X_n$ be continuous and linear map.

Consider a sequence of operators $B_n \subseteq X_n \times X_n$. We define the *extended limit* $ex - \lim B_n$ of the sequence of operators by

$$\{(x,y) \in X \times X \mid \exists (x_n, y_n) \in B_n : \|\eta_n x - x_n\| + \|\eta_n y - y_n\| \to 0\} .$$

The following theorem can be proven as in Proposition 5.5 in Feng and Kurtz [2006] using Theorem 3.2 of Kurtz [1974].

Theorem 2.1.14. *For every $n \geq 1$ let $\{T_n(t)\}_{t\geq 0}$ be a strongly continuous semigroup on a Banach space $(X_n, \|\cdot\|)$. For every $n \geq 1$, let $\eta_n : X \to X_n$ be a continuous linear map.*

Then (a) implies (b).

(a) There exists a densely defined dissipative operator $(A, \mathcal{D}(A))$ on X such that $A \subseteq ex - \lim A_n$ and such that the range condition holds for $(A, \mathcal{D}(A))$.

(b) The closure of $(A, \mathcal{D}(A))$ generates a strongly continuous contraction semigroup $\{T(t)\}_{t\geq 0}$ on X. Additionally, if $x_n \in X_n$ and $x \in X$ such that $\|x_n - \eta_n x\| \to 0$, then

$$\lim_{n\to\infty} \sup_{t\leq T} \|T_n(t)x_n - \eta_n T(t)x\| = 0$$

for all $T \geq 0$.

We will explore this extension only for functions spaces. First however, we will try to get around the range condition on the operator A. This will be achieved by using the theory of viscosity solutions.

2.2 VISCOSITY SOLUTIONS

In this section, we let $E \subseteq \mathbb{R}^d$ be some closed set. Consider a function $F : E \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. It is known that for many equations it is not possible to solve

$$F(x, u(x), \nabla u(x)) = 0, \qquad x \in E \tag{2.2.1}$$

classically. For example, consider the Eikonal equation on $E = [-1, 1]$ given by

$$\begin{cases} |u'(x)| - 1 = 0, \\ u(-1) = u(1) = 0. \end{cases} \tag{2.2.2}$$

Classical solutions to this problem do not exist by Rolle's theorem, so we have to resort to weak solutions. Clearly, there exists infinitely many 'solutions' that solve the Eikonal equation almost everywhere. For example, consider $u_1(x) = 1 - |x|$ and $u_2(x) = |x| - 1$.

Definition 2.2.1. We say that u is a *(viscosity) subsolution* of equation (2.2.1) if u is bounded, upper semi-continuous and if for every $f \in C^1(E)$ and $x_0 \in E$ such that $u(x_0) - f(x_0) = \sup_x u(x) - f(x)$, we have

$$F(x, u(x), \nabla f(x)) \leq 0.$$

We say that u is a *(viscosity) supersolution* of equation (2.2.1) if u is bounded, lower semi-continuous and if for every $f \in C^1(E)$ and $x_0 \in E$ such that $u(x_0) - f(x_0) = \inf_x u(x) - f(x)$, we have

$$F(x, u(x), \nabla f(x)) \geq 0.$$

We say that u is a *(viscosity) solution* of equation (2.2.1) if it is both a sub and a super solution.

Note that a solution u must be bounded and continuous, which is in contrast with the weak solution methods based on Sobolev spaces. This property turns out to be of use later. In the case that E is non-compact, there exists various other definitions of viscosity solutions in the literature. Because we will mainly focus on compact spaces, we stick to this definition. The motivation for changing the definition is the possibility that points $x_0 \in E$ such that $u(x_0) - f(x_0) = \sup_x u(x) - f(x)$ might not exist. The definition therefore ends up to be to weak.

Returning to the Eikonal equation, we check whether u_1, u_2 are viscosity solutions. Note that u_1 and u_2 are differentiable everywhere except in $x = 0$. So the point of interest is $x = 0$.

We start with u_1. Any function $f \in C^1(-1, 1)$ such that $u_1(0) - f(0) = \sup_x u_1(x) - f(x)$ satisfies $f'(0) \in [-1, 1]$ which implies that u_1 is a viscosity subsolution to the Eikonal equation. On the other hand there exists no $f \in C^1(-1,1)$ such that $u_1(0) - f(0) = \inf_x u_1(x) - f(x)$, which implies that u_1 is also a viscosity supersolution to the Eikonal equation.

Similarly to the argument that shows that u_1 is a supersolution, we find that u_2 is a subsolution to the Eikonal equation. However, for $f \in C^1(-1,1)$ such that $u_1(0) - f(0) = \inf_x u_1(x) - f(x)$ and $f'(0) \in (-1,1)$, we find $|f'(0)| - 1 \leq 0$, which implies that u_2 is not a supersolution.

In fact, one can show that u_1 is the unique solution to the Eikonal equation. This fact is established via the comparison principle.

Definition 2.2.2. We say that equation (2.2.1) satisfies the *comparison principle* if for a subsolution u and supersolution v we have $u \leq v$.

Note that if the comparison principle is satisfied, then a viscosity solution is unique. In Chapter 3, we will verify the comparison principle for the resolvent equation for some specific operators A. In these examples, the underlying state-space will be a compact subset of $\mathbb{R}^d$. We will proceed now with the discussion of the generation of semigroups under the assumption that the comparison principle is satisfied.

2.2.1 *Viscosity solutions to solve the resolvent equation*

We return to the situation where our goal is to show that an operator $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ generates a semigroup. Recall from Theorems 2.1.10 and 2.1.14 that we need to verify the range condition. In other words, for any fixed $\lambda > 0$, we need to find for a dense set of functions $h \subseteq C_b(E)$ a function $f \in \mathcal{D}(A)$ such that

$$(\mathbb{1} - \lambda A)f = h.$$

An alternative approach, noted in Section 5 of Crandall et al. [1984] and suggested as a starting point in Feng and Kurtz [2006] is to extend the domain of the generator. The goal of this extension is to obtain an operator that satisfies the range condition by construction. On the other hand, the extension must be such that it also satisfies the positive maximum principle. It turns out that viscosity solutions are especially suitable for this goal.

Pick some $h \in C_b(E)$ and $\lambda > 0$ and consider

$$u - \lambda A u = h. \tag{2.2.3}$$

Definition 2.2.3. We say that u is a *(viscosity) subsolution* of equation (2.2.3) if u is bounded, upper semi-continuous and if for every $f \in \mathcal{D}(A)$ and $x_0 \in E$ such that $u(x_0) - f(x_0) = \sup_x u(x) - f(x)$, we have

$$u - \lambda A f \leq h.$$

We say that u is a *(viscosity) supersolution* of equation (2.2.3) if u is bounded, lower semi-continuous and if for every $f \in \mathcal{D}(A)$ and $x_0 \in E$ such that $u(x_0) - f(x_0) = \inf_x u(x) - f(x)$, we have

$$u - \lambda A f \geq h.$$

We say that u is a *(viscosity) solution* of equation (2.2.3) if it is both a sub and a super solution.

To understand the relation between viscosity solutions of (2.2.3) and the positive maximum principle, consider a viscosity solution u to (2.2.3). This means that u is a candidate for the, for now undefined, resolvent $(\mathbb{1} - \lambda A)^{-1} h$. If this were the case, then $Au = \lambda^{-1}(u - h)$. The conditions for u being a viscosity solution, exactly turn out to show that the operator $\hat{A}$, defined by $A \cup (u, \lambda^{-1}(u-h))$ as a graph, satisfies the positive maximum principle. We check condition (a) of definition 2.1.11 for the extension. Let $(f, x_0) \in \mathcal{D}(A) \times E$ be such that $u(x_0) - f(x_0) = \sup_x u(x) - f(x)$. Because u is a viscosity subsolution, we obtain

$$\begin{aligned}\lambda \left[\hat{A}u(x_0) - \hat{A}f(x_0)\right] &= \lambda \left[\frac{u(x_0) - h(x_0)}{\lambda} - Af(x_0)\right] \\ &= u(x_0) - \lambda A f(x_0) - h(x_0) \leq 0,\end{aligned}$$

which proves that $\hat{A}u(x_0) - \hat{A}f(x_0) \leq 0$.

This indicates that if for every $h \in C_b(E)$ and $\lambda > 0$ there exists a unique viscosity solution to (2.2.3), the extension

$$\hat{A} := \bigcup_{\substack{\lambda > 0, \\ h \in C_b(E)}} \left\{ (u, \lambda^{-1}(u-h)) \,\middle|\, u - \lambda A u = h \text{ in the viscosity sense} \right\}.$$

is a suitable candidate for the construction of the semigroup associated to A. A priori, it is not clear, however, that $\hat{A}$ satisfies the positive maximum principle, or that $\hat{A}$ is the graph of an operator. For the first issue, note that we have only checked the positive maximum principle for pairs of functions (f, g) where the first is a viscosity solution and the second a classical solution. However, if one can find an explicit family of viscosity solutions to the family of equations (2.2.3), these issues can be resolved in a straightforward way.

2.2.2 *Approximation of semigroups*

Combining the discussion of last section with Theorem 2.1.14, we obtain a more elaborate approximation theorem on function spaces.

Let $\{E_n\}_{n\geq 1}$ be a sequence of compact metric spaces and let E be a compact metric space. For each n, we have some continuous map $\eta_n : E_n \to E$. This defines a map $\eta_n : C(E) \to C(E_n)$ by $\eta_n f = f \circ \eta_n$. We assume that $\lim_n E_n = E$, in the sense that for every $x \in E$, there exists $x_n \in E_n$ such that $\eta_n x_n \to x$.

The range condition in Theorem 2.1.14 will be replaced by the comparison principle for the resolvent equation. This replacement is very important because the verification of the range condition is often difficult or even impossible. The domain of the operator might be too small to be able to solve the resolvent equation. Even if this is the case, if the comparison principle is satisfied, there exists at most one unique extension of the operator that satisfies the range condition. This extension generates a semigroup via the Crandall-Liggett theorem. The result below is a special case of [Feng and Kurtz, 2006, Theorem 6.13].

Theorem 2.2.4. *Suppose that* $\lim_n E_n = E$. *For every* n, *let* $\{T_n\}_{n\geq 1}$ *be strongly continuous semigroups on* $(C(E_n), \|\cdot\|)$ *that have generators* $A_n \subseteq C(E_n)\times C(E_n)$ *in the sense of the Crandall-Liggett theorem 2.1.10, i.e.* $\{A_n\}_{n\geq 1}$ *are dissipative and satisfy the range condition. Suppose for every* n *that if* $(f, g) \in A_n$, *then* $(f + c, g) \in A$ *for all* $c \in \mathbb{R}$.

Suppose that $A \subseteq C(E) \times B(E)$ *such that* $A \subseteq ex - \lim A_n$. *Furthermore, assume that for all* $0 < \lambda < \lambda_0$, *there exists a dense set* $D_\lambda \subseteq C(E)$ *such that for* $h \in D_\lambda$ *the comparison principle holds for*

$$u - \lambda A u = h. \tag{2.2.4}$$

Then, we have

(a) For $h \in D_\lambda$, *there exits a unique viscosity solution of* (2.2.4), *which we will denote by* $R_\lambda h$.

(b) The map R_λ *is contractive and, hence, extends to a continuous map* $R_\lambda : C(E) \to C(E)$.

(c) The operator $\hat{A}$, *defined by*

$$\hat{A} := \left\{ (R_\lambda h, \lambda^{-1}(R_\lambda h - h)) \,\middle|\, \lambda > 0, h \in C_b(E) \right\}$$

extends A, *is dissipative and satisfies the range condition.*

(d) $\hat{A}$ *generates a strongly continuous semigroup* $\{T(t)\}_{t\geq 0}$ *in the sense of 2.1.10 and*

$$T(t)f = \lim_{n\to\infty} R^n_{\lambda/n} f$$

uniformly for t *in compact intervals.*

(e) *For* $f_n \in C(E_n)$ *and* $f \in C(E)$ *such that* $\|\eta_n f - f_n\| \to 0$*, we have*

$$\lim_{n\to\infty} \sup_{t\leq T} \|T_n(t)f_n - \eta_n T(t)f\| = 0$$

for all $T \geq 0$.

The theorem is proved using the following two lemmas. The first one gives the existence of viscosity sub- and supersolutions to (2.2.4), the second gives the contractiveness of the resolvent operator. The rest of the Theorem then essentially follows from Theorem 2.1.14.

Lemma 2.2.5. *Suppose that* $\lim_n E_n = E$. *Suppose for every* n *we have a dissipative operator* $A_n \subseteq C(E_n) \times C(E_n)$. *Suppose for every* n *that if* $(f, g) \in A_n$, *then* $(f + c, g) \in A$ *for all* $c \in \mathbb{R}$.

Now consider an operator $A \subseteq C(E) \times B(E)$ *such that* $A \subseteq ex - \lim A_n$. *Pick some* $\lambda > 0$ *and* $h \in C(E)$. *Let* $(f_n, g_n) \in H_n$ *and define* $h_n := f_n - \lambda g_n$. *Suppose that* $\|h_n - \eta_n h\| \to 0$, *then* $\overline{f}$ *and* $\underline{f}$ *defined by*

$$\overline{f}(x) = \inf_k \sup_{n\geq k} \left\{ f_n(z) \,\middle|\, z \in E_n : \, d(x, \eta_n(z)) \leq \frac{1}{k} \right\}$$

$$\underline{f}(x) = \sup_k \inf_{n\geq k} \left\{ f_n(z) \,\middle|\, z \in E_n : \, d(x, \eta_n(z)) \leq \frac{1}{k} \right\}$$

are sub, respectively super solutions to $f - \lambda A f = h$. *If the comparison principle holds for this equation, then* $f := \overline{f} = \underline{f}$ *and* $\|\eta_n f - f_n\| \to 0$. *Additionally, if* $(f_0, g_0) \in H$, *then* $\|f - f_0\| \leq \|h - (f_0 - \lambda g_0)\|$.

The last statement $\|f - f_0\| \leq \|h - (f_0 - \lambda g_0)\|$ implies that the resolvent $(\mathbb{1} - \lambda\hat{A})^{-1}$ of the operator $A \cup \{f, \lambda^{-1}(f - h)\}$ is contractive. To obtain a contractive resolvent for an extension with more than one viscosity solution, we need the following lemma.

Lemma 2.2.6. *Suppose the conditions of Lemma 2.2.5 are satisfied. Suppose that* $h^1, h^2 \in C(E)$ *and that there exists, for* $i \in \{1, 2\}$ *functions* $(f^i_n, g^i_n) \in A_n$ *such that* $h^i_n := f^i_n - \lambda g^i_n$ *satisfy* $\left\|h^i_n - \eta_n h^i\right\| \to 0$. *Then the unique viscosity solutions* f^i *to* $(\mathbb{1} - \lambda A)f = h^i$ *satisfy* $\left\|f^1 - f^2\right\| \leq \left\|h^1 - h^2\right\|$.

The proofs of these Lemmas can be found in Feng and Kurtz [2006]. The existence of viscosity sub- and super-solutions in some specific cases can also be obtained via variational methods, see Section 2.4.2.

2.3 MARKOV PROCESSES

We now turn to the theory of Markov processes, for which we follow the notation of Ethier and Kurtz [1986]. In this thesis, we will only consider Markov processes that take values in the Skorokhod space. To be well prepared to study Markov processes, we start with some general results on the space of probability measures and on the Skorokhod space.

2.3.1 *The space of probability measures*

Let (E, d) be a complete separable space. For the study of collections of measures in $\mathcal{P}(E)$, we equip $\mathcal{P}(E)$ with the *Prohorov* metric

$$\rho(\mu, \nu) = \inf \{\varepsilon > 0 \,|\, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \textit{ for all closed sets } A \subseteq E\},$$

where A^ε is the ε blow-up of E:

$$E^\varepsilon := \left\{ x \in E \,\middle|\, \inf_{y \in A} d(x, y) \leq \varepsilon \right\}.$$

The Prohorov metric inherits nice properties from d.

Theorem 2.3.1 (Theorem 3.1.7, Ethier and Kurtz [1986]). *If (E, d) is separable, then $(\mathcal{P}(E), \rho)$ is separable. If (E, d) is complete, then $(\mathcal{P}(E), \rho)$ is complete.*

Definition 2.3.2 (Tightness). We say that a collection of measures $M \subseteq \mathcal{P}(E)$ is *tight* if for every $\varepsilon > 0$ there exists a compact set $K \subseteq E$ such that

$$\sup_{\mu \in M} \mu(K^c) \leq \varepsilon.$$

Prohorov's celebrated theorem shows us that tightness of a family of measures is equivalent to compactness for the topology induced by the Prohorov metric.

Theorem 2.3.3 (Prohorov). *Let (E, d) be complete and separable and let $M \subseteq \mathcal{P}(E)$ be a collection of probability measures. Then the following are equivalent.*

(a) M *is tight.*

(b) For every $\varepsilon > 0$ *there exists a compact set* $K \subseteq E$ *such that*

$$\sup_{\mu \in M} \mu((K^\varepsilon)^c) \leq \varepsilon.$$

(c) The closure of M *in* $(\mathcal{P}(E), \rho)$ *is compact.*

We say that a net μ_α converges to μ *weakly* if we have for all $f \in C_b(E)$ that

$$\int f \mathrm{d}\mu_\alpha \to \int f \mathrm{d}\mu.$$

The weak topology and the Prohorov metric are nicely connected by the following theorem.

Theorem 2.3.4 (Portmanteau). *Let (E, d) be complete and separable. Then the weak topology is metrizable by the Prohorov metric. Furthermore, let $\mu_n \in \mathcal{P}(E)$ be a sequence of probability measures and let $\mu \in \mathcal{P}(E)$. The following are equivalent.*

(a) $\lim_{n\to\infty} \rho(\mu_n, \mu) = 0$.

(b) μ_n *converges to* μ *weakly.*

(c) For all closed sets $A \subseteq E$*, we have* $\limsup_{n\to\infty} \mu_n(A) \leq \mu(A)$.

(d) For all open sets $A \subseteq E$*, we have* $\liminf_{n\to\infty} \mu_n(A) \geq \mu(A)$.

We will say that a net of E valued random variables X_α converges to a random variable X in distribution or weakly if their push-forward measures μ_α on E converge weakly to the push-forward μ of X. We will, however, not distinguish between these two different definitions of convergence and use them interchangeably. For example, if we say that a family of random variables is weakly compact, we technically mean that the family of push-forward measures is weakly compact.

2.3.2 *The Skorokhod space*

For a complete separable metric space (E, d), we denote by $D_E(\mathbb{R}^+)$ the space of all functions $x : \mathbb{R}^+ \to E$ that are right continuous and have left limits. We will denote by $x(t-) = \lim_{r\uparrow t} x(r)$ the left limit of x at t.

Next, we equip $D_E(\mathbb{R}^+)$ with a metric that will turn $D_E(\mathbb{R}^+)$ into a complete separable metric space. Denote by Λ' the collection of strictly increasing functions $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ that are also surjective. Denote by $\Lambda \subseteq \Lambda'$ the set of Lipschitz continuous $\lambda \in \Lambda'$ such that

$$\gamma(\lambda) := \sup_{s>t\geq 0} \left| \log \frac{\lambda(s) - \lambda(t)}{s-t} \right| < \infty.$$

First define $q = d \wedge 1$, to obtain a bounded metric that is equivalent to d. Then, define for $x, y \in D_E(\mathbb{R}^+)$

$$r(x,y) := \inf_{\lambda \in \Lambda} \left[\gamma(\lambda) \vee \int_0^\infty e^{-u} r(x,y,\lambda,u) \mathrm{d}u \right],$$

where

$$r(x,y,\lambda,u) = \sup_{t\geq 0} q(x(t\wedge u), y(\lambda(t) \wedge u)).$$

The metric r inherits desirable properties from d.

Theorem 2.3.5 (Theorem 3.5.6, Ethier and Kurtz [1986]). *If (E,d) is separable, then $(D_E(\mathbb{R}^+), r)$ is separable. If (E,d) is complete, then $(D_E(\mathbb{R}^+), r)$ is complete.*

Additionally, even though we allow for jumps in the trajectories in $D_E(\mathbb{R}^+)$, this does not happen to often.

Lemma 2.3.6 (Lemmas 3.5.1 and 3.7.7 Ethier and Kurtz [1986]). *If $x \in D_E(\mathbb{R}^+)$, then x only has at most countable points of discontinuity. If X is a process with sample paths in $D_E(\mathbb{R}^+)$ then the complement of*

$$D(X) := \{t \geq 0 \,|\, \mathbb{P}[X(t) = X(t-)] = 1\}$$

is at most countable.

The following result is the basis under the study of weak convergence of Markov processes via their transition semigroups.

Theorem 2.3.7 (Theorem 3.7.8 Ethier and Kurtz [1986]). *Let E be separable and let $\{X_n\}_{n\geq 1}$ be processes with sample paths in $D_E(\mathbb{R}^+)$.*

(a) If $X_n \to X$ in distribution then

$$(X_n(t_1)\ldots, X_n(t_k)) \to (X(t_1), \ldots, X(t_k)) \tag{2.3.1}$$

in distribution for all finite sets $\{t_1, \ldots, t_k\} \subseteq D(X)$.

(b) *If the sequence* $\{X_n\}$ *is relatively compact and there exists a dense set* $D \subseteq D(X)$ *such that* (2.3.1) *holds for all* $\{t_1, \ldots, t_k\} \subseteq D$*, then* $X_n \to X$ *in distribution.*

For the convergence processes (b) is of interest as it implies a reduction in difficulty.

The verification of relative compactness of a sequence of processes X_n is a technical issue, that is normally carried out in two steps. First, compact containment is verified. This entails proving that up to a fixed time $T \geq 0$, the laws of the of $X_n(t)$, for $t \leq T$ and $n \geq 1$, are uniformly tight. The second step is to verify that for each f in some dense set in $C_b(E)$, the processes $t \mapsto f(X_n(t))$ are relatively compact in $D_{\mathbb{R}}(\mathbb{R}^+)$; an issue that is verifiable explicitly. See Sections 3.8 and 3.9 in Ethier and Kurtz [1986].

Proving that (2.3.1) holds goes well together with the structure of Markov processes. In particular, using the Markov property (2.3.1) can be reduced to the convergence of the distribution at time 0 and the convergence of the transition semigroups that we will introduce in Section 2.3.3 below.

2.3.3 *The semigroup of transition operators of a Feller process*

A filtration $\mathbb{F}$ on $D_E(\mathbb{R}^+)$is a collection of σ-algebras $\{\mathcal{F}_t\}_{t\geq 0}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{B}(D_E(\mathbb{R}^+))$ if $s \leq t$. Let X denote the coordinate process on $D_E(\mathbb{R}^+)$. The filtration $\mathbb{F}^X = \{\mathcal{F}_t^X\}_{t\geq 0}$ on $D_E(\mathbb{R}^+)$ generated by X is defined by

$$\mathcal{F}_t^X := \sigma\left\{X(s) \,|\, s \leq t\right\}.$$

We say that X is a *Markov process* if

$$\mathbb{P}\left[X(t+s) \in B \,\middle|\, \mathcal{F}_t^X\right] = \mathbb{P}\left[X(t+s) \in B \,|\, X(t)\right] \tag{2.3.2}$$

for all $s, t \geq 0$ and $B \in \mathcal{B}(E)$. Additionally, we say that X is *strongly Markov* if

$$\mathbb{P}\left[X(\tau+s) \in B \,\middle|\, \mathcal{F}_\tau^X\right] = \mathbb{P}\left[X(\tau+s) \in B \,|\, X(\tau)\right] \tag{2.3.3}$$

for all $s \geq 0$, $B \in \mathcal{B}(E)$ and stopping times τ, such that $\tau < \infty$ almost surely.

We say that $P(t, x, B)$ on $\mathbb{R}^+ \times E \times \mathcal{B}(E)$ is a *time-homogeneous transition function* if $P(t, x, \cdot) \in \mathcal{P}(E)$, $P(0, x, \cdot) = \delta_x$, $P(\cdot, \cdot, B)$ is measurable in the first two coordinates for all $B \in \mathcal{B}(E)$ and $P(t + s, x, B) =$

$\int P(s, y, B)P(t, x, \mathrm{d}y)$ for all $s, t \geq 0$, $x \in E$ and $B \in \mathcal{B}(E)$. Finally, we say that P is a transition function for a time-homogeneous Markov process X if

$$\mathbb{P}\left[X(t+s) \in B \,\middle|\, \mathcal{F}_t^X\right] = P(t, X(t), B)$$

for all $s, t \geq 0$ and $B \in \mathcal{B}(E)$.

In the case that E is compact, there is a well known functional analytic point of view on the transition function of a Markov process. Even though this approach can be extended to Polish spaces as we will show in Chapter 9, we restrict ourselves here to compact spaces.

Let $\{X(t)\}_{t\geq 0}$ be a Markov process on E whose trajectories take values in $D_E(\mathbb{R}^+)$. For $f \in M_b(E)$, we define the semigroup $\{S(t)\}_{t\geq 0}$ by

$$S(t)f(x) = \mathbb{E}\left[f(X(t)) \,|\, X(0) = x\right].$$

If $S(t)C(E) \subseteq C(E)$, we call X a *Feller process*. Clearly, $S(t)$ is contractive and the property that X takes values in $D_E(\mathbb{R}^+)$ combined with the fact that E is compact, yields the fact that $\{S(t)\}_{t\geq 0}$ is a strongly continuous semigroup on $(C(E), \|\cdot\|)$.

Analogously to the question posed for general semigroups, here the question can be asked whether for a given operators $(A, \mathcal{D}(A))$ one can find a Markov process such that the transition semigroup has A as its generator. The construction of a process via the Lumer-Phillips theorem for discrete interacting particle systems is carried out in Liggett [1985]. The construction of diffusion semigroups can be found in Ethier and Kurtz [1986] and Engel and Nagel [2000].

The theory of semigroups can also be applied to study approximation questions. Suppose that we have a sequence of Feller processes X_n with trajectories in $D_{E_n}(\mathbb{R}^+)$. We suppose that there are continuous maps $\eta_n : E_n \to E$ and we suppose that $\lim_n \eta_n E_n = E$ in the sense that for every $x \in E$, there are $x_n \in E_n$ such that $\eta_n(x_n) \to x$.

The map η_n induces a continuous map $\eta_n : (C(E), \|\cdot\|) \to (C(E_n), \|\cdot\|)$ by $\eta_n f(x) = f(\eta_n x)$. We find ourselves in the setting of Theorem 2.1.14 and 2.2.4. Thus, suppose that the semigroups and generators of X_n are denoted by $\{T_n(t)\}_{t\geq 0}$ and $(A_n, \mathcal{D}(A_n))$, and suppose there is some limiting operator $(A, \mathcal{D}(A))$ such that $A \subseteq ex - \lim A_n$ that generates a semigroup $\{T(t)\}_{t\geq 0}$. It follows that for $f_n \in C(E_n)$ and $f \in C(E)$ such that $\|\eta_n f - f\| \to 0$, we have

$$\lim_{n\to\infty} \sup_{t\leq T} \|T_n(t)f_n - \eta_n T(t)f\| = 0.$$

Suppose X is the process on $D_E(\mathbb{R}^+)$ with semigroup $\{T(t)\}_{t\geq 0}$ and suppose that $\eta_n X_n(0) \to X(0)$. By repeated conditioning of our processes, the convergence of the semigroups of conditional expectations yields the convergence of the finite dimensional distributions as required in (b) of Theorem 2.3.7. Moreover, the verification that the processes $t \mapsto f_n(X_n(t))$ are relatively compact in $D_{\mathbb{R}}(\mathbb{R}^+)$ follows as a consequence of the convergence $A \subseteq ex - \lim A_n$.

For non-compact spaces E the transition semigroup is in general not strongly continuous for the norm topology on $C_b(E)$.

Example 2.3.8. Consider standard Brownian Motion on $\mathbb{R}$. The transition functions $P(t, x, \cdot)$ are given in terms of their Radon-Nikodym derivative with respect to the Lebesgue measure:

$$\frac{P(t, x, \mathrm{d}y)}{\mathrm{d}y} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}.$$

The corresponding transition semigroup $\{S(t)\}_{t\geq 0}$ is strongly continuous on $(C_0(\mathbb{R}), \|\cdot\|)$. On the other hand, it is not strongly continuous on $C_b(E)$. Consider for example the function $f(x) = \sin(x^2)$. Because the oscillations of the function increase as $|x| \to \infty$, we do not have $\|S(t)f - f\| \to 0$.

A different, probabilistic, approach has been proposed by Stroock and Varadhan [1969a,b] and is based on the following observation.

Lemma 2.3.9. *Let E be a compact metric space and let $\mathbb{P}$ be a Markov measure on $D_E(\mathbb{R}^+)$. Denote by $\{S(t)\}_{t\geq 0}$ the transition semigroup and by $(A, \mathcal{D}(A))$ the generator of the process. Let $f \in \mathcal{D}(A)$, then*

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s$$

is a mean 0 $\mathbb{F}^X$ martingale.

Note that this result follows from Lemma 2.1.3 (d).

2.3.4 *The martingale problem*

The insight by Stroock and Varadhan was that the result of Lemma 2.3.9 can be taken as a starting point for the construction of a Markov process on Polish spaces E. For more information on this approach, see Stroock and Varadhan [1979], Ethier and Kurtz [1986].

Definition 2.3.10 (The martingale problem)**.** Let $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ be a linear operator. For $(A, \mathcal{D}(A))$ and a measure $\nu \in \mathcal{P}(E)$, we say that $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ solves the *martingale problem* for (A, ν) if for all $f \in \mathcal{D}(A)$

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s$$

is a mean 0 $\mathbb{F}^X$ martingale under $\mathbb{P}$, and if the law of $X(0)$ under $\mathbb{P}$ equals ν.

We denote the set of all solutions to the martingale problem, for varying initial measures ν, by $\mathcal{M}_A$. We say that *uniqueness* holds for the martingale problem if for every $\nu \in \mathcal{P}(E)$ the set $\{\mathbb{P} \in \mathcal{M}_A \,|\, \mathbb{P}X(0)^{-1} = \nu\}$ is empty or a singleton. Furthermore, we say that the martingale problem is *well-posed* if this set is a singleton.

Regarding well-posedness, we have the following result [Ethier and Kurtz, 1986, Theorem 4.5.11].

Theorem 2.3.11. *Let $A \subset C_b(E) \times C_b(E)$ and suppose that $\overline{\mathcal{D}(A)}$ contains an algebra that separates points and vanishes nowhere. Suppose that for each compact $K \subset E$, $\varepsilon > 0$ and $T > 0$, there exists a compact $K' = K'(K, \varepsilon, T)$ such that*

$$\mathbb{P}\left[X(t) \in K' \text{ for all } t < T, X(0) \in K\right] \geq (1-\varepsilon)\mathbb{P}\left[X(0) \in K\right]$$

for all $\mathbb{P} \in \mathcal{M}_A$. Additionally, assume that the martingale problem for A is well posed. Then the solutions to the Martingale problem are strong Markov processes corresponding to a semigroup that maps $C_b(E)$ into $C_b(E)$.

Note that unless E is compact, the result does not imply that the semigroup is strongly continuous for the supremum norm topology. In Chapter 9, we show that by using a suitably changed topology, the strong continuity can also be obtained for Markov processes on Polish E.

2.4 LARGE DEVIATIONS

We first start with some basic definitions on the large deviation principle. Afterwards, we will focus on large deviations for Markov processes.

Definition 2.4.1 (Rate function)**.** We say that $I : E \to [0, \infty]$ is a *rate function* if I is lower semi-continuous, i.e. for all $\alpha \in \mathbb{R}^+$, the set

$\{x \in E \,|\, I(x) \leq \alpha\}$ is closed. We say that I is *good* if the level sets $\{x \in E \,|\, I(x) \leq \alpha\}$ are compact.

Now consider a sequence of measures $\mu_n \in \mathcal{P}(E)$.

Definition 2.4.2 (Large deviation principle)**.** We say that the sequence $\{\mu_n\}_{n\geq 1}$ satisfies the *large deviation principle* with rate function I and normalisation r_n if the following inequalities hold for any set $A \in \mathcal{B}(E)$

$$\liminf_{n\to\infty} \frac{1}{r_n} \log \mu_n(A) \leq - \inf_{x\in A^\circ} I(x),$$
$$\limsup_{n\to\infty} \frac{1}{r_n} \log \mu_n(A) \leq - \inf_{x\in \overline{A}} I(x).$$

The study of large deviations involves many concepts that are analogous to concepts in the study of weak convergence. In a sense the large deviation principle is an exponential version of properties (c) and (d) of the Portmanteau theorem, 2.3.4. Note however that we assume both an upper and lower bound for the large deviation principle as opposed to the equivalence that holds in the case of weak convergence.

The functional form of the large deviation principle is given by Varadhan's lemma, see Theorem 4.3.1 in Dembo and Zeitouni [1998].

Theorem 2.4.3 (Varadhan)**.** *Suppose that the random variables Z_n, with laws $\mu_n \in \mathcal{P}(E)$ satisfy the large deviation principle with normalisation r_n and good rate function I. Let $f : E \to \mathbb{R}$ be continuous. Assume either the tail condition*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{r_n} \log \mathbb{E}\left[e^{r_n f(Z_n)} \mathbb{1}_{\{f(Z_n)\geq M\}}\right] = -\infty,$$

or the following moment condition for some $\gamma > 1$,

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \mathbb{E}\left[e^{\gamma r_n f(Z_n)}\right] < \infty.$$

Then

$$\lim_{n\to\infty} \frac{1}{r_n} \log \mathbb{E}\left[e^{r_n f(Z_n)}\right] = \sup_{x\in E} \{f(x) - I(x)\}.$$

As a corollary, we can prove large deviation principles for tilted measures.

Corollary 2.4.4. *Suppose that the sequence $\mu_n \in \mathcal{P}(E)$ satisfies the large deviation principle with normalisation r_n and good rate function I. Let $f : E \to \mathbb{R}$ be bounded and continuous. Then the sequence of probability measures ν_n, defined by*

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n} = e^{r_n f} Z_{n,f}^{-1},$$

where $Z_{n,f}$ is a normalising constant, satisfies the large deviation principle with normalisation r_n and good rate function

$$J(x) = I(x) - f(x) - \inf_{y \in E} \{I(y) - f(y)\}.$$

An inverse to Varadhan's lemma is given by Bryc's Theorem, see Theorem 4.4.2 in Dembo and Zeitouni [1998]. This shows that, as in the weak convergence setting, the functional and probabilistic versions of the theory are equivalent.

Also the concept of tightness has an exponential variant.

Definition 2.4.5 (Exponential tightness)**.** We say that a family of measures $\{\mu_n\}_{n \geq 1}$ on E is *exponentially tight* with rate r_n if for every $\alpha < \infty$ there exists a compact set $K_\alpha \subseteq E$ such that

$$\limsup_{n\to\infty} \frac{1}{r_n} \log \mu_n(K_\alpha^c) < -\alpha.$$

The contraction principle shows that a large deviation principle can be pushed forward, Theorem 4.2.1 in Dembo and Zeitouni [1998].

Theorem 2.4.6 (Contraction principle)**.** *Let E, F be Hausdorff spaces and let $f : E \to F$ be a continuous function. Consider a good rate function $I : E \to [0, \infty]$.*

(a) For each $y \in F$, define

$$I'(y) = \inf \{I(x) \,|\, x \in X, y = f(x)\},$$

where we set $\inf \emptyset = \infty$*. Then I' is a good rate function on Y.*

(b) If a sequence of probability measures μ_n satisfies the large deviation principle with rate function I on E, then the probability measures $f_\# \mu_n = \mu_n \circ f^{-1}$ satisfy the large deviation principle on F with rate function I'.

2.4.1 *Large deviations for Markov processes*

A main question in this thesis is whether the large deviation principle can be proven for the trajectories of Markov processes. In particular, we consider a sequence of Markov processes Y_n on spaces E_n. We assume the existence of some Polish space E, and connecting maps $\eta_n : E_n \to E$. We define the processes X_n by $X_n(t) := \eta_n(Y_n(t))$. Note that the process $X_n(t)$ is not necessarily Markov and might live in a space that is lower-dimensional than the original processes. For example Y_n could model spin flip dynamics on $\{-1,1\}^n$, whereas $X_n = n^{-1}\sum_{i\leq n} Y_n(i)$ models the empirical magnetisation in $[-1,1]$.

As in the study of weak convergence of processes, the large deviation principle on the Skorokhod space is a consequence of the large deviation principle of the finite dimensional distributions and exponential tightness. The following result is the large deviation analogue of Theorem 2.3.7.

Theorem 2.4.7 (Theorem 4.28 Feng and Kurtz [2006]). *Suppose $\{X_n\}_{n\geq 1}$ is exponentially tight in $D_E(\mathbb{R}^+)$. Suppose that for each set $\{t_1 = 0, t_1, \dots, t_k\}$ the random variables $\{X_n(t_1), \dots, X_n(t_k)\}_{n\geq 1}$ satisfy the large deviation principle in E^k with rate function $I_{t_1,\dots,t_k}$. Then $\{X_n\}_{n\geq 1}$ satisfy the large deviation principle in $D_E(\mathbb{R}^+)$ with good rate function*

$$I(x) = \sup_{\{t_i\}\subseteq \Delta_x^c} I_{t_1,\dots,t_k}(x(t_1),\dots,x(t_k)),$$

where $\{t_i\}$ is shorthand for all sets of the form $\{t_1,\dots,t_k\}$ and where Δ_x is the set of times where x is discontinuous.

As in the setting of weak convergence of Markov processes, the finite dimensional rate functions can be treated by conditioning.

Proposition 2.4.8 (Proposition 3.25 in Feng and Kurtz [2006]). *Suppose $\{X_n, Y_n\}_{n\geq 1}$ is exponentially tight in the product space $(F_1 \times F_2)$, where F_1, F_2 are Polish. Suppose $\mu_n \in \mathcal{P}(E_1 \times E_2)$ is the law of (X_n, Y_n) and suppose that $\mu_n(\mathrm{d}x, \mathrm{d}y) = \mu_n(\mathrm{d}y \,|\, x)\mu_n(\mathrm{d}x)$, where $\mu_n(\mathrm{d}y \,|\, x)$ is a version of a regular conditional probability. For $f \in C_b(F_2)$, denote*

$$\Lambda_{2,n}(f \,|\, x) := \frac{1}{n}\log\int e^{nf(y)}\mu_n(\mathrm{d}y \,|\, x).$$

Suppose there exists a continuous function $\Lambda_2(f \,|\, x)$ such that $\Lambda_{2,n}(f \,|\, \cdot) \to \Lambda_2(f \,|\, \cdot)$ uniformly on compact sets. For $x \in F_1$ and $y \in F_2$, define

$$I_2(y \,|\, x) = \sup_{f\in C_b(F_2)} f(y) - \Lambda_2(f \,|\, x).$$

If $\{X_n\}_{n\geq 1}$ *satisfies the large deviation principle on* F_1 *with good rate function* I_1*, then* $\{X_n, Y_n\}$ *satisfies the large deviation principle on* $F_1 \times F_2$ *with good rate function* $I(x,y) = I_1(x) + I_2(y \mid x)$*.*

The same result holds under the condition that there exists a sequence of sets $K_n \subseteq F_1$*, such that*

$$\lim_{n\to\infty} \sup_{x\in K_n} |\Lambda_2(f \mid x) - \Lambda_{2,n}(f \mid x)| = 0 \tag{2.4.1}$$

and $\lim_n n^{-1} \log \mu_n(K_n^c \times F_2) = -\infty$.

Given exponential tightness on the Skorokhod space, this proposition gives us the means to reduce the study of a finite dimensional large deviation problem to the study of the large deviations of the first marginal and a semigroup approximation issue.

Denote by $\{S_n(t)\}_{t\geq 0}$ the transition semigroups of the processes Y_n on E_n and denote their exponential transforms by $\{V_n(t)\}_{t\geq 0}$, i.e. $V_n(t) = n^{-1} \log S_n(t) e^{nf}$. Note that $\{V_n(t)\}_{t\geq 0}$ inherits the semigroup property of $\{S_n(t)\}_{t\geq 0}$. Suppose that there exists a limiting semigroup $\{V(t)\}_{t\geq 0}$ in the sense that

$$\lim_{n\to\infty} \|V_n(t)\eta_n f - \eta_n V(t) f\| = 0 \tag{2.4.2}$$

for all $f \in C_b(E)$. Note that the use of the uniform topology can be relaxed to the topology of uniform convergence on compact sets. Given the fact that we work with different spaces E_n, we restrict ourselves here to the uniform topology.

To prove the large deviation principle of $\{X_n\}_{n\geq 1}$ in $D_E(\mathbb{R}^+)$, where $X_n = \eta_n(Y_n)$ and $E = \lim_n \eta_n(E_n)$, we take an arbitrary finite collection of times $\{t_1 = 0, t_1, \ldots, t_k\}$. To apply Proposition 2.4.8, we start by decomposing E^k into $F_1 = E^{k-1}$ and $F_2 = E$. Taking $\Lambda_{2,n}(f, \cdot) = V_n(t_k - t_{k-1})f$, the condition in (2.4.1) follows from (2.4.2) where we have taken $K_n = (\eta_n^{-1}(E_n))^{k-1}$. Because the process Y_n takes values on E_n, we have $\mu_n(K_n^c) = 0$. Hence, the result follows if we know the large deviation principle for the set of times $\{t_1, \ldots, t_{k-1}\}$.

Iterating this process gives the large deviation principle under the condition that we have exponential tightness, the large deviation principle for the time 0 marginal, and the existence of a limiting semigroup.

The chain rule gives us that the generator of $V_n(t)$ should be $H_n f := \frac{1}{n} e^{-nf} A_n e^{nf}$. As in the case of linear semigroups, the existence of a limiting operator $H \subseteq ex - \lim H_n$ that generates a semigroup, is a major step towards the large deviation principle. The operators H_n can also be used to obtain exponential tightness.

Theorem 2.4.9 (Corollary 4.17 Feng and Kurtz [2006]). *Let E_n, E be compact. Let $F \subseteq C_b(E)$ separate points and let F be closed under addition. Suppose that for each $\lambda \in \mathbb{R}$ and $f \in F$, there exists $(f_n, g_n) \in H_n$ such that $\sup_n \|f_n\| < \infty$ and*

$$\lim_n \|f_n - \eta_n f\| = 0 \qquad \sup_n \sup_{x \in E_n} g_n(x) = C_\lambda(f) < \infty.$$

Then $\{X_n\}_{n\geq 1}$ is exponentially tight.

The full large deviation principle can, thus, also be established. The following result is a special case of Theorem 6.14 in Feng and Kurtz [2006].

Theorem 2.4.10. *Suppose that $\{E_n\}_{n\geq 1}$ is a sequence of compact separable metric spaces, and let E be a compact separable metric space. Let $\eta_n : E_n \to E$ be measurable maps. Assume that $E = \lim_n \eta_n(E_n)$. Let Y_n be a Markov process on E_n with generator A_n and transition semigroup $\{S_n(t)\}_{t\geq 0}$.*

Denote by $H_n f = e^{-nf} A_n e^{nf}$ and let $H \subseteq ex - \lim H_n$ where $\mathcal{D}(H)$ is closed under addition and dense in $(C(E), \|\cdot\|)$. Furthermore, suppose that the comparison principle is satisfied for $u - \lambda H u = h$ for all $\lambda > 0$ and $h \in C_b(E)$.

Set $X_n = \eta_n(Y_n)$ and suppose that $\{X_n(0)\}_{n\geq 1}$ satisfies the large deviation principle in E with good rate function I_0.

Then $(H, \mathcal{D}(H))$ generates a strongly continuous semigroup $\{V(t)\}_{t\geq 0}$ as in Theorem 2.2.4 such that

$$\sup_{t\leq T} \|V_n(t)\eta_n f - \eta_n V(t) f\| = 0$$

for all $T \geq 0$ and $f \in C(E)$.

Additionally, $\{X_n\}$ satisfies the large deviation principle in $D_E(\mathbb{R}^+)$ with good rate function I given by

$$I(x) = \sup_{t_1=0, t_2, \ldots, t_k} I_0(x(0)) + \sum_{i=1}^{k} I_{t_i - t_{i-1}}(x(t_i) \,|\, x(t_{i-1}))$$

and where $I_t(y \,|\, x) = \sup_{f \in C(E)} f(y) - V(t) f(x)$.

Proof. That $(H, \mathcal{D}(H))$ generates a semigroup $\{V(t)\}_{t \geq 0}$ and that we have convergence of the semigroups V_n to the limiting semigroup is a consequence of Theorem 2.2.4. Exponential tightness of $\{X_n\}_{n\geq 1}$ follows by Theorem 2.4.9. The large deviation principle and the form of the rate function follow from Theorem 2.4.7 and Proposition 2.4.8. □

2.4.2 *Variational representation of the rate function*

The following non-rigorous introductory section on variational representations of the rate function only considers the large deviation behaviour of sequences of processes $\{X_n(t)\}_{t\geq 0}$ that take values in a compact subset $E \subseteq \mathbb{R}^d$. We assume that all conditions for Theorem 2.4.10 are satisfied, but that $E \subseteq \mathbb{R}^d$. Additionally, we assume that $\mathcal{D}(H) = C^1(E)$ and that the operator $(H, \mathcal{D}(H))$ is of the form $Hf(x) = H(x, \nabla f(x))$, where $H : E \times \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and where $p \mapsto H(x,p)$ is convex for every fixed x and strictly convex for $x \in E^\circ$.

There are some additional technical assumptions on H that we will not mention here. The statements given below will be made rigorous in Chapter 3 below. In Chapter 6, we will prove a similar representation in a more abstract setting.

We introduce a new semigroup, the *Nisio semigroup* $\mathbf{V}(t)$, for which we will prove that $\mathbf{V}(t)f = V(t)f$ if $f \in C(E)$. The new semigroup is given as a variational problem where one optimises a pay-off $f(\gamma(t))$, but where a cost is paid that depends on the whole trajectory $\{\gamma(s)\}_{0\leq s\leq t}$. This cost is accumulated over time and is given by the Lagrangian. We define this Lagrangian by taking the Legendre-Fenchel transform of H:

$$\begin{aligned}\mathcal{L}(x,u) &:= \sup_{p\in\mathbb{R}^d} \{\langle p, u\rangle - H(x,p)\} \\ &= \sup_{f\in C^1(E)} \{\langle \nabla f(x), u\rangle - Hf(x)\}\end{aligned}$$

Because $p \mapsto H(x,p)$ is convex and continuous, it follows by the Fenchel Moreau theorem that also

$$Hf(x) = H(x, \nabla f(x)) = \sup_{u\in\mathbb{R}^d} \{\langle \nabla f(x), u\rangle - \mathcal{L}(x,u)\}.$$

Let $\mathcal{AC}$ be the space of absolutely continuous trajectories in $E \subseteq \mathbb{R}^d$, and set $\mathcal{AC}_x$ be the trajectories in $\mathcal{AC}$ that start in x. Using $\mathcal{L}$, we define the Nisio semigroup for measurable functions f on E:

$$\mathbf{V}(t)f(x) = \sup_{\gamma \in \mathcal{AC}_x} f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s.$$

We expect the semigroup $\mathbf{V}(t)$ to be related to $V(t)$ because of the following non-rigorous calculation. Consider a continuously differentiable function f, then

$$\begin{aligned}
\left[\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{V}(t)f(x)\right]_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}t}\left[\sup_{\gamma \in \mathcal{AC}_x} f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s\right]_{t=0} \\
&= \sup_{\gamma \in \mathcal{AC}_x} \frac{\mathrm{d}}{\mathrm{d}t}\left[f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s\right]_{t=0} \\
&= \sup_{\gamma \in \mathcal{AC}_x} \langle \nabla f(\gamma(0)), \dot{\gamma}(0)\rangle - \mathcal{L}(\gamma(0), \dot{\gamma}(0)) \\
&= \sup_u \langle \nabla f(x), u\rangle - \mathcal{L}(x, u) \\
&= Hf(x).
\end{aligned}$$

In other words, formally the generator of $\{\mathbf{V}(t)\}_{t\geq 0}$ equals the generator of $\{V(t)\}_{t\geq 0}$ which should imply that the semigroups coincide.

A rigorous approach to proving that $\mathbf{V}(t)f = V(t)f$ for $f \in C(E)$ is via the resolvent. By Theorem 2.2.4, we know that there exists a resolvent operator $\{R(\lambda, H)\}_{\lambda>0}$ such that for all h in a dense set $D_\lambda \subseteq C(E)$ we have that $R(\lambda, H)h$ is a viscosity solution to $u - \lambda H u = h$ and such that

$$V(t)f = \lim_n R\left(\frac{\lambda}{n}, H\right)^n f.$$

To connect the variational semigroup to the resolvent, we define the following variational resolvent, using the intuition that the resolvent is related to the behaviour of the system at an exponential random time:

$$\mathbf{R}(\lambda)h(x) = \sup_{\gamma \in \mathcal{AC}_x} \int_0^\infty \frac{1}{\lambda} e^{-\lambda^{-1}t}\left(h(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\right) \mathrm{d}t.$$

Following the first part of the proof of Theorem 8.27 in Feng and Kurtz [2006], we obtain the following important result.

Lemma 2.4.11. *For $\lambda > 0$, we have $\mathbf{R}(\lambda)C(E) \subseteq C(E)$ and $\mathbf{R}(\lambda)h$ is a viscosity solution of $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$.*

As a consequence of the last lemma, we see that if $f(x) - \lambda H(x, \nabla f(s)) - h(x) = 0$ satisfies the comparison principle for all $\lambda > 0$ and $h \in C(E)$, then $\mathbf{R}(\lambda)h = R(\lambda, H)h$. Additionally, in Lemma 8.18 in Feng and Kurtz [2006] it is proven for $f \in C(E)$ that

$$\mathbf{V}(t)f(x) = \lim_{n\to\infty} \mathbf{R}(n^{-1})^{\lfloor nt \rfloor} f(x),$$

which yields that $V(t)f = \mathbf{V}(t)f$. Using this identification, Feng and Kurtz obtain the following result. Note that their result holds in a more general setting.

Theorem 2.4.12 (Feng and Kurtz [2006], Corollary 8.29). *Let the Assumptions for Theorem 2.4.10 be satisfied. Let $E \subseteq \mathbb{R}^d$ be compact and let H be of the form $Hf(x) = H(x, \nabla f(x))$, where $H : E \times \mathbb{R}^d \to \mathbb{R}$ is continuous and convex in the second coordinate.*

Let $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$ satisfy the comparison principle for every $\lambda > 0$ and $h \in C(E)$.

Suppose that the sequence $X_n(0)$ satisfies the large deviation principle with good rate function I_0. Then, $\{X_n\}_{n\geq 1}$ is exponentially tight in $D_E(\mathbb{R}^+)$ and satisfies the large deviation principle with rate function I given by

$$I(\gamma) := \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{if } \gamma \notin \mathcal{AC}. \end{cases}$$

2.5 LOCALLY CONVEX SPACES

As noted in Section 2.3.3, the space $(C_b(E), \|\cdot\|)$ is not suitable for the study of Markov transition semigroups in the case that E is non-compact. This problem becomes clear on a more abstract level in the counterpart of the Riesz representation theorem. For a compact metric space E, the continuous dual space of $(C(E), \|\cdot\|)$ is given by the space $\mathcal{M}(E)$ of Radon measures of bounded total variation on E. For non-compact E, however, this identification breaks down and the dual space of $(C_b(E), \|\cdot\|)$ is strictly larger than the space of Radon measures of bounded total variation.

To restore the connection of $C_b(E)$ with $\mathcal{M}(E)$ for Polish E, we need to consider the strict topology β, which is a weaker locally convex topology on $C_b(E)$.

We will introduce some general notation from the theory of locally convex spaces, see Köthe [1969] or Treves [1967].

Let X be some vector space. We say that $p : X \to \mathbb{R}^+$ is a *semi-norm* if for $\lambda \in \mathbb{R}$ and $x, y \in X$, we have

(a) $p(\lambda x) = |\lambda| p(x)$,

(b) $p(x + y) \leq p(x) + p(y)$.

Definition 2.5.1 (Locally convex spaces)**.** A *locally convex space* (X, τ) is a vector space X, equipped with a topology τ which is generated by a family of semi-norms $\{p_i\}_{i \in I}$ for some index set I. In other words, a basis for the topology is given by the collection of sets

$$\{x \in X \,|\, p_i(x_0 + x) < c\}, \quad x_0 \in X, c > 0, i \in I.$$

We will always assume that τ is Hausdorff, which is the case if $\bigcap_{i \in I} \bigcap_{c > 0} \{x \in X \,|\, p_i(x) < c\} = \{0\}$.

We say that a set $A \subseteq X$ is convex if for all $x, y \in A$ and λ such that $|\lambda| \leq 1$ the element $\lambda x + (1 - \lambda) y \in A$. We say that A is absolutely convex if it is convex and if for every $\lambda \in [-1, 1]$ and $x \in A$, we have $\lambda x \in A$. We say that A is bounded if for any sequence $\{x_n\}_{n \geq 1}$ in A and sequence of non-negative real numbers λ_n such that $\lambda_n \to 0$, we have that $\lambda_n x_n \to 0$.

We will say that (X, τ) is complete if it is complete as a uniform space. In other words, if every Cauchy net converges. A Cauchy net $\{x_\alpha\}_{\alpha \in J}$ is a net such that for every τ continuous semi-norm p, there exists $\beta \in J$ such that for all $\alpha_1, \alpha_2 \geq \beta$, we have $p(x_{\alpha_1} - x_{\alpha_2}) \leq 1$. We say that (X, τ) is sequentially complete, if every Cauchy sequence converges.

By X^* we denote the algebraic dual of X, the space of all linear maps $x^* : X \to \mathbb{R}$. By X' we denote the continuous dual of X, the space of all maps $x' \in X^*$ that are continuous for τ. Finally, X^+ is the sequential dual of X:

$$X^+ := \{f \in X^* \,|\, f(x_n) \to 0, \text{ for every sequence } x_n \in X \text{ converging to } 0\}.$$

For any element $x^* \in X^*$, we write $\langle x, x^* \rangle = x^*(x)$ for the canonical pairing between X and X^*.

We say that a family $\mathfrak{S} \subseteq X'$ is τ equi-continuous on X if there exists a τ continuous semi-norm p such that

$$\sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle| \leq p(x).$$

The dual paring of X and X' can be used to construct topologies on both spaces. Denote by $\sigma(X, X')$ the weak topology on X, which is generated by all semi-norms $p_{x'}(x) = |\langle x, x'\rangle|$. Note that this topology is Hausdorff by the Hahn-Banach theorem and that it is weaker than τ. Similarly, one can define the weak topology $\sigma(X', X)$ on X'. A first result on the weak topology on X' is that $(X', \sigma(X', X))' = X$, which symmetrises many of the results that will follow in X and X'.

Let $A \subseteq X$. We denote by $A^\circ \subseteq X'$ the *polar* of A, which is defined by

$$A^\circ := \left\{x' \in X' \,\middle|\, \forall x \in A : |\langle x, x'\rangle| \le 1\right\}.$$

The notation for the polar is the same as for the interior of a set. These two notions are distinct and usually it is clear from the context which of the two is meant. Otherwise, we will explicitly state which of two notions is used. Note that A° is absolutely convex. Similarly, we define the polar $B^\circ \subseteq X$ of a set $B \subseteq X'$. For $A \subseteq X$, we denote $A^{\circ\circ} = (A^\circ)^\circ$ for the bipolar of A in X. The next theorem is a special case of the Fenchel-Moreau theorem in convex analysis.

Theorem 2.5.2 (Bipolar theorem). *Let $A \subseteq X$ be weakly closed and absolutely convex. Then $A = A^{\circ\circ}$. In particular, if p is a τ continuous semi-norm on X, then we have*

$$p(x) = \sup_{x' \in \{x | p(x) \le 1\}^\circ} |\langle x, x'\rangle|,$$

and the set $\{x \,|\, p(x) \le 1\}^\circ$ is τ equi-continuous.

Note that this implies that the polar A° of a neighbourhood A of 0 is always equi-continuous. In fact, we have the following result.

Lemma 2.5.3. *The following are equivalent.*

(a) $B \subseteq X'$ is τ equi-continuous.

(b) There exists a τ-neighbourhood A of 0 in X such that $B \subseteq A^\circ$.

The following is a well known theorem on weak compactness of polars in the dual space.

Theorem 2.5.4 (Bourbaki-Alaoglu). *If $A \subseteq X$ is a neighbourhood of 0 for the topology τ, then U° is compact in $(X', \sigma(X', X))$.*

Combining the Bipolar and the Bourbaki-Aloaglu theorem, we see that all τ continuous semi-norms are of the form

$$p(x) = \sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle|$$

for some collection $\mathfrak{S}$ of weakly compact and absolutely convex sets in X'. It is not true however, that all weakly compact and absolutely convex sets generate a τ continuous semi-norm in this way. This can be seen by considering $\sigma(X, X')$ in the case that (X, τ) is a Banach space.

2.5.1 *Admissible topologies*

To study the question of topologies generated by absolutely convex sets in the dual, we identify the largest possible class of semi-norms. For a set $B \subseteq X'$, we set $p_B(x) := \sup_{x' \in B} |\langle x, x' \rangle|$. Note that a set $B \subseteq X'$ is weakly bounded if for every $x \in X$, we have $\sup_{x' \in B} |\langle x, x' \rangle| < \infty$. It follows that p_B is a semi-norm if and only if B is weakly bounded.

We say that a collection $\mathcal{B}$ of sets $B \subseteq X'$ is *total* if the linear span of the union of these subsets is weakly dense in X'.

Lemma 2.5.5. *Suppose that $\mathcal{B}$ is a collection of weakly bounded subsets of X', then the collection of semi-norms $\{p_B\}_{B \in \mathcal{B}}$ defined a locally convex Hausdorff space if and only if $\mathcal{B}$ is total.*

Using total collections of weakly bounded subsets, we can define various topologies on X. We single out three special cases:

- The weak topology $\sigma(X, X')$ where $\mathcal{B}$ is the collection singletons in X'.
- The Mackey topology $\mu(X, X')$, where $\mathcal{B}$ is the collection of all weakly compact absolutely convex sets in X'.
- The strong topology $\beta(X, X')$ where $\mathcal{B}$ is the collection of all weakly bounded sets in X'.

Similarly, we can define $\sigma(X', X)$, $\mu(X', X)$ and $\beta(X', X)$.

By the Bourbaki-Alaoglu theorem, we have that the topology τ satisfies $\sigma(X, X') \subseteq \tau \subseteq \mu(X, X')$. Moreover, we have the following result.

Theorem 2.5.6 (Mackey-Arens)**.** *Consider the locally convex space (X, τ). The weak topology on X is the weakest locally convex topology on X such that $(X, \sigma(X, X'))' = X'$, whereas the Mackey topology is the strongest locally convex topology such that $(X, \mu(X, X'))' = X'$.*

We conclude this section with two general definitions.

Definition 2.5.7. We say that (X, τ) is *barrelled* if $\mu(X, X') = \beta(X, X')$. We say that (X, τ) is *strong Mackey* if every weakly compact set in X' is contained in a absolutely convex weakly compact set.

Note that both properties are of the form that a topology in general stronger than the Mackey topology, in fact, coincides with the Mackey topology. The first property is very strong and holds for example for Banach spaces. The second property is quite a bit weaker but occurs naturally as well. It is particularly interesting as the notions of weak compactness and equicontinuity coincide. This fact is for example used for the proof of Lemma 8.2.2.

2.5.2 *The strict topology on the space of continuous and bounded functions*

Having introduced the general terminology of locally convex spaces, we are able to introduce a particularly interesting space for the purposes of measure theory.

We return to the setting where (E, d) is a Polish space. For every compact set $K \subseteq E$, define the semi-norm $p_K(f) := \sup_{x \in K} |f(x)|$. The *compact-open* topology κ on $C_b(E)$ is generated by the semi-norms $\{p_K \mid K \text{ compact}\}$. We define a new collection of semi-norms in the following way. Pick a non-negative sequence a_n in $\mathbb{R}$ such that $a_n \to 0$. Also pick an arbitrary sequence of compact sets $K_n \subseteq E$. Define

$$p_{(K_n),(a_n)}(f) := \sup_n a_n p_{K_n}(f). \tag{2.5.1}$$

The *strict* topology β, defined on $C_b(E)$ is generated by the semi-norms

$$\left\{p_{(K_n),(a_n)} \mid K_n \text{ compact}, 0 < a_n \to 0\right\},$$

see Theorem 3.1.1 in Wiweger [1961] and Theorem 2.4 in Sentilles [1972]. Note that in the latter paper, the topology introduced here is called the substrict topology. However, Sentilles shows in Theorem 9.1 that the strict and the substrict topology coincide when the underlying space E is Polish. Note that if additionally (E, d) is locally compact, then the strict topology can also be given by the collection of semi-norms

$$p_g(f) := \|fg\|$$

where g ranges over $C_0(E)$.

The strict topology is the 'right' generalisation of the norm topology on $C(E)$ for compact metric E to the more general context of Polish spaces. We give some of the properties of β.

Theorem 2.5.8. *Let E be Polish. The locally convex space $(C_b(E), \beta)$ satisfies the following properties.*

(a) $(C_b(E), \beta)$ is complete, strong Mackey and the continuous dual space coincides with the space of Radon measures on E of bounded total variation.

(b) $(C_b(E), \beta)$ is separable.

(c) For any locally convex space (Y, τ_Y) and β to τ_Y sequentially equi-continuous family $\{T_i\}_{i \in I}$ of maps $T_i : (C_b(E), \beta) \to (Y, \tau_Y)$, the family I is β to τ_Y equi-continuous.

(d) The norm bounded and β bounded sets coincide. Furthermore, on norm bounded sets β and κ coincide.

(e) Stone-Weierstrass: Let M be an algebra of functions in $C_b(E)$. If M vanishes nowhere and separates points, then M is β dense in $C_b(E)$.

(f) Arzelà-Ascoli: A set $M \subset C_b(E)$ is β compact if and only if M is norm bounded and M is an equi-continuous family of functions.

Proof. (a) and (c) follow from Theorems 9.1 and 8.1 in Sentilles [1972], Theorem 7.4 in Wilansky [1981], Corollary 3.6 in Webb [1968] and Krein's theorem[Köthe, 1969, 24.5.(4)]. (b) follows from Theorem 2.1 in Summers [1972]. (d) follows by Theorems 4.7, 2.4 in Sentilles [1972] and 2.2.1 in Wiweger [1961]. (e) is proven in Theorem 2.1 and Corollary 2.4 in Haydon [1976]. (f) follows by the Arzelà-Ascoli theorem for the compact-open topology, Theorem 8.2.10 in Engelking [1989], and (d). □

The strict topology is used in Chapter 9, where we show that if the martingale problem on a Polish space is well-posed and the associated process satisfies a compact containment condition, then the corresponding transition semigroup is strongly continuous for the strict topology and the generator of this semigroup extends the operator in the martingale problem.

Part II

LARGE DEVIATIONS OF MARKOV PROCESSES AND APPLICATIONS

3 LARGE DEVIATIONS FOR JUMP PROCESSES WITH MEAN-FIELD INTERACTION

In this chapter, we consider the path-space large deviations of two models of Markov jump processes with mean-field interaction. The results presented are based on:

Richard Kraaij. Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton–Jacobi equation. *Journal of Statistical Physics*, 164(2):321–345, 2016b. ISSN 1572-9613. doi: 10.1007/s10955-016-1542-8.

In both examples, we have n particles or spins that evolve as a pure jump process, where the jump rates of the individual particles depend on the empirical distribution of all n particles. We prove the large deviation principle (LDP) for the trajectory of these empirical quantities and show that the rate function is in Lagrangian form. The first set of models that we consider are conservative models that generalize the Ehrenfest model. In the one dimensional setting, this model is also known as the Moran model without mutation or selection. For these models, the empirical quantity of interest for large n is the empirical magnetisation. The second class of models are jump processes of Glauber type such as Curie-Weiss spin flip dynamics. In this case, the empirical measure is given by

$$\mu_n(t) := \frac{1}{n} \sum_{i \leq n} \delta_{\sigma_i(t)},$$

where $\sigma_i(t) \in \{1, \ldots, d\}$ is the state of the i-th spin at time t. Under some appropriate conditions, the trajectory $\mu_n(t)$ converges as $n \to \infty$ to $\mu(t)$, the solution of a McKean-Vlasov equation, which is a generalization of the linear Kolmogorov forward equation which would appear in the case of independent particles.

For the second class of models, we obtain a large deviation principle for the trajectory of these empirical measures on the space $D_{\mathcal{P}(\{1,\ldots,d\})}(\mathbb{R}^+)$ of càdlàg paths on $E := \mathcal{P}(\{1, \ldots, d\})$ of the form

$$\mathbb{P}\left[\{\mu_n(t)\}_{t \geq 0} \approx \gamma\right] \approx e^{-nI(\gamma)}$$

where

$$I(\gamma) = I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s$$

for trajectories γ that are absolutely continuous and $I(\gamma) = \infty$ otherwise. In particular, $I(\gamma) = 0$ for the solution γ of the limiting McKean-Vlasov equation. The *Lagrangian* $\mathcal{L} : E \times \mathbb{R}^d \to \mathbb{R}^+$ is defined as the Legendre transform of a *Hamiltonian* $H : E \times \mathbb{R}^d \to \mathbb{R}$ that can be obtained via a limiting procedure

$$H(x, \nabla f(x)) = Hf(x) = \lim_n \frac{1}{n} e^{-nf} A_n e^{nf}. \tag{3.0.1}$$

Here A_n is the generator of the Markov process of $\{\mu_n(t)\}_{t\geq 0}$. More details on the models and definitions follow shortly in Section 3.1.

Recent applications of the path-space large deviation principle are found in the study of mean-field Gibbs-non-Gibbs transitions, see e.g. Ermolaev and Külske [2010], van Enter et al. [2010] or the microscopic origin of gradient flow structures, see e.g. Adams et al. [2013], Mielke et al. [2014]. Other authors have considered the path-space LDP in various contexts before, see for example Freidlin and Wentzell [1998], Comets [1989], Léonard [1995], Dai Pra and den Hollander [1996], Feng [1994], Budhiraja et al. [2011], Borkar and Sundaresan [2012]. A comparison with these results follows in Section 3.1.6.

The novel aspect of the results in this chapter with respect to large deviations for jump processes is an approach via a class of *Hamilton-Jacobi* equations. In Feng and Kurtz [2006], a general strategy is proposed for the study for large deviations of trajectories which is based on an extension of the theory of convergence of non-linear semigroups by the theory of viscosity solutions.As in the theory of weak convergence of Markov processes, this program is carried out in three steps, first one proves convergence of the generators, i.e. (3.0.1), secondly one shows that H is indeed the generator of a semigroup. The third step is the verification of the exponential compact containment condition, which for our compact state-spaces is immediate, that yields, given the convergence of generators, exponential tightness on the Skorokhod space. This final step reduces the proof of the large deviation principle on the Skorokhod space to that of the finite dimensional distributions, which can then be proven via the first two steps.

Showing that H generates a semigroup is non-trivial and follows for example by showing that the Hamilton-Jacobi equation

$$f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0 \tag{3.0.2}$$

has a unique solution f for all $h \in C(E)$ and $\lambda > 0$ in the viscosity sense. It is exactly this problem that is the main focus of this chapter. An extra bonus of this approach is that the conditions on the Markov processes for finite N are weaker than in previous studies, and allow for singular behaviour in the jump rate for a particle to move from a to b in boundary regions when the empirical average $\mu(a)$ is close to 0.

This approach via the Hamilton-Jacobi equation has been carried out in Feng and Kurtz [2006] for Levy processes on $\mathbb{R}^d$, systems with multiple time scales and for stochastic equations in infinite dimensions. In Deng et al. [2011], the LDP for a diffusion process on $(0, \infty)$ is treated with singular behaviour close to 0.

As a direct consequence of our large deviation principle, we obtain a straightforward method to find Lyapunov functions for the limiting McKean-Vlasov equation. If A_n is the linear generator of the empirical quantity of interest of the n-particle process, the operator A obtained by $Af = \lim_n A_n f$ can be represented by $Af(\mu) = \langle \nabla f(\mu), \mathbf{F}(\mu) \rangle$ for some vector field $\mathbf{F}$. If solutions to

$$\dot{\mu}(t) = \mathbf{F}(\mu(t)) \tag{3.0.3}$$

are unique for a given starting point and if the empirical measures $\mu_n(0)$ converges to $\mu(0)$, the empirical measures $\{\mu_n(t)\}_{t \geq 0}$ converge almost surely to a solution $\{\mu(t)\}_{t \geq 0}$ of (3.0.3). In Section 3.1.4, we will show that if the stationary measures of A_n satisfy a large deviation principle on $\mathcal{P}(\{1, \ldots, d\})$ with rate function I_0, then I_0 is a Lyapunov function for (3.0.3).

This chapter is organised as follows. In Section 3.1, we introduce the models and state our results. Additionally, we give some examples to show how to apply the theorems. In Section 3.2, we recall the main results from Feng and Kurtz [2006] that relate the Hamilton-Jacobi equations (3.0.2) to the large deviation problem. Additionally, we verify conditions from Feng and Kurtz [2006] that are necessary to obtain our large deviation result with a rate function in Lagrangian form, in the case that we have uniqueness of solutions to the Hamilton-Jacobi equations. Finally, in Section 3.3 we prove uniqueness of viscosity solutions to (3.0.2).

3.1 MAIN RESULTS

3.1.1 *Two models of interacting jump processes*

We do a large deviation analysis of the trajectory of the empirical magnetization or distribution for two models of interacting spin-flip systems. We replace the notation of the state-space E by E_1 and E_2. The first setting is a d-dimensional Ehrenfest model.

Generalized Ehrenfest model in d-dimensions.

Consider d-dimensional spins $\sigma = (\sigma(1), \dots, \sigma(n)) \in (\{-1,1\}^d)^n$. For example, we can interpret this as n individuals with d types, either being -1 or 1. For $k \leq n$, we denote the i-th coordinate of $\sigma(k)$ by $\sigma_i(k)$. Set $x_n = (x_{n,1}, \dots, x_{n,d}) \in E_1 := [-1,1]^d$, where $x_{n,i} = x_{n,i}(\sigma) = \frac{1}{n}\sum_{j=1}^n \sigma_i(j)$ the empirical magnetisation in the i-th spin. For later convenience, denote by $E_{1,n}$ the discrete subspace of E_1 which is the image of $(\{-1,1\}^d)^n$ under the map $\sigma \mapsto x_n(\sigma)$, i.e. $E_{1,n} = x_n((\{-1,1\}^d)^n)$. The spins evolve according to mean-field Markovian dynamics with generator $\mathcal{A}_n$:

$$\begin{aligned}\mathcal{A}_n f(\sigma) = &\sum_{i=1}^d \sum_{j=1}^n \mathbb{1}_{\{\sigma_i(j)=-1\}} r_{n,+}^i(x_n(\sigma)) \left[f(\sigma^{i,j}) - f(\sigma)\right] \\ &+ \sum_{i=1}^d \sum_{j=1}^n \mathbb{1}_{\{\sigma_i(j)=1\}} r_{n,-}^i(x_n(\sigma)) \left[f(\sigma^{i,j}) - f(\sigma)\right].\end{aligned}$$

The configuration $\sigma^{i,j}$ is obtained by flipping the i-th coordinate of the j-th spin. The functions $r_{n,+}^i, r_{n,-}^i$ are non-negative and represent the jump rate of the i-th spin flipping from a -1 to 1 or vice-versa.

The empirical magnetisation x_n itself also behaves Markovian and has generator $A_n : C(E_{1,n}) \to C(E_{1,n})$ which satisfies $A_n f(x_n(\sigma)) := \mathcal{A}_n(f \circ x_n)(\sigma)$ and is given by

$$\begin{aligned}A_n f(x) = \sum_{i=1}^d \Bigg\{ &n\frac{1-x_i}{2} r_{n,+}^i(x) \left[f\left(x + \frac{2}{n}e_i\right) - f(x)\right] \\ &+ n\frac{1+x_i}{2} r_{n,-}^i(x) \left[f\left(x - \frac{2}{n}e_i\right) - f(x)\right]\Bigg\},\end{aligned}$$

where e_i the vector consisting of 0's, and a 1 in the i-th component.

Under suitable conditions on the rates $r^i_{n,+}$ and $r^i_{n,-}$, we will derive a large deviation principle for the trajectory $\{x_n(t)\}_{t\geq 0}$ in the Skorokhod space $D_{E_1}(\mathbb{R}^+)$ of right continuous E_1 valued paths that have left limits.

Systems of Glauber type with d states.

We will also study the large deviation behaviour of copies of a Markov process on $\{1,\dots,d\}$ that evolve under the influence of some mean-field interaction. Here $\sigma = (\sigma(1),\dots,\sigma(n)) \in \{1,\dots,d\}^n$ and the empirical distribution μ is given by $\mu_n(\sigma) = \frac{1}{n}\sum_{i\leq n}\delta_{\sigma(i)}$ which takes values in

$$E_{2,n} := \left\{ \mu \in \mathcal{P}(E_2) \,\middle|\, \mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \text{ for some } x_i \in \{1,\dots,d\} \right\}.$$

Of course, this set can be seen as discrete subset of $E_2 := \mathcal{P}(\{1,\dots,d\}) = \{\mu \in \mathbb{R}^d \,|\, \mu_i \geq 0, \sum_i \mu_i = 1\}$. We take some n-dependent family of jump kernels $r_n : \{1,\dots,d\} \times \{1,\dots,d\} \times E_{2,n} \to \mathbb{R}^+$ and define Markovian evolutions for σ by

$$\mathcal{A}_n f(\sigma(1),\dots,\sigma(n)) \\ = \sum_{i=1}^n \sum_{b=1}^d r_n\left(\sigma(i), b, \frac{1}{n}\sum_{i=1}^n \delta_{\sigma(i)}\right)\left[f(\sigma^{i,b}) - f(\sigma)\right],$$

where $\sigma^{i,b}$ is the configuration obtained from σ by changing the i-th coordinate to b. Again, we have an effective evolution for μ_n, which is governed by the generator

$$A_n f(\mu) = n\sum_{a,b}\mu(a) r_n(a,b,\mu)\left[f\left(\mu - n^{-1}\delta_a + n^{-1}\delta_b\right) - f(\mu)\right].$$

As in the first model, we will prove, under suitable conditions on the jump kernels r_n a large deviation principle in n for $\{\mu_n(t)\}_{t\geq 0}$ in the Skorokhod space $D_{E_2}(\mathbb{R}^+)$.

3.1.2 *Large deviation principles*

The main results in this chapter are the two large deviation principles for the two sets of models introduced above. To be precise, we say that the sequence $x_n \in D_{E_1}(\mathbb{R}^+)$, or for the second case $\mu_n \in D_{E_2}(\mathbb{R}^+)$, satisfies the large deviation principle with rate function $I : D_{E_1}(\mathbb{R}^+) \to [0,\infty]$ if I is lower semi-continuous and the following two inequalities hold:

(a) For all closed sets $G \subseteq D_{E_1}(\mathbb{R}^+)$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}[\{x_n(t)\}_{t \geq 0} \in G] \leq - \inf_{\gamma \in G} I(\gamma).$$

(b) For all open sets $U \subseteq D_{E_1}(\mathbb{R}^+)$, we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}[\{x_n(t)\}_{t \geq 0} \in U] \geq - \inf_{\gamma \in U} I(\gamma).$$

For the definition of the Skorokhod topology defined on $D_{E_1}(\mathbb{R}^+)$, see for example Ethier and Kurtz [1986]. We say that I is *good* if the level sets $I^{-1}[0, a]$ are compact for all $a \geq 0$.

Carrying out the procedure in (3.0.1) for our two sets of models, we obtain, see Lemma 3.2.1 below, operators $(H, \mathcal{D}(H))$, $\mathcal{D}(H) = C^1(E)$ that are of the form $Hf(x) = H(x, \nabla f(x))$, $H : E \times \mathbb{R}^d \to \mathbb{R}$. These are the Hamiltonians that appear in Theorems 3.1.1 and 3.1.3.

For a trajectory $\gamma \in D_{E_1}(\mathbb{R})$, we say that $\gamma \in \mathcal{AC}$ if the trajectory is absolutely continuous. For the d-dimensional Ehrenfest model, we have the following result.

Theorem 3.1.1. *Suppose that there exists a family of continuous functions $v_+^i, v_-^i : E_1 \to \mathbb{R}^+$, $1 \leq i \leq d$, such that*

$$\lim_{n \to \infty} \sup_{x \in E_{1,n}} \sum_{i=1}^d \left| \frac{1 - x_i}{2} r_{n,+}^i(x) - v_+^i(x) \right| + \left| \frac{1 + x_i}{2} r_{n,-}^i(x) - v_-^i(x) \right| = 0. \tag{3.1.1}$$

Suppose that for every i, the functions v_+^i and v_-^i satisfy the following.

The rate v_+^i is identically zero or we have the following set of conditions.

(a) $v_+^i(x) > 0$ if $x_i \neq 1$.

(b) For $z \in [-1, 1]^d$ such that $z_i = 1$, we have $v_+^i(z) = 0$ and for every such z there exists a neighbourhood U_z of z on which there exists a decomposition $v_+^i(x) = v_{+,z,\dagger}^i(x_i) v_{+,z,\ddagger}^i(x)$, where $v_{+,z,\dagger}^i$ is decreasing and where $v_{+,z,\ddagger}^i$ is continuous and satisfies $v_{+,z,\ddagger}^i(z) \neq 0$.

The rate v_-^i is identically zero or we have the following set of conditions.

(a) $v_-^i(x) > 0$ if $x_i \neq -1$.

(b) For $z \in [-1, 1]^d$ such that $z_i = -1$, we have $v_-^i(z) = 0$ and for every such z there exists a neighbourhood U_z of z on which there exists a decomposition $v_-^i(x) = v_{-,z,\dagger}^i(x_i) v_{-,z,\ddagger}^i(x)$, where $v_{+,z,\dagger}^i$ is increasing and where $v_{-,z,\ddagger}^i$ is continuous and satisfies $v_{-,z,\ddagger}^i(z) \neq 0$.

Furthermore, suppose that $\{x_n(0)\}_{n\geq 1}$ *satisfies the large deviation principle on* E_1 *with good rate function* I_0*. Then,* $\{x_n\}_{n\geq 1}$ *satisfies the large deviation principle on* $D_{E_1}(\mathbb{R}^+)$ *with good rate function* I *given by*

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{otherwise} \end{cases}$$

where the Lagrangian $\mathcal{L}(x,v) : E_1 \times \mathbb{R}^d \to \mathbb{R}$ *is given by the Legendre transform* $\mathcal{L}(x,v) = \sup_{p\in\mathbb{R}^d} \langle p, v\rangle - H(x,p)$ *of the Hamiltonian* $H : E_1 \times \mathbb{R}^d \to \mathbb{R}$*, defined by*

$$H(x,p) = \sum_{i=1}^d v_+^i(x)\left[e^{2p_i} - 1\right] + v_-^i(x)\left[e^{-2p_i} - 1\right]. \tag{3.1.2}$$

Remark 3.1.2. Note that the functions v_+^i and v_-^i do not have to be of the form $v_+^i(x) = \frac{1-x_i}{2} r_+^i(x)$, $v_-^i(x) = \frac{1+x_i}{2} r_-^i(x)$ for some bounded functions r_+^i, r_-^i. This we call singular behaviour, as such a rate cannot be obtained the large deviation principle for independent particles via Varadhan's lemma and the contraction principle as in Léonard [1995] or Dai Pra and den Hollander [1996].

Theorem 3.1.3. *Suppose there exists a continuous function* $v : \{1,\ldots,d\} \times \{1,\ldots,d\} \times E_2 \to \mathbb{R}^+$ *such that for all* $a, b \in \{1,\ldots,d\}$*, we have*

$$\lim_{n\to\infty} \sup_{\mu\in E_{1,n}} |\mu(a) r_n(a,b,\mu) - v(a,b,\mu)| = 0. \tag{3.1.3}$$

Suppose that for each a, b*, the map* $\mu \mapsto v(a,b,\mu)$ *is either identically equal to zero or satisfies the following two properties.*

(a) $v(a,b,\mu) > 0$ *for all* μ *such that* $\mu(a) > 0$*.*

(b) For ν *such that* $\nu(a) = 0$*, there exists a neighbourhood* U_ν *of* ν *on which there exists a decomposition* $v(a,b,\mu) = v_{\nu,\dagger}(a,b,\mu(a)) v_{\nu,\ddagger}(a,b,\mu)$ *such that* $v_{\nu,\dagger}$ *is increasing in the third coordinate and such that* $v_{\nu,\ddagger}(a,b,\cdot)$ *is continuous and satisfies* $v_{\nu,\ddagger}(a,b,\nu) \neq 0$*.*

Additionally, suppose that $\{\mu_n(0)\}_{n\geq 1}$ *satisfies the large deviation principle on* E_2 *with good rate function* I_0*. Then,* $\{\mu_n\}_{n\geq 1}$ *satisfies the large deviation principle on* $D_{E_2}(\mathbb{R}^+)$ *with good rate function* I *given by*

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC} \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{L} : E_2 \times \mathbb{R}^d \to \mathbb{R}^+$ is the Legendre transform of $H : E_2 \times \mathbb{R}^d \to \mathbb{R}$ given by

$$H(\mu, p) = \sum_{a,b} v(a, b, \mu) \left[e^{p_b - p_a} - 1\right]. \tag{3.1.4}$$

3.1.3 *The comparison principle*

The main results in this chapter are the two large deviation principles as stated above. However, the main step in the proof of these principles is the verification of the comparison principle for a set of Hamilton-Jacobi equations. As this result is of independent interest, we state these results here as well, and leave explanation on why the comparison principle is relevant for the large deviation principles for later. We start with some definitions.

For E equals E_1 or E_2, let $H : E \times \mathbb{R}^d \to \mathbb{R}$ be some continuous map. For $\lambda > 0$ and $h \in C(E)$. Set $F_{\lambda,h} : E \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ by

$$F_{\lambda,h}(x, \alpha, p) = \alpha - \lambda H(x, p) - h(x).$$

We will solve the *Hamilton-Jacobi* equation

$$F_{\lambda,h}(x, f(x), \nabla f(x)) = f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0 \qquad x \in E, \tag{3.1.5}$$

in the *viscosity* sense.

Definition 3.1.4. We say that u is a *(viscosity) subsolution* of equation (3.1.5) if u is bounded, upper semi-continuous and if for every $f \in C^1(E)$ and $x \in E$ such that $u - f$ has a maximum at x, we have

$$F_{\lambda,h}(x, u(x), \nabla f(x)) \leq 0.$$

We say that u is a *(viscosity) supersolution* of equation (3.1.5) if u is bounded, lower semi-continuous and if for every $f \in C^1(E)$ and $x \in E$ such that $u - f$ has a minimum at x, we have

$$F_{\lambda,h}(x, u(x), \nabla f(x)) \geq 0.$$

We say that u is a *(viscosity) solution* of equation (3.1.5) if it is both a sub and a super solution.

There are various other definitions of viscosity solutions in the literature. This definition is the standard one for continuous H and compact state-space E.

Definition 3.1.5. We say that equation (3.1.5) satisfies the *comparison principle* if for a subsolution u and supersolution v we have $u \leq v$.

Note that if the comparison principle is satisfied, then a viscosity solution is unique.

Theorem 3.1.6. *Suppose that $H : E_1 \times \mathbb{R}^d \to \mathbb{R}$ is given by* (3.1.2) *and that the family of functions $v_+^i, v_-^i : E_1 \to \mathbb{R}^+$, $1 \leq i \leq d$, satisfy the conditions of Theorem 3.1.1.*

Then, for every $\lambda > 0$ and $h \in C(E_1)$, the comparison principle holds for $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$.

Theorem 3.1.7. *Suppose that $H : E_2 \times \mathbb{R}^d \to \mathbb{R}$ is given by* (3.1.4) *and that function $v : \{1, \ldots, d\} \times \{1, \ldots, d\} \times E_2 \to \mathbb{R}^+$ satisfies the conditions of Theorem 3.1.3.*

Then, for every $\lambda > 0$ and $h \in C(E_2)$, the comparison principle holds for $f(\mu) - \lambda H(\mu, \nabla f(\mu)) - h(\mu) = 0$.

The main consequence of the comparison principle for the Hamilton-Jacobi equations stems from the fact, as we will see below, that the operator H generates a strongly continuous contraction semigroup on $C(E)$.

The proof of the large deviation principle is, in a sense, a problem of semigroup convergence. At least for linear semigroups, it is well known that semigroup convergence can be proven via the convergence of their generators. The main issue in this approach is to prove that the limiting generator H generates a semigroup. It is exactly this issue that the comparison principle takes care of.

Hence, the independent interest of the comparison principle comes from the fact that we have semigroup convergence whatever the approximating semigroups are, as long as their generators converge to H, i.e. this holds not just for the specifically chosen approximating semigroups that we consider in Section 3.2.

3.1.4 *A Lyapunov function for the limiting dynamics*

As a corollary to the large deviation results, we show how to obtain a Lyapunov function for the solutions of

$$\dot{x}(t) = \mathbf{F}(x(t)), \tag{3.1.6}$$

where $\mathbf{F}(x) := H_p(x,0)$ for a Hamiltonian as in (3.1.4) or (3.1.2). Here $H_p(x,p)$ is interpreted as the vector of partial derivatives of H in the second coordinate.

We will see in Example 3.1.11 that the trajectories that solve this differential equation are the trajectories with 0 Lagrangian cost: $\dot{x} = \mathbf{F}(x)$ if and only if $\mathcal{L}(x,\dot{x}) = 0$. Additionally, the limiting operator $(A, C^1(E))$ obtained by

$$\sup_{x \in E_n \cap K} |A_n f(x) - Af(x)| \to 0$$

for all $f \in C^1(E)$ and compact sets $K \subseteq E$ has the form by $Af(x) = \langle \nabla f(x), \mathbf{F}(x) \rangle$ for the same vector field $\mathbf{F}$. This implies that the 0-cost trajectories are solutions to the McKean-Vlasov equation (3.1.6). Solutions to 3.1.6 are not necessarily unique, see Example 3.1.11. Uniqueness holds for example under a one-sided Lipschitz condition: if there exists $M > 0$ such that $\langle \mathbf{F}(x) - \mathbf{F}(y), x - y \rangle \leq M|x-y|^2$ for all $x, y \in E$.

For non-interacting systems, it is well known that the relative entropy with respect to the stationary measure is a Lyapunov function for solutions of (3.1.6). The large deviation principle explains this fact and gives a method to obtain a suitable Lyapunov function, also for interacting dynamics.

Proposition 3.1.8. *Suppose the conditions for Theorem 3.1.1 or Theorem 3.1.3 are satisfied. Suppose there exists measures $\nu_n \in \mathcal{P}(E_n) \subseteq \mathcal{P}(E)$ that are invariant for the dynamics generated by A_n. Furthermore, suppose that the measures ν_n satisfy the large deviation principle on E with good rate function S.*

Then S is increasing along any solution of $\dot{x}(t) = \mathbf{F}(x(t))$.

Note that we do not assume that (3.1.6) has a unique solution for a given starting point.

3.1.5 *Examples*

We give a series of examples to show the extent of Theorems 3.1.1 and 3.1.3.

For the Ehrenfest model, we start with the basic case, of spins flipping under the influence of some mean-field potential.

Example 3.1.9. Fix some continuously differentiable $V : [-1,1]^d \to \mathbb{R}$ and set for every $n \geq 1$ and $i \in \{1,\ldots,d\}$ the rates

$$r^+_{n,i}(x) = \exp\left\{-n2^{-1}\left(V\left(x + \frac{2}{n}e_i\right) - V(x)\right)\right\},$$
$$r^-_{n,i}(x) = \exp\left\{-n2^{-1}\left(V\left(x - \frac{2}{n}e_i\right) - V(x)\right)\right\}.$$

The limiting objects v^i_+ and v^i_- are given by

$$v^i_+(x) = \frac{1-x_i}{2}e^{-\nabla_i V(x)}, \qquad v^i_-(x) = \frac{1+x_i}{2}e^{\nabla_i V(x)},$$

which already have the decomposition as required in the conditions of the Theorem 3.1.1. For example, condition (b) for v^i_+ is satisfied by

$$v^i_{+,z,\dagger}(x_i) := \frac{1-x_i}{2}, \qquad v^i_{+,z,\ddagger}(x) := e^{-\nabla_i V(x)}.$$

For $d = 1$, we give two extra notable examples, the first one exhibits unbounded jump rates for the individual spins if the empirical magnetisation is close to one of the boundary points. The second example shows a case where we have multiple trajectories γ with $I(\gamma) = 0$ that start from $x_0 = 0$. Because $d = 1$, we drop all sub- and super-scripts $i \in \{1,\ldots,d\}$ for the these two examples.

Example 3.1.10. Consider the one-dimensional Ehrenfest model with

$$r_{n,+}(x) = \frac{2}{\sqrt{1-x}} \wedge n, \qquad r_{n,-}(x) = \frac{2}{\sqrt{1+x}} \wedge n.$$

Set $v_+(x) = \sqrt{1-x}$, $v_-(x) = \sqrt{1+x}$. By Dini's theorem, we have

$$\sup_{x\in[-1,1]}\left|\frac{1-x}{2}r_{n,+(x)} - v_+(x)\right| = 0,$$
$$\sup_{x\in[-1,1]}\left|\frac{1+x}{2}r_{n,-(x)} - v_-(x)\right| = 0.$$

And additionally, conditions (a) and (b) of Theorem 3.1.1 are satisfied, e.g. take $v_{+,1,\dagger}(x) = \sqrt{1-x}$, $v_{+,1,\ddagger}(x) = 1$.

Example 3.1.11. Consider the one-dimensional Ehrenfest model with some rates $r_{n,+}$, $r_{n,-}$ and functions $v_+(x) > 0, v_-(x) > 0$ such that $\frac{1}{2}(1-x)r_{n,+}(x) \to v_+(x)$ and $\frac{1}{2}(1+x)r_{n,-}(x) \to v_-(x)$ uniformly in $x \in [-1,1]$.

Now suppose that there is a neighbourhood U of 0 on which v_+, v_- have the form

$$v_+(x) = \begin{cases} 1+\sqrt{x} & x \geq 0, \\ 1 & x < 0, \end{cases} \qquad v_-(x) = 1.$$

Consider the family of trajectories $t \mapsto \gamma_a(t)$, $a \geq 0$, defined by

$$\gamma_a(t) := \begin{cases} 0 & \text{for } t \leq a, \\ (t-a)^2 & \text{for } t \geq a. \end{cases}$$

Let $T > 0$ be small enough such that $\gamma_0(t) \in U$, and hence $\gamma_a(t) \in U$, for all $t \leq T$. A straightforward calculation yields $\int_0^T \mathcal{L}(\gamma_a(t), \dot{\gamma}_a(t))\mathrm{d}t = 0$ for all $a \geq 0$. So we find multiple trajectories starting at 0 that have zero Lagrangian cost.

Indeed, note that $\mathcal{L}(x,v) = 0$ is equivalent to $v = H_p(x,0) = 2\left[v_+(x) - v_-(x)\right] = 2\sqrt{(x)}$. This yields that trajectories that have 0 Lagrangian cost are the trajectories, at least in U, that solve

$$\dot{\gamma}(t) = 2\sqrt{\gamma(t)}$$

which is the well-known example of a differential equation that allows for multiple solutions.

We end with an example for Theorem 3.1.3 and Proposition 3.1.8 in the spirit of Example 3.1.9.

Example 3.1.12 (Glauber dynamics for the Potts-model)**.** Fix some continuously differentiable function $V : \mathbb{R}^d \to \mathbb{R}$. Define the Gibbs measures

$$\nu_n(\mathrm{d}\sigma) := \frac{e^{-V(\mu_n(\sigma))}}{Z_n} P^{\otimes,n}(\mathrm{d}\sigma)$$

on $\{1,\ldots,d\}^n$, where $P^{\otimes,n}$ is the n-fold product measure of the uniform measure P on $\{1,\ldots,d\}$ and where Z_n are normalizing constants.

Let $S_0(\mu \mid P)$ denote the relative entropy of $\mu \in \mathcal{P}(\{1,\ldots,d\})$ with respect to P:

$$S_0(\mu \mid P) = \sum_a \log(d\mu(a))\mu(a).$$

By Sanov's theorem and Varadhan's lemma, the empirical measures under the laws ν_n satisfy a large deviation principle with rate function $S(\mu) = S_0(\mu \mid P) + V(\mu)$.

Fix some function $r : \{1,\ldots,d\} \times \{1,\ldots,d\} \to \mathbb{R}^+$. Set

$$r_n(a,b,\mu) = r(a,b)\exp\left\{-n2^{-1}\left(V\left(\mu - n^{-1}\delta_a + n^{-1}\delta_b\right) - V(\mu)\right)\right\}.$$

As n goes to infinity, we have uniform convergence of $\mu(a)r_n(a,b,\mu)$ to

$$v(a,b,\mu) := \mu(a)r(a,b)\exp\left\{\frac{1}{2}\nabla_a V(\mu) - \frac{1}{2}\nabla_b V(\mu)\right\},$$

where $\nabla_a V(\mu)$ is the derivative of V in the a-th coordinate. As in Example 3.1.9, condition (b) of Theorem 3.1.3 is satisfied by using the obvious decomposition.

By Proposition 3.1.8, we obtain that $S(\mu) = S_0(\mu \mid P) + V(\mu)$ is Lyapunov function for

$$\dot{\mu}(a) = \sum_b [v(b,a,\mu) - v(a,b,\mu)] \qquad a \in \{1,\ldots,d\}.$$

3.1.6 *Discussion and comparison to the existing literature*

We discuss our results in the context of the existing literature that cover our situation. Additionally, we consider a few cases where the large deviation principle(LDP) is proven for diffusion processes, because the proof techniques could possibly be applied in this setting.

LDP: Approach via non-interacting systems, Varadhan's lemma and the contraction principle. In Léonard [1995], Dai Pra and den Hollander [1996], Borkar and Sundaresan [2012], the first step towards the LDP of the trajectory of some mean-field statistic of n interacting particles is the LDP for non-interacting particles on some large product space obtained via Sanov's theorem. Varadhan's lemma then gives the LDP in this product space for interacting particles, after which the contraction principle gives the LDP on the desired trajectory space. In Léonard [1995], Dai Pra and den Hollander [1996], the set-up is more general compared to ours in the sense that in Léonard [1995] the behaviour of the particles depends on their spatial location, and in Dai Pra and den Hollander [1996] the behaviour of a particle depends on some external random variable.

On the other hand, systems as in Example 3.1.10 fall outside of the conditions imposed in the three papers, if we disregard spatial dependence or external randomness.

The approach via Varadhan's lemma, which needs control over the size of the perturbation, does not work, at least naively, for the situation where the jump rate for individual particles is diverging to ∞, or converging to 0, if the mean is close to the boundary, see Remark 3.1.2.

LDP: Explicit control on the probabilities. For another approach considering interacting spins that have a spatial location, see Comets [1987]. The jump rates are taken to be explicit and the large deviation principle is proven via explicit control on the Radon-Nikodym derivatives. This method should in principle work also in the case of singular v. The approach via the generators H_n in this chapter, avoids arguments based on explicit control. This is an advantage for processes where the functions r_n and v are not very regular. Also in the classical Freidlin-Wentzell approach Freidlin and Wentzell [1998] for dynamical systems with Gaussian noise the explicit form of the Radon-Nikodym derivatives is used to prove the LDP.

LDP: Direct comparison to a process of independent particles. The main reference concerning large deviations for the trajectory of the empirical mean for interacting diffusion processes on $\mathbb{R}^d$ is Dawson and Gärtner [1987]. In this chapter, the large deviation principle is also first established for non-interacting particles. An explicit rate function is obtained by showing that the desired rate is in between the rate function obtained via Sanov's theorem and the contraction principle and the projective limit approach. The large deviation principle for interacting particles is then obtained via comparing the interacting process with a non-interacting process that has a suitably chosen drift. For related approaches, see Feng [1994] for large deviations of interacting jump processes on $\mathbb{N}$, where the interaction is unbounded and depends on the average location of the particles. See Boualem and Ingemar [1995] for mean-field jump processes on $\mathbb{R}^d$.

Again, the comparison with non-interacting processes would fail in our setting due the singular interaction terms.

LDP: Stochastic control. A more recent approach using stochastic control and weak convergence methods has proposed in the context of both jump and diffusion processes in Budhiraja et al. [2011, 2012]. A direct application of the results in Budhiraja et al. [2011] fails for jump processes in the setting of singular behaviour at the boundary.

LDP: Proof via operator convergence and the comparison principle. Regarding our approach based on the comparison principle, see [Feng and Kurtz, 2006, Section 13.3], for an approach based on the comparison principle in the setting of Dawson and Gärtner [1987] and Budhiraja et al. [2012]. See Deng et al. [2011] for an example of large deviations of a diffusion processes on $(0, \infty)$ with vanishing diffusion term with singular behaviour at the boundary. The methods to prove the comparison principle in Sections 9.2 and 9.3 in Feng and Kurtz [2006] do not apply in our setting due to the different nature of our Hamiltonians.

LDP: Comparison of the approaches The method of obtaining exponential tightness in Feng and Kurtz [2006], and thus employed for this chapter, is via density of the domain of the limiting generator $(H, \mathcal{D}(H))$. Like in the theory of weak convergence, functions $f \in \mathcal{D}(H)$ in the domain of the generator, and functions $f_n \in \mathcal{D}(H_n)$ that converge to f uniformly, can be used to bound the fluctuations in the Skorokhod space. This method is similar to the approaches taken in Comets [1989], Freidlin and Wentzell [1998], Dawson and Gärtner [1987].

The approach using operator convergence is based on a variant of the projective limit theorem for the Skorokhod space proven in Feng and Kurtz [2006] by direct calculations. Because we have exponential tightness on the Skorokhod space, it suffices to prove the large deviation principle for all finite dimensional distributions. This is done via the convergence of the logarithmic moment generating functions for the finite dimensional distributions. The Markov property reduces this to the convergence of the logarithmic moment generating function for time 0 and convergence of the conditional moment generating functions, that form a semigroup $V_n(t)f(x) = \frac{1}{n} \log \mathbb{E}[e^{nf(X_n(t))} \mid X_n(0) = x]$. Thus, the problem is reduced to proving convergence of semigroups $V_n(t)f \to V(t)f$. As in the theory of linear semigroups, this comes down to two steps. First one proves convergence of the generators $H_n \to H$. Then one shows that the limiting semigroup generates a semigroup. The verification of the comparison principle implies that the domain of the limiting operator is sufficiently large to pin down a limiting semigroup.

This can be compared to the same problem for linear semigroups and the martingale problem. If the domain of a limiting linear generator is too small, multiple solutions to the martingale problem can be found, giving rise to multiple semigroups, see Chapter 12 in Stroock and Varadhan [1979] or Section 4.5 in Ethier and Kurtz [1986].

The convergence of $V_n(t)f(x) \to V(t)f(x)$ uniformly in x corresponds to having sufficient control on the Doob-h transforms corresponding to the change of measures

$$\frac{\mathrm{d}\mathbb{P}_{n,x}^{f,t}}{\mathrm{d}\mathbb{P}_{n,x}}(X_n) = \exp\left\{nf(X_n(t))\right\},$$

where $\mathbb{P}_{n,x}$ is the measure corresponding to the process X_n started in x at time 0. An argument based on the projective limit theorem and control on the Doob h-transforms for independent particles is also used in Dawson and Gärtner [1987], whereas the methods in Comets [1989], Freidlin and Wentzell [1998] are based on direct calculation of the probabilities being close to a target trajectories.

Large deviations for large excursions in large time. A notable second area of comparison is the study of large excursions in large time in the context of queuing systems, see e.g. Dupuis et al. [1990], Dupuis and Ellis [1995], Atar and Dupuis [1999] and references therein. Here, it is shown that the rate functions themselves, varying in space and time, are solutions to a Hamilton-Jacobi equation. As in our setting, one of the main problems is the verification of the comparison principle. The notable difficulty in these papers is a discontinuity of the Hamiltonian at the boundary, but in their interior the rates are uniformly bounded away from infinity and zero.

Lyapunov functions. In Budhiraja et al. [2015a,b], Lyapunov functions are obtained for the McKean-Vlasov equation corresponding to interacting Markov processes in a setting similar to the setting of Theorem 3.1.3. Their discussion goes much beyond Proposition 3.1.8, which is perhaps best compared to Theorem 4.3 in Budhiraja et al. [2015b]. However, the proof of Proposition 3.1.8 is interesting in its own right, as it gives an intuitive explanation for finding a relative entropy as a Lyapunov functional and is not based on explicit calculations. In particular, the proof of Proposition 3.1.8 in principle works for any setting where the path-space large deviation principle holds.

3.2 LARGE DEVIATION PRINCIPLE VIA AN ASSOCIATED HAMILTON-JACOBI EQUATION

In this section, we will summarize the main results of Feng and Kurtz [2006]. Additionally, we will verify the main conditions of their results, except for the comparison principle of an associated Hamilton-Jacobi equation. This verification needs to be done for each individual model separately and this is the main contribution of this chapter. We verify the comparison principle for our two models in Section 3.3.

3.2.1 *Operator convergence*

Let E_n and E denote either of the spaces $E_{n,1}, E_1$ or $E_{n,2}, E_2$. Furthermore, denote by $C(E)$ the continuous functions on E and by $C^1(E)$ the functions that are continuously differentiable on a neighbourhood of E in $\mathbb{R}^d$.

Assume that for each $n \in \mathbb{N}$, we have a jump process X_n on E_n, generated by a bounded infinitesimal generator A_n. For the two examples, this process is either x_n or μ_n. We denote by $\{S_n(t)\}_{t\geq 0}$ the transition semigroups $S_n(t)f(y) = \mathbb{E}\left[f(X_n(t)) \,|\, X_n(0) = y\right]$ on $C(E_n)$. Define for each n the exponential semigroup

$$V_n(t)f(y) := \frac{1}{n} \log S_n(t)e^{nf}(y) = \frac{1}{n} \log \mathbb{E}\left[e^{nf(X_n(t))} \,\middle|\, X_n(0) = y\right].$$

As in the theory of weak convergence, given that the processes X_n satisfy a exponential compact containment condition on the Skorokhod space, which in this setting is immediate, Feng and Kurtz [2006] show that the existence of a strongly continuous limiting semigroup $\{V(t)\}_{t\geq 0}$ on $C(E)$ in the sense that for all $f \in C(E)$ and $T \geq 0$, we have

$$\lim_{n\to\infty} \sup_{t\leq T} \sup_{x\in E_n} |V(t)f(x) - V_n(t)f(x)| = 0, \tag{3.2.1}$$

allows us to study the large deviation behaviour of the process X_n. We will consider this question from the point of view of the generators H_n of

$\{V_n(t)\}_{t\geq 0}$, where $H_n f$ is defined by the norm limit of $t^{-1}(V_n(t)f - f)$ as $t \downarrow 0$. Note that $H_n f = n^{-1}e^{-nf}A_n e^{nf}$, which for our first model yields

$$\begin{aligned} &H_n f(x) \\ &= \sum_{i=1}^{d} \left\{ \frac{1-x_i}{2} r^i_{n,+}(x) \left[\exp\left\{ n \left(f\left(x + \frac{2}{n} e_i \right) - f(x) \right) \right\} - 1 \right] \right. \\ &\qquad \left. + \frac{1+x_i}{2} r^i_{n,-}(x) \left[\exp\left\{ n \left(f\left(x - \frac{2}{n} e_i \right) - f(x) \right) \right\} - 1 \right] \right\}. \end{aligned}$$

For our second model, we have

$$H_n f(\mu) = \sum_{a,b=1}^{d} \mu(a) r_n(a,b,\mu) \left[e^{n\left(f\left(\mu - n^{-1}\delta_a + n^{-1}\delta_b\right) - f(\mu)\right)} - 1 \right].$$

In particular, Feng and Kurtz show that, as in the theory of weak convergence of Markov processes, the existence of a limiting operator $(H, \mathcal{D}(H))$, such that for all $f \in \mathcal{D}(H)$

$$\lim_{n\to\infty} \sup_{x\in E_n} |Hf(x) - H_n f(x)| = 0, \tag{3.2.2}$$

for which one can show that $(H, \mathcal{D}(H))$ generates a semigroup $\{V(t)\}_{t\geq 0}$ on $C(E)$ via the Crandall-Liggett theorem, Crandall and Liggett [1971], then (3.2.1) holds.

Lemma 3.2.1. *For either of our two models, assuming* (3.1.1) *or* (3.1.3)*, we find that $H_n f \to Hf$, as in* (3.2.2) *holds for $f \in C^1(E)$, where Hf is given by $Hf(x) := H(x, \nabla f(x))$ and where $H(x,p)$ is defined in* (3.1.2) *or* (3.1.4).

The proof of the lemma is straightforward using the assumptions and the fact that f is continuously differentiable.

Thus, the problem is reduced to proving that $(H, C^1(E))$ generates a semigroup. The verification of the conditions of the Crandall-Liggett theorem is in general very hard, or even impossible. Two conditions need to be verified, the first is the *dissipativity* of H, which can be checked via the positive maximum principle. The second condition is the *range condition*: one needs to show that for $\lambda > 0$, the range of $(\mathbb{1} - \lambda H)$ is dense in $C(E)$. In other words, for $\lambda > 0$ and sufficiently many fixed $h \in C(E)$, we need to solve $f - \lambda Hf = h$ with $f \in C^1(E)$. An alternative is to solve this equation in the *viscosity sense*. If a viscosity solution exists and is unique, we denote it by $\tilde{R}(\lambda)h$. Using these solutions, we can extend the domain of the operator $(H, C^1(E))$ by adding all pairs of the form $(\tilde{R}(\lambda)h, \lambda^{-1}(\tilde{R}(\lambda)h - h))$ to the graph of H to obtain an operator $\hat{H}$ that satisfies the conditions for the Crandall-Liggett theorem. This is part of the content of Theorem 3.2.2 stated below.

As a remark, note that any concept of weak solutions could be used to extend the operator. However, viscosity solutions are special in the sense that the extended operator remains dissipative.

The next result is a direct corollary of Theorem 6.14 in Feng and Kurtz [2006].

Theorem 3.2.2. *For either of our two models, assume that* (3.1.1) *or* (3.1.3) *holds. Additionally, assume that the comparison principle is satisfied for* (3.1.5) *for all* $\lambda > 0$ *and* $h \in C(E)$.

Then, the operator

$$\hat{H} := \bigcup_{\lambda > 0} \left\{ \left(\tilde{R}(\lambda)h, \lambda^{-1}(\tilde{R}(\lambda)h - h) \right) \,\middle|\, h \in C(E) \right\}$$

generates a semigroup $\{V(t)\}_{t \geq 0}$ *as in the Crandall-Liggett theorem and we have* (3.2.1).

Additionally, suppose that $\{X_n(0)\}$ *satisfies the large deviation principle on* E *with good rate function* I_0. *Then* X_n *satisfies the large deviation principle on* $D_E(\mathbb{R}^+)$ *with good rate function* I *given by*

$$I(\gamma) = I_0(\gamma(0)) + \sup_m \sup_{0 = t_0 < t_1 < \cdots < t_m} \sum_{k=1}^{m} I_{t_k - t_{k-1}}(\gamma(t_k) \,|\, \gamma(t_{k-1})),$$

where $I_s(y \,|\, x) := \sup_{f \in C(E)} f(y) - V(s)f(x)$.

Note that to prove Theorem 6.14 in Feng and Kurtz [2006], one needs to check that viscosity sub- and super-solutions to (3.1.5) exist. Feng and Kurtz construct these sub- and super-solutions explicitly, using the approximating operators H_n, see the proof of Lemma 6.9 in Feng and Kurtz [2006].

Proof. We check the conditions for Theorem 6.14 in Feng and Kurtz [2006]. In our models, the maps $\eta_n : E_n \to E$ are simply the embedding maps. Condition (a) is satisfied as all our generators A_n are bounded. The conditions for convergence of the generators follow by Lemma 3.2.1. □

The additional assumptions in Theorems 3.1.1 and 3.1.3 are there to make sure we are able to verify the comparison principle. This is the major contribution of the chapter and will be carried out in Section 3.3.

The final steps to obtain Theorems 3.1.1 and 3.1.3 are to obtain the rate function as the integral over a Lagrangian. Also this is based on results in Chapter 8 of Feng and Kurtz [2006].

3.2.2 *Variational semigroups*

In this section, we introduce the *Nisio semigroup* $\mathbf{V}(t)$, of which we will show that it equals $V(t)$ on $C(E)$. This semigroup is given as a variational problem where one optimises a pay-off $f(\gamma(t))$ that depends on the state $\gamma(t) \in E$, but where a cost is paid that depends on the whole trajectory $\{\gamma(s)\}_{0\leq s\leq t}$. The cost is accumulated over time and is given by a 'Lagrangian'. Given the continuous and convex operator $Hf(x) = H(x, \nabla f(x))$, we define this Lagrangian by taking the Legendre-Fenchel transform:

$$\mathcal{L}(x,u) := \sup_{p\in\mathbb{R}^d} \left\{\langle p, u\rangle - H(x,p)\right\}.$$

As $p \mapsto H(x,p)$ is convex and continuous, it follows by the Fenchel - Moreau theorem that also

$$Hf(x) = H(x,\nabla f(x)) = \sup_{u\in\mathbb{R}^d} \left\{\langle \nabla f(x), u\rangle - \mathcal{L}(x,u)\right\}.$$

Using $\mathcal{L}$, we define the Nisio semigroup for measurable functions f on E:

$$\mathbf{V}(t)f(x) = \sup_{\substack{\gamma\in\mathcal{AC}\\ \gamma(0)=x}} f(\gamma(t)) - \int_0^t \mathcal{L}(\gamma(s),\dot\gamma(s))\mathrm{d}s. \tag{3.2.3}$$

To be able to apply the results from Chapter 8 in Feng and Kurtz [2006], we need to verify Conditions 8.9 and 8.11 of Feng and Kurtz [2006].

For the semigroup to be well behaved, we need to verify Condition 8.9 in Feng and Kurtz [2006]. In particular, this condition implies Proposition 8.13 in Feng and Kurtz [2006] that ensures that the Nisio semigroup is in fact a semigroup on the upper semi-continuous functions that are bounded above. Additionally, it implies that all absolutely continuous trajectories up to time T, that have uniformly bounded Lagrangian cost, are a compact set in $D_E([0,T])$.

Lemma 3.2.3. *For the Hamiltonians in* (3.1.2) *and* (3.1.4)*, Condition 8.9 in Feng and Kurtz [2006] is satisfied.*

Proof. For (1),take $U = \mathbb{R}^d$ and set $Af(x,v) = \langle \nabla f(x), v\rangle$. Considering Definition 8.1 in Feng and Kurtz [2006], if $\gamma \in \mathcal{AC}$, then

$$f(\gamma(t)) - f(\gamma(0)) = \int_0^t Af(\gamma(s), \dot{\gamma}(s))\mathrm{d}s$$

by definition of A. In Definition 8.1, however, relaxed controls are considered, i.e. instead of a fixed speed $\dot{\gamma}(s)$, one considers a measure $\lambda \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^+)$, such that $\lambda(\mathbb{R}^d \times [0,t]) = t$ for all $t \geq 0$ and

$$f(\gamma(t)) - f(\gamma(0)) = \int_0^t Af(\gamma(s), v)\lambda(\mathrm{d}v, \mathrm{d}s).$$

These relaxed controls are then used to define the Nisio semigroup in equation (8.10). Note however, that by convexity of H in the second coordinate, also $\mathcal{L}$ is convex in the second coordinate. It follows that a deterministic control $\lambda(\mathrm{d}v, \mathrm{d}t) = \delta_{v(t)}(\mathrm{d}v)\mathrm{d}t$ is always the control with the smallest cost by Jensen's inequality. We conclude that we can restrict the definition (8.10) to curves in $\mathcal{AC}$. This motivates our changed definition in equation (3.2.3).

For this chapter, it suffices to set $\Gamma = E \times \mathbb{R}^d$, so that (2) is satisfied. By compactness of E, (4) is clear.

We are left to prove (3) and (5). For (3), note that $\mathcal{L}$ is lower semi-continuous by construction. We also have to prove compactness of the level sets. By lower semi-continuity, it is sufficient to show that the level sets $\{\mathcal{L} \leq c\}$ are contained in a compact set.

Set $\mathcal{N} := \cap_{x \in E} \left\{ p \in \mathbb{R}^d \,\middle|\, H(x,p) \leq 1 \right\}$. First, we show that $\mathcal{N}$ has non-empty interior, i.e. there is some $\varepsilon > 0$ such that the open ball $B(0,\varepsilon)$ of radius ε around 0 is contained in $\mathcal{N}$. Suppose not, then there exists x_n and p_n such that $p_n \to 0$ and for all n: $H(x_n, p_n) = 1$. By compactness of E and continuity of H, we find a value $H(x,0) = 1$, which contradicts our definitions of H, where $H(y,0) = 0$ for all $y \in E$.

Let $(x,v) \in \{\mathcal{L} \leq c\}$, then

$$\langle p, v\rangle \leq \mathcal{L}(x,v) + H(x,p) \leq c + 1$$

for all $p \in B(0,\varepsilon) \subseteq \mathcal{N}$. It follows that v is contained in some bounded ball in $\mathbb{R}^d$. It follows that $\{\mathcal{L} \leq c\}$ is contained in some compact set by the Heine-Borel theorem.

Finally, (5) can be proven as Lemma 10.21 in Feng and Kurtz [2006] or as in Lemma 6.4.19. □

The last property necessary for the equality of $V(t)f$ and $\mathbf{V}(t)f$ on $C(E)$ is the verification of Condition 8.11 in Feng and Kurtz [2006]. This condition is key to proving that a variational resolvent, see equation (8.22), is a viscosity super-solution to (3.1.5). As the variational resolvent is also a sub-solution to (3.1.5) by Young's inequality, the variational resolvent is a viscosity solution to this equation. If viscosity solutions are unique, this yields, after an approximation argument that $V(t) = \mathbf{V}(t)$.

Lemma 3.2.4. *Condition 8.11 in Feng and Kurtz [2006] is satisfied. In other words, for all $g \in C^1(E)$ and $x_0 \in E$, there exists a trajectory $\gamma \in \mathcal{AC}$ such that $\gamma(0) = x_0$ and for all $T \geq 0$:*

$$\int_0^T Hg(\gamma(t))\mathrm{d}t = \int_0^T \langle \nabla g(\gamma(t)), \dot\gamma(t)\rangle - \mathcal{L}(\gamma(t), \dot\gamma(t))\mathrm{d}t. \quad (3.2.4)$$

Proof. Fix $T > 0$, $g \in C^1(E)$ and $x_0 \in E$. We introduce a vector field $\mathbf{F}^g : E \to \mathbb{R}^d$, by

$$\mathbf{F}^g(x) := H_p(x, \nabla g(x)),$$

where $H_p(x,p)$ is the vector of partial derivatives of H in the second coordinate. Note that in our examples, H is continuously differentiable in the p-coordinates. For example, for the $d = 1$ case of Theorem 3.1.1, we obtain

$$\mathbf{F}^g(x) := 2v_+(x)e^{2\nabla g(x)} - 2v_-(x)e^{-2\nabla g(x)}.$$

As $\mathbf{F}^g$ is a continuous vector field, we can find a local solution $\gamma^g(t)$ in E to the differential equation

$$\begin{cases} \dot\gamma(t) = \mathbf{F}^g(\gamma(t)), \\ \gamma(0) = x_0, \end{cases}$$

by an extended version of Peano's theorem Crandall [1972]. The result in Crandall [1972] is local, however, the length of the interval on which the solution is constructed depends inversely on the norm of the vector field, see his equation (2). As our vector fields are globally bounded in size, we can iterate the construction in Crandall [1972] to obtain a global existence result, such that $\dot\gamma^g(t) = \mathbf{F}^g(\gamma(t))$ for almost all times in $[0,\infty)$.

We conclude that on a subset of full measure of $[0,T]$ that

$$\begin{aligned}\mathcal{L}(\gamma^g(t),\dot{\gamma}^g(t)) &= \mathcal{L}(\gamma^g(t),\mathbf{F}^g(\gamma^g(t)))\\ &= \sup_{p\in\mathbb{R}^d}\langle p,\mathbf{F}^g(\gamma^g(t))\rangle - H(\gamma^g(t),p)\\ &= \sup_{p\in\mathbb{R}^d}\langle p,H_p(\gamma^g(t),\nabla g(\gamma^g(t)))\rangle - H(\gamma^g(t),p).\end{aligned}$$

By differentiating the final expression with respect to p, we find that the supremum is taken for $p=\nabla g(\gamma^g(t))$. In other words, we find

$$\begin{aligned}&\mathcal{L}(\gamma^g(t),\dot{\gamma}^g(t))\\ &\quad= \langle\nabla g(\gamma^g(t)),H_p(\gamma^g(t),\nabla g(\gamma^g(t)))\rangle - H(\gamma^g(t),\nabla g(\gamma^g(t)))\\ &\quad= \langle\nabla g(\gamma^g(t)),\dot{\gamma}^g(t)\rangle - Hg(\gamma^g(t)).\end{aligned}$$

By integrating over time, the zero set does not contribute to the integral, we find (3.2.4). □

The following result follows from Corollary 8.29 in Feng and Kurtz [2006].

Theorem 3.2.5. *For either of our two models, assume that* (3.1.1) *or* (3.1.3) *holds. Assume that the comparison principle is satisfied for* (3.1.5) *for all* $\lambda > 0$ *and* $h\in C(E)$. *Finally, suppose that* $\{X_n(0)\}$ *satisfies the large deviation principle on* E *with good rate function* I_0.

Then, we have $V(t)f = \mathbf{V}(t)f$ *for all* $f\in C(E)$ *and* $t\geq 0$. *Also,* X_n *satisfies the large deviation principle on* $D_E(\mathbb{R}^+)$ *with good rate function* I *given by*

$$I(\gamma) := \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s),\dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma\in\mathcal{AC},\\ \infty & \text{if } \gamma\notin\mathcal{AC}.\end{cases}$$

Proof. We check the conditions for Corollary 8.29 in Feng and Kurtz [2006]. Note that in our setting $H=\mathbf{H}$. Therefore, condition (a) of Corollary 8.29 is trivially satisfied. Furthermore, we have to check the conditions for Theorems 6.14 and 8.27. For the first theorem, these conditions were checked already in the proof of our Theorem 3.2.2. For Theorem 8.27, we need to check Conditions 8.9, 8.10 and 8.11 in Feng and Kurtz [2006]. As $H1=0$, Condition 8.10 follows from 8.11. 8.9 and 8.11 have been verified in Lemmas 3.2.3 and 3.2.4. □

The last theorem shows us that we have Theorems 3.1.1 and 3.1.3 if we can verify the comparison principle, i.e. Theorems 3.1.6 and 3.1.7. This will be done in the section below.

Proof of Theorems 3.1.1 and 3.1.3. The comparison principles for equation (3.1.5) are verified in Theorems 3.1.6 and 3.1.7. The two theorems now follow from Theorem 3.2.5. □

Proof of Proposition 3.1.8. We give the proof for the system considered in Theorem 3.1.1. Fix $t \geq 0$ and some starting point x_0. Let $x(t)$ be any solution of $\dot{x}(t) = \mathbf{F}(x(t))$ with $x(0) = x_0$. We show that $S(x(t)) \leq S(x_0)$.

Let $X_n(0)$ be distributed as ν_n. Then it follows by Theorem 3.1.1 that the large deviation principle holds for $\{X_n\}_{n\geq 0}$ on $D_E(\mathbb{R}^+)$.

As ν_n is invariant for the Markov process generated by A_n, also the sequence $\{X_n(t)\}_{n\geq 0}$ satisfies the large deviation principle on E with good rate function S. Combining these two facts, the Contraction principle[Dembo and Zeitouni, 1998, Theorem 4.2.1] yields

$$\begin{aligned} S(x(t)) = \inf_{\gamma \in \mathcal{AC}: \gamma(t) = x(t)} S(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s \\ \leq S(x(0)) + \int_0^t \mathcal{L}(x(s), \dot{x}(s)) \mathrm{d}s = S(x(0)). \end{aligned}$$

Note that $\mathcal{L}(x(s), \dot{x}(s)) = 0$ for all s as was shown in Example 3.1.11. □

3.3 THE COMPARISON PRINCIPLE

We proceed with checking the comparison principle for equations of the type $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$. In other words, for subsolutions u and supersolutions v we need to check that $u \leq v$. We start with some known results. First of all, we give the main tool to construct sequences x_α and y_α that converge to a maximising point $z \in E$ such that $u(z) - v(z) = \sup_{z' \in E} u(z') - v(z')$. This result can be found for example as Proposition 3.7 in Crandall et al. [1992].

Lemma 3.3.1. *Let E be a compact subset of $\mathbb{R}^d$, let u be upper semi-continuous, v lower semi-continuous and let $\Psi : E^2 \to \mathbb{R}^+$ be a lower semi-continuous function such that $\Psi(x, y) = 0$ if and only if $x = y$. For $\alpha > 0$, let $x_\alpha, y_\alpha \in E$ such that*

$$u(x_\alpha) - v(y_\alpha) - \alpha \Psi(x_\alpha, y_\alpha) = \sup_{x,y \in E} \left\{ u(x) - v(y) - \alpha \Psi(x, y) \right\}.$$

Then the following hold

(i) $\lim_{\alpha \to \infty} \alpha \Psi(x_\alpha, y_\alpha) = 0$.

(ii) All limit points of (x_α, y_α) *are of the form* (z, z) *and for these limit points we have* $u(z) - v(z) = \sup_{x \in E} \{u(x) - v(x)\}$.

We say that $\Psi : E^2 \to \mathbb{R}^+$ is a *good penalization function* if $\Psi(x, y) = 0$ if and only if $x = y$, it is continuously differentiable in both components and if $(\nabla \Psi(\cdot, y))(x) = -(\nabla \Psi(x, \cdot))(y)$ for all $x, y \in E$. The next two results can be found as Lemma 9.3 in Feng and Kurtz [2006]. We will give the proofs of these results for completeness.

Proposition 3.3.2. *Let* $(H, \mathcal{D}(H))$ *be an operator such that* $\mathcal{D}(H) = C^1(E)$ *of the form* $Hf(x) = H(x, \nabla f(x))$. *Let* u *be a subsolution and* v *a supersolution to* $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$, *for some* $\lambda > 0$ *and* $h \in C(E)$. *Let* Ψ *be a good penalization function and let* x_α, y_α *satisfy*

$$u(x_\alpha) - v(y_\alpha) - \alpha \Psi(x_\alpha, y_\alpha) = \sup_{x, y \in E} \{u(x) - v(y) - \alpha \Psi(x, y)\}.$$

Suppose that

$$\liminf_{\alpha \to \infty} H\left(x_\alpha, \alpha (\nabla \Psi(\cdot, y_\alpha))(x_\alpha)\right) - H\left(y_\alpha, \alpha (\nabla \Psi(\cdot, y_\alpha))(x_\alpha)\right) \leq 0,$$

then $u \leq v$. *In other words,* $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$ *satisfies the comparison principle.*

Proof. Fix $\lambda > 0$ and $h \in C(E)$. Let u be a subsolution and v a supersolution to

$$f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0. \tag{3.3.1}$$

We argue by contradiction and assume that $\delta := \sup_{x \in E} u(x) - v(x) > 0$. For $\alpha > 0$, let x_α, y_α be such that

$$u(x_\alpha) - v(y_\alpha) - \alpha \Psi(x_\alpha, y_\alpha) = \sup_{x, y \in E} \{u(x) - v(y) - \alpha \Psi(x, y)\}.$$

Thus Lemma 3.3.1 yields $\alpha \Psi(x_\alpha, y_\alpha) \to 0$ and for any limit point z of the sequence x_α, we have $u(z) - v(z) = \sup_{x \in E} u(x) - v(x) = \delta > 0$. It follows that for α large enough, $u(x_\alpha) - v(y_\alpha) \geq \frac{1}{2}\delta$.

For every $\alpha > 0$, the map $\Phi^1_\alpha(x) := v(y_\alpha) + \alpha \Psi(x, y_\alpha)$ is in $C^1(E)$ and $u(x) - \Phi^1_\alpha(x)$ has a maximum at x_α. On the other hand, $\Phi^2_\alpha(y) := u(x_\alpha) -$

$\alpha\Psi(x_\alpha, y)$ is also in $C^1(E)$ and $v(y) - \Phi^2_\alpha(y)$ has a minimum at y_α. As u is a sub- and v a super solution to (3.3.1), we have

$$\begin{aligned}
\frac{u(x_\alpha) - h(x_\alpha)}{\lambda} &\leq H(x_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)) \\
\frac{v(y_\alpha) - h(y_\alpha))}{\lambda} &\geq H(y_\alpha, -\alpha(\nabla\Psi(x_\alpha, \cdot))(y_\alpha)) \\
&= H(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)),
\end{aligned}$$

where the last equality follows as Ψ is a good penalization function. It follows that for α large enough, we have

$$\begin{aligned}
0 &< \frac{\delta}{2\lambda} \leq \frac{u(x_\alpha) - v(y_\alpha)}{\lambda} \qquad (3.3.2)\\
&= \frac{u(x_\alpha) - h(x_\alpha)}{\lambda} - \frac{v(y_\alpha) - h(y_\alpha)}{\lambda} + \frac{1}{\lambda}\left(h(x_\alpha) - h(y_\alpha)\right) \\
&\leq H(x_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)) - H(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)) \\
&\qquad + \frac{1}{\lambda}\left(h(x_\alpha) - h(y_\alpha)\right).
\end{aligned}$$

As h is continuous, we obtain $\lim_{\alpha\to\infty} h(x_\alpha) - h(y_\alpha) = 0$. Together with the assumption of the proposition, we find that the $\liminf$ as $\alpha \to \infty$ of the third line in (3.3.2) is bounded above by 0, which contradicts the assumption that $\delta > 0$. □

The next lemma gives additional control on the sequences x_α, y_α.

Lemma 3.3.3. *Let $(H, \mathcal{D}(H))$ be an operator such that $\mathcal{D}(H) = C^1(E)$ of the form $Hf(x) = H(x, \nabla f(x))$. Let u be a subsolution and v a supersolution to $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$, for some $\alpha > 0$ and $h \in C(E)$. Let Ψ be a good penalization function and let x_α, y_α satisfy*

$$u(x_\alpha) - v(y_\alpha) - \alpha\Psi(x_\alpha, y_\alpha) = \sup_{x,y\in E} \left\{u(x) - v(y) - \alpha\Psi(x, y)\right\}.$$

Then we have that

$$\sup_\alpha H\left(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) < \infty. \qquad (3.3.3)$$

Proof. Fix $\lambda > 0$, $h \in C(E)$ and let u and v be sub- and super-solutions to $f(x) - \lambda H(x, f(x)) - h(x) = 0$. Let Ψ be a good penalization function and let x_α, y_α satisfy

$$u(x_\alpha) - v(y_\alpha) - \alpha\Psi(x_\alpha, y_\alpha) = \sup_{x,y\in E} \left\{u(x) - v(y) - \alpha\Psi(x, y)\right\}.$$

As y_α is such that

$$v(y_\alpha) - (u(x_\alpha) - \Psi(x_\alpha, y_\alpha)) = \inf_y v(y) - (u(x_\alpha) - \Psi(x_\alpha, y)),$$

and v is a super-solution, we obtain

$$H\left(y_\alpha, -\alpha(\nabla\Psi(x_\alpha,\cdot))(y_\alpha)\right) \leq \frac{v(y_\alpha) - h(y_\alpha)}{\lambda}$$

As Ψ is a good penalization function, we have $-(\nabla\Psi(x_\alpha,\cdot))(y_\alpha) = (\nabla\Psi(\cdot, y_\alpha))(x_\alpha)$. The boundedness of v and h imply

$$\sup_\alpha H\left(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) \leq \frac{1}{\lambda}\left(v(y_\alpha) - h(y_\alpha)\right) \leq \frac{\|v-h\|}{\lambda} < \infty.$$

□

3.3.1 *One-dimensional Ehrenfest model*

To single out the important aspects of the proof of the comparison principle for equation (3.1.5), we start by proving it for the $d = 1$ case of Theorem 3.1.1.

Proposition 3.3.4. *Let $E = [-1, 1]$ and let*

$$H(x,p) = v_+(x)\left[e^{2p} - 1\right] + v_-(x)\left[e^{-2p} - 1\right],$$

where v_+, v_- are continuous and satisfy the following properties:

(a) $v_+(x) = 0$ for all x or v_+ satisfies the following properties:
   (i) $v_+(x) > 0$ for $x \neq 1$.
   (ii) $v_+(1) = 0$ and there exists a neighbourhood U_1 of 1 on which there exists a decomposition $v_+(x) = v_{+,\dagger}(x)v_{+,\ddagger}(x)$ such that $v_{+,\dagger}$ is decreasing and where $v_{+,\ddagger}$ is continuous and satisfies $v_{+,\ddagger}(1) \neq 0$.

(b) $v_-(x) = 0$ for all x or v_- satisfies the following properties:
   (i) $v_-(x) > 0$ for $x \neq -1$.
   (ii) $v_+(-1) = 0$ and there exists a neighbourhood U_{-1} of 1 on which there exists a decomposition $v_-(x) = v_{-,\dagger}(x)v_{-,\ddagger}(x)$ such that $v_{-,\dagger}$ is increasing and where $v_{-,\ddagger}$ is continuous and satisfies $v_{-,\ddagger}(-1) \neq 0$.

Let $\lambda > 0$ and $h \in C(E)$. Then the comparison principle holds for $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$.

Proof. Fix $\lambda > 0$, $h \in C(E)$ and pick a sub- and super-solutions u and v to $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$. We check the condition for Proposition 3.3.2. We take the good penalization function $\Psi(x,y) = 2^{-1}(x-y)^2$ and let x_α, y_α satisfy

$$u(x_\alpha) - v(y_\alpha) - \frac{\alpha}{2}|x_\alpha - y_\alpha|^2 = \sup_{x,y \in E} \left\{ u(x) - v(y) - \frac{\alpha}{2}|x-y|^2 \right\}.$$

We need to prove that

$$\liminf_{\alpha \to \infty} H(x_\alpha, \alpha(x_\alpha - y_\alpha)) - H(y_\alpha, \alpha(x_\alpha - y_\alpha)) \leq 0. \tag{3.3.4}$$

By Lemma 3.3.1, we know that $\alpha|x_\alpha - y_\alpha|^2 \to 0$ as $\alpha \to \infty$ and any limit point of (x_α, y_α) is of the form (z,z) for some z such that $u(z) - v(z) = \max_{z' \in E} u(z') - v(z')$. Restrict α to the sequence $\alpha \in \mathbb{N}$ and extract a subsequence, which we will also denote by α, such that $\alpha \to \infty$ x_α and y_α converge to some z. The rest of the proof depends on whether $z = -1$, $z = 1$ or $z \in (-1,1)$.

First suppose that $z \in (-1,1)$. By Lemma 3.3.3, we have

$$\sup_\alpha v_+(y_\alpha)\left[e^{2\alpha(x_\alpha - y_\alpha)} - 1\right] + v_-(y_\alpha)\left[e^{-2\alpha(x_\alpha - y_\alpha)} - 1\right] < \infty.$$

As $e^c - 1 > -1$, we see that the $\limsup$ of both terms of the sum individually are bounded as well. Using that $y_\alpha \to z \in (-1,1)$, and the fact that v_+, v_- are bounded away from 0 on a closed interval around z, we obtain from the first term that $\sup_\alpha \alpha(x_\alpha - y_\alpha) < \infty$ and from the second that $\sup_\alpha \alpha(y_\alpha - x_\alpha) < \infty$. We conclude that $\alpha(x_\alpha - y_\alpha)$ is a bounded sequence. Therefore, there exists a subsequence $\alpha(k)$ such that $\alpha(k)(x_{\alpha(k)} - y_{\alpha(k)})$ converges to some p_0. We find that

$$\begin{aligned}
&\liminf_{\alpha \to \infty} H(x_\alpha, \alpha(x_\alpha - y_\alpha)) - H(y_\alpha, \alpha(x_\alpha - y_\alpha)) \\
&\leq \lim_{k \to \infty} H(x_{\alpha(k)}, \alpha(x_{\alpha(k)} - y_{\alpha(k)})) - H(y_{\alpha(k)}, \alpha(x_{\alpha(k)} - y_{\alpha(k)})) \\
&= H(z, p_0) - H(z, p_0) = 0.
\end{aligned}$$

We proceed with the proof in the case that $x_\alpha, y_\alpha \to z = -1$. The case where $z = 1$ is proven similarly. Again by Lemma 3.3.3, we obtain the bounds

$$\sup_\alpha v_+(y_\alpha)\left[e^{2\alpha(x_\alpha - y_\alpha)} - 1\right] < \infty,$$

$$\sup_\alpha v_-(y_\alpha)\left[e^{-2\alpha(x_\alpha - y_\alpha)} - 1\right] < \infty. \tag{3.3.5}$$

As v_+ is bounded away from 0 near -1, we obtain by the left hand bound that $\sup_\alpha \alpha(x_\alpha - y_\alpha) < \infty$. As in the proof above, it follows that if $\alpha|x_\alpha - y_\alpha|$ is bounded, we are done. This leaves the case where there exists a subsequence of α, denoted by $\alpha(k)$, such that $\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)}) \to \infty$. Then clearly, $e^{2\alpha(k)(x_{\alpha(k)}-y_{\alpha(k)})} - 1$ is bounded and contains a converging subsequence. We obtain as in the proof where $z \in (-1,1)$ that

$$\begin{aligned}
&\liminf_{\alpha\to\infty} H(x_\alpha, \alpha(x_\alpha - y_\alpha)) - H(y_\alpha, \alpha(x_\alpha - y_\alpha)) \\
&\quad = \liminf_{\alpha\to\infty} \left[v_+(x_\alpha) - v_+(y_\alpha)\right]\left[e^{2\alpha(x_\alpha - y_\alpha)} - 1\right] \\
&\qquad\qquad + \left[v_-(x_\alpha) - v_-(y_\alpha)\right]\left[e^{2\alpha(y_\alpha - x_\alpha)} - 1\right] \\
&\quad \leq \liminf_{k\to\infty} \left[v_-(x_{\alpha(k)}) - v_-(y_{\alpha(k)})\right]\left[e^{2\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)})} - 1\right].
\end{aligned}$$

Note that as $\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)}) \to \infty$, we have $y_{\alpha(k)} > x_{\alpha(k)} \geq -1$, which implies $v_-(y_{\alpha(k)}) > 0$. Also for k sufficiently large, $y_{\alpha(k)}, x_{\alpha(k)} \in U_{-1}$. Thus, we can write

$$\begin{aligned}
&\left[v_-(x_{\alpha(k)}) - v_-(y_{\alpha(k)})\right]\left[e^{2\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)})} - 1\right] \\
&= \left[\frac{v_{-,\dagger}(x_{\alpha(k)})}{v_{-,\dagger}(y_{\alpha(k)})}\frac{v_{-,\ddagger}(x_{\alpha(k)})}{v_{-,\ddagger}(y_{\alpha(k)})} - 1\right] v_-(y_{\alpha(k)}) \left[e^{2\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)})} - 1\right].
\end{aligned}$$

By the bound in (3.3.5), and the obvious lower bound, we see that the non-negative sequence

$$u_k := v_-(y_{\alpha(k)}) \left[e^{2\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)})} - 1\right]$$

contains a converging subsequence $u_{k'} \to c$. As $y_{\alpha(k)} > x_{\alpha(k)}$ and $v_{-,\dagger}$ is increasing:

$$\begin{aligned}
&\limsup_k \frac{v_{-,\dagger}(x_{\alpha(k)})}{v_{-,\dagger}(y_{\alpha(k)})}\frac{v_{-,\ddagger}(x_{\alpha(k)})}{v_{-,\ddagger}(y_{\alpha(k)})} \\
&\qquad \leq \left(\limsup_k \frac{v_{-,\dagger}(x_{\alpha(k)})}{v_{-,\dagger}(y_{\alpha(k)})}\right)\left(\lim_k \frac{v_{-,\ddagger}(x_{\alpha(k)})}{v_{-,\ddagger}(y_{\alpha(k)})}\right) \leq \frac{v_{-,\ddagger}(-1)}{v_{-,\ddagger}(-1)} = 1.
\end{aligned}$$

As a consequence, we obtain

$$\begin{aligned}
&\liminf_k \left[\frac{v_-(x_{\alpha(k)})}{v_-(y_{\alpha(k)})} - 1\right] v_-(y_{\alpha(k)}) \left[e^{2\alpha(k)(y_{\alpha(k)} - x_{\alpha(k)})} - 1\right] \\
&\quad \leq \left(\limsup_k \left[\frac{v_{-,\dagger}(x_{\alpha(k)})}{v_{-,\dagger}(y_{\alpha(k)})}\frac{v_{-,\ddagger}(x_{\alpha(k)})}{v_{-,\ddagger}(y_{\alpha(k)})} - 1\right]\right)\left(\liminf_{k'} u_{k'}\right) \leq 0.
\end{aligned}$$

This concludes the proof of (3.3.4) for the case that $z = -1$. □

3.3.2 *Multi-dimensional Ehrenfest model*

Proof of Theorem 3.1.6. Let u be a subsolution and v a supersolution to $f(x) - \lambda H(x, \nabla f(x)) - h(x) = 0$. As in the proof of Proposition 3.3.4, we check the condition for Proposition 3.3.2. Again, for $\alpha \in \mathbb{N}$ let x_α, y_α satisfy

$$u(x_\alpha) - v(y_\alpha) - \frac{\alpha}{2}|x_\alpha - y_\alpha|^2 = \sup_{x,y \in E} \left\{ u(x) - v(y) - \frac{\alpha}{2}|x-y|^2 \right\}.$$

and without loss of generality let z be such that $x_\alpha, y_\alpha \to z$.

Denote with $x_{\alpha,i}$ and $y_{\alpha,i}$ the i-th coordinate of x_α respectively y_α. We prove

$$\begin{aligned} &\liminf_{\alpha \to \infty} H(x_\alpha, \alpha(x_\alpha - y_\alpha)) - H(y_\alpha, \alpha(x_\alpha - y_\alpha)) \\ &= \liminf_{\alpha \to \infty} \sum_i \Bigg\{ \left[v_+^i(x_\alpha) - v_+^i(y_\alpha)\right] \left[e^{\alpha(x_{\alpha,i} - y_{\alpha,i})} - 1\right] \\ &\qquad\qquad + \left[v_-^i(x_\alpha) - v_-^i(y_\alpha)\right] \left[e^{\alpha(y_{\alpha,i} - x_{\alpha,i})} - 1\right] \Bigg\} \leq 0, \end{aligned}$$

by constructing a subsequence $\alpha(n) \to \infty$ such that the first term in the sum converges to 0. From this sequence, we find a subsequence such that the second term converges to zero, and so on.

Therefore, we will assume that we have a sequence $\alpha(n) \to \infty$ for which the first $i-1$ terms of the difference of the two Hamiltonians vanishes and prove that we can find a subsequence for which the i-th term

$$\begin{aligned} &\left[v_+^i(x_\alpha) - v_+^i(y_\alpha)\right] \left[e^{\alpha(x_{\alpha,i} - y_{\alpha,i})} - 1\right] \\ &\qquad + \left[v_-^i(x_\alpha) - v_-^i(y_\alpha)\right] \left[e^{\alpha(y_{\alpha,i} - x_{\alpha,i})} - 1\right] \qquad (3.3.6) \end{aligned}$$

vanishes. This follows directly as in the proof of Proposition 3.3.4, arguing depending on the situation $z_i \in (-1, 1)$, $z_i = -1$ or $z_i = -1$. □

3.3.3 *Mean field Markov jump processes on a finite state space*

The proof of Theorem 3.1.7 follows along the lines of the proofs of Proposition 3.3.4 and Theorem 3.1.6. The proof however needs one important adaptation because of the appearance of the difference $p_b - p_a$ in the exponents of the Hamiltonian.

Naively copying the proofs using the penalization function $\Psi(\mu,\nu) = \frac{1}{2}\sum_a (\mu(a) - \nu(a))^2$ one obtains by Lemma 3.3.3 , for suitable sequences μ_α and ν_α, that

$$\sup_\alpha v(a,b,\nu_\alpha)\left[e^{\alpha((\mu_\alpha(b)-\nu_\alpha(b))-(\mu_\alpha(a)-\nu_\alpha(a)))} - 1\right] < \infty.$$

One sees that the control on the sequences $\alpha(\nu_\alpha(a) - \mu_\alpha(a))$ obtained from this bound is not very good, due to the compensating term $\alpha(\mu_\alpha(b) - \nu_\alpha(b))$.

The proof can be suitably adapted using a different penalization function. For $x \in \mathbb{R}$, let $x^- := x \wedge 0$ and $x^+ = x \vee 0$. Define $\Psi(\mu,\nu) = \frac{1}{2}\sum_a ((\mu(a) - \nu(a))^-)^2 = \frac{1}{2}\sum_a ((\nu(a) - \mu(a))^+)^2$. Clearly, Ψ is differentiable in both components and satisfies $(\nabla\Psi(\cdot,\nu))(\mu) = -(\nabla\Psi(\mu,\cdot))(\nu)$. Finally, using the fact that $\sum_i \mu(i) = \sum_i \nu(i) = 1$, we find that $\Psi(\mu,\nu) = 0$ implies that $\mu = \nu$. We conclude that Ψ is a good penalization function.

The bound obtained from Lemma 3.3.3 using this Ψ yields

$$\sup_\alpha v(a,b,\nu_\alpha)\left[e^{\alpha\left((\mu_\alpha(b)-\nu_\alpha(b))^- - (\mu_\alpha(a)-\nu_\alpha(a))^-\right)} - 1\right] < \infty.$$

We see that if $(\mu_\alpha(b) - \nu_\alpha(b))^- - (\mu_\alpha(a) - \nu_\alpha(a))^- \to \infty$ it must be because $\alpha(\nu_\alpha(a) - \mu_\alpha(a)) \to \infty$. This puts us in the position to use the techniques from the previous proofs.

Proof of Theorem 3.1.7. Set $\Psi(\mu,\nu) = \frac{1}{2}\sum_a ((\mu(a) - \nu(a))^-)^2$, as above. We already noted that Ψ is a good penalization function.

Let u be a subsolution and v be a supersolution to $f(\mu) - \lambda H(\mu, \nabla f(\mu)) - h(\mu) = 0$. For $\alpha \in \mathbb{N}$, pick μ_α and ν_α such that

$$u(\mu_\alpha) - v(\nu_\alpha) - \alpha\Psi(\mu_\alpha,\nu_\alpha) = \sup_{\mu,\nu \in E}\{u(\mu) - v(\nu) - \alpha\Psi(\mu,\nu)\}$$

Furthermore, assume without loss of generality that $\mu_\alpha, \nu_\alpha \to z$ for some z such that $u(z) - v(z) = \sup_{z' \in E} u(z') - v(z')$. By Proposition 3.3.2, we need to bound

$$\begin{aligned} &H(\mu_\alpha, \alpha(\nabla\Phi(\cdot,\nu_\alpha))(\mu_\alpha)) - H(\nu_\alpha, \alpha(\nabla\Phi(\mu_\alpha,\cdot))(\mu_\alpha)) \\ &= \sum_{a,b}\left[v(a,b,\mu_\alpha) - v(a,b,\nu_\alpha)\right] \\ &\qquad \times \left[e^{\alpha\left((\mu_\alpha(b)-\nu_\alpha(b))^- - (\mu_\alpha(a)-\nu_\alpha(a))^-\right)} - 1\right]. \end{aligned} \tag{3.3.7}$$

As in the proof of Theorem 3.1.6, we will show that each term in the sum above can be bounded above by 0 separately. So pick some ordering of the ordered pairs (i,j), $i,j \in \{1,\dots,n\}$ and assume that we have some sequence α such that the $\liminf_{\alpha\to\infty}$ of the first k terms in equation (3.3.7) are bounded above by 0. Suppose that (i,j) is the pair corresponding to the $k+1$-th term of the sum in (3.3.7).

Clearly, if $v(i,j,\pi) = 0$ for all π then we are done. Therefore, we assume that $v(i,j,\pi) \neq 0$ for all π such that $\pi(i) > 0$.

In the case that $\mu_\alpha, \nu_\alpha \to \pi^*$, where $\pi^*(i) > 0$, we know by Lemma 3.3.3, using that $v(i,j,\cdot)$ is bounded away from 0 on a neighbourhood of π^*, that

$$\sup_\alpha e^{\alpha\left((\mu_\alpha(j)-\nu_\alpha(j))^- - (\mu_\alpha(i)-\nu_\alpha(i))^-\right)} - 1 < \infty.$$

Picking a subsequence $\alpha(n)$ such that this term above converges and using that $\pi \to v(i,j,\pi)$ is uniformly continuous, we see

$$\begin{aligned}
&\liminf_{\alpha\to\infty} \left[v(i,j,\mu_\alpha) - v(i,j,\nu_\alpha)\right] \times \\
&\qquad\qquad \left[e^{\alpha\left((\mu_\alpha(j)-\nu_\alpha(j))^- - (\mu_\alpha(i)-\nu_\alpha(i))^-\right)} - 1\right] \\
&= \lim_{n\to\infty} \left[v(i,j,\mu_{\alpha(n)}) - v(i,j,\nu_{\alpha(n)})\right] \times \\
&\qquad \left[e^{\alpha(n)\left(\left(\mu_{\alpha(n)}(j)-\nu_{\alpha(n)}(j)\right)^- - \left(\mu_{\alpha(n)}(i)-\nu_{\alpha(n)}(i)\right)^-\right)} - 1\right] = 0
\end{aligned}$$

For the second case, suppose that $\mu_\alpha(i), \nu_\alpha(i) \to 0$. By Lemma 3.3.3, we get

$$\sup_\alpha v(i,j,\nu_\alpha) \left[e^{\alpha\left((\mu_\alpha(j)-\nu_\alpha(j))^- - (\mu_\alpha(i)-\nu_\alpha(i))^-\right)} - 1\right] < \infty. \tag{3.3.8}$$

First of all, if $\sup_\alpha \alpha\left((\mu_\alpha(j) - \nu_\alpha(j))^- - (\mu_\alpha(i) - \nu_\alpha(i))^-\right) < \infty$, then the argument given above also takes care of this situation. So suppose that this supremum is infinite. Clearly, the contribution $(\mu_\alpha(j) - \nu_\alpha(j))^-$ is negative, which implies that $\sup_\alpha \alpha\left(\nu_\alpha(i) - \mu_\alpha(i)\right)^+ = \infty$. This means that we can assume without loss of generality that

$$\alpha\left(\nu_\alpha(i) - \mu_\alpha(i)\right) \to \infty, \qquad \nu_\alpha(i) > \mu_\alpha(i). \tag{3.3.9}$$

We rewrite the term $a = i$, $b = j$ in equation (3.3.7) as

$$\left[\frac{v(i,j,\mu_\alpha)}{v(i,j,\nu_\alpha)} - 1\right] v(i,j,\nu_\alpha) \left[e^{\alpha\left((\mu_\alpha(j)-\nu_\alpha(j))^- - (\mu_\alpha(i)-\nu_\alpha(i))^-\right)} - 1\right].$$

The right hand side is bounded above by (3.3.8) and bounded below by -1, so we take a subsequence of α, also denoted by α, such that the right hand side converges. Also note that for α large enough the right hand side is non-negative. Therefore, it suffices to show that

$$\liminf_{\alpha\to\infty} \frac{v(i,j,\mu_\alpha)}{v(i,j,\nu_\alpha)} \leq 1,$$

which follows as in the proof of Proposition 3.3.4. □

4 EXPONENTIAL DECAY OF ENTROPY AND ENTROPIC INTERPOLATIONS

In Chapter 3, we considered Markov processes x_n with generators A_n that take values in some closed convex set $E_n \subseteq \mathbb{R}^d$. In particular, we saw in Proposition 3.1.8 that the rate function S of the stationary measures of the processes A_n is a Lyapunov function for the McKean-Vlasov equation $\dot{x} = H_p(x, 0)$.

In this chapter, we study the decay of this rate function, or entropy, S along the flow of the McKean-Vlasov equation in more detail and in a more general context. We will give conditions for exponential decay

$$S(x(t)) \leq e^{-\alpha t} S(x(0))$$

and verify these conditions for Glauber dynamics on the Curie-Weiss model and for the Wright-Fisher diffusion model with parent-independent mutations.

Afterwards, we extend the definition of entropic-interpolations introduced by Léonard [2013]. Léonard considers entropic interpolations for the measure valued flow of a Markov-process and are defined via a (f, g) transform, which is essentially an extension of the classical Doob-h transform.

Noting, however, as in our introduction, that the flow of laws of a Markov process can be seen as the solution to the Kolmogorov forward equation and thus, as the minimal cost trajectory of a path-space large deviation principle, the entropic interpolation can also be seen as a trajectory that connects two points with minimal Lagrangian cost.

This is our starting point, and we give conditions for the convexity of the entropy along an entropic interpolation. This concept is well known in the theory of displacement interpolations, see von Renesse and Sturm [2005], Chapter 16 in Villani [2009] or Chapter 9 in Bakry et al. [2014], and can be used to study the curvature of the underlying space. Because displacement interpolations are not suited to study processes on discrete spaces, a different kind of interpolation is needed.

A powerful interpolation method for discrete spaces is introduced in Erbar and Maas [2012], but this definition is not directly suitable for the dynamics obtained from interacting systems. Here, we make a first step for entropic interpolations in this general setting.

4.1 LARGE DEVIATIONS AND THE MCKEAN-VLASOV EQUATION

For clarity, we state the setting for the results in this chapter. Given a closed subset E of $\mathbb{R}^d$, we assume the existence of a sequence of measures $\mathbb{P}_n \in \mathcal{P}(D_E(\mathbb{R}^+))$ so that the large deviation principle holds for the trajectories:

$$\mathbb{P}_n\left[\{x(t)\}_{t\geq 0} \approx \{\gamma(t)\}_{t\geq 0}\right] \approx e^{-nu(\gamma)}. \tag{4.1.1}$$

We assume that the *rate function* $u : D_E(\mathbb{R}^+) \to [0,\infty]$ has compact level sets and has the form

$$u(\gamma) = \begin{cases} u_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot\gamma(s))\mathrm{d}s & \text{if } \gamma \text{ is absolutely continuous} \\ \infty & \text{otherwise,} \end{cases}$$

where the *Lagrangian* $\mathcal{L} : E \times \mathbb{R}^d \to [0,\infty)$ is lower semi-continuous and for each $x \in E$, $v \mapsto \mathcal{L}(x,v)$ is convex. We define the *Hamiltonian* H as $H(x,p) = \inf_v \langle p, v\rangle - \mathcal{L}(x,v)$. Such large deviation principles are obtained in various contexts, notably in Chapter 3 and Chapter 5, but also more generally for Freidlin-Wentzell theory Freidlin and Wentzell [1998], for Levy processes Feng and Kurtz [2006], for other interacting jump processes Dupuis et al. [2016] or for the Wright-Fisher model for population dynamics Dawson and Feng [1998].

In the setting that the measures $\mathbb{P}_n$ correspond to Markov processes for which there exists stationary measures μ_n that satisfy the large deviation principle with rate function S, this S is a Lyapunov function for the *McKean-Vlasov* equation $\dot{x} = H_p(x,0)$. Here H_p denotes the vector of derivatives of H in the second coordinate. To be precise, it was found in Chapter 3, but also in Roeck et al. [2006] that $S(x(t)) \leq S(x(0))$ for any $t \geq 0$.

In this chapter, we analyse the decay of the entropy S along the flow of the McKean-Vlasov equation in more detail. We will give conditions for exponential decay

$$S(x(t)) \leq e^{-\alpha t} S(x(0)).$$

Afterwards, we will extend the definition of entropic-interpolations introduced in Léonard [2013] and give conditions for the convexity of the entropy along these entropic interpolations.

The standing assumption on H and S for the results in this paper are the following.

Assumption 4.1.1. We assume that the continuous Hamiltonian $H : E \times \mathbb{R}^d \to \mathbb{R}$ satisfies

H(a) H is twice continuously differentiable,

H(b) for every $x \in E$, the map $p \mapsto H(x,p)$ is convex and for every x in the interior of E, the map $p \mapsto H(x,p)$ is strictly convex.

There exists a continuous function $S : E \to [0,\infty)$ such that

S(a) S is twice continuously differentiable on the interior of E,

S(b) for $x \in E^\circ$, we have $H(x, DS(x)) = 0$,

S(c) S is a Lyapunov function for the McKean-Vlasov equation: if $x(t)$ solves $\dot{x}(t) = H_p(x(t), 0)$ then $S(x(t)) \leq S(x(s))$ for all $0 \leq s \leq t$.

Remark 4.1.2. The assumption that S is twice continuously differentiable can be relaxed to once continuously differentiable in various situations.

The following three results verify that in the setting where $\mathbb{P}_n$ are Markovian, the stationary measures satisfy the large deviation principle with rate function S and where S is differentiable on the interior of E, the conditions on S of the assumption above are satisfied.

The proof of Proposition 3.1.8 generalizes and we obtain the following result.

Proposition 4.1.3. *Suppose the measures $\mathbb{P}_n$ correspond to Markov processes for which there exists stationary measures μ_n that satisfy the large deviation principle with rate function S.*

Let $\{x(t)\}_{t\geq 0}$ be a solution to the McKean-Vlasov equation $\dot{x} = H_p(x,0)$, then $S(x(t)) \leq S(x(s))$ for all $0 \leq s \leq t$.

In the following proposition we show that S is a solution to $Hf = 0$ in the viscosity sense.

Lemma 4.1.4. *Suppose the measures $\mathbb{P}_n$ correspond to Markov processes for which there exists stationary measures μ_n that satisfy the large deviation principle with rate function S. Then S is a viscosity solution to $HS = 0$.*

Proof. By a standard argument using dynamic programming, cf. Theorem 6.4.5 in Cannarsa and Sinestrari [2004], we find that for any function u_0, the function

$$u(x,t) = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(t)=x}} u_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s$$

is a viscosity solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} u(x,t) + H(x, \nabla u(x,t)) = 0.$$

on $E \times (0,T)$. In the case that $u_0 = S$ is the large deviation rate function of the stationary measures, it follows that $u(\cdot, t) = S$ for all $t \geq 0$ by the contraction principle. As a direct consequence, we find that S is a viscosity solution of $H(x, \nabla S(x)) = 0$ on E. □

Viscosity solutions have the property that the equation is satisfied at any point where the viscosity solution is differentiable. This gives the following lemma.

Lemma 4.1.5. *Suppose the measures $\mathbb{P}_n$ correspond to Markov processes for which there exists stationary measures μ_n that satisfy the large deviation principle with rate function S. Then if x is in the interior of E and is such that S is differentiable at x, then $H(x, DS(x)) = 0$.*

4.2 EXPONENTIAL DECAY OF ENTROPY

We start by studying the decay of S along the solutions of the McKean-Vlasov equation. Motivated by the analogous quantities in the theory of (modified) logarithmic Sobolev inequalities, we define the concept of information.

Definition 4.2.1. Let H and S satisfy Assumption 4.1.1. We define the *information* $I : E \to \mathbb{R}^+$ by

$$I(x) = -\langle DS(x), H_p(x, 0) \rangle.$$

We say that H satisfies a *entropy-information inequality* (EII) with constant $\alpha > 0$ if for all $x \in E$:

$$\alpha S(x) \leq I(x).$$

Note that as $I(x) = -\frac{\mathrm{d}}{\mathrm{d}t}|_{t=0}S(x(t))$ for the solution $x(t)$ to the McKean-Vlasov equation with $x(0) = x$, it follows that $I(x) \geq 0$ by Proposition 4.1.3.

This entropy-information inequality is a naturally connected to similar inequalities present in the literature.

In the setting of the measure-valued flow generated by the Kolmogorov forward equation of a diffusion operator, the derivative of the entropy along the flow is called the *Fisher information*. Thus, the entropy information inequality is analogous to the log-Sobolev inequality, we refer to Section 5.2 in Bakry et al. [2014].

For the measure valued flow generated by the Kolmogorov forward equation of a Markov jump process, the entropy-information inequality coincides with the modified logarithmic Sobolev inequality. See for example Caputo et al. [2009] where this inequality is connected to the decay of entropy along the Kolmogorov forward equation of jump processes.

Lemma 4.2.2. *Let H and S satisfy Assumption 4.1.1. Let $x(t)$ solve the McKean-Vlasov equation: $\dot{x} = H_p(x,0)$. Then H satisfies (EII)-α if and only if*

$$S(x(t)) \leq e^{-\alpha t}S(x(0)).$$

Proof. By (EII)-α, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}S(x(t)) = \langle DS(x(t)), H_p(x(t),0)\rangle = -I(x) \leq -\alpha S(x(t)).$$

It follows by Grönwall's inequality that

$$S(x(t)) \leq e^{-\alpha t}S(x(0)).$$

The reverse implication follows by differentiation. □

It is well known that control on the second derivative of the entropy along solutions of the McKean-Vlasov equation yields stronger control on the decay of the entropy, see for example Lemma 2.1 in Caputo et al. [2009]. The second derivative of the entropy of S gives:

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2}S(x(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle DS(x(t)), H_p(x(t),0)\rangle \\ &= \langle D^2S(x(t))H_p(x(t),0), H_p(x(t),0)\rangle \\ &\qquad + \langle DS(x(t)), H_{px}(x(t),0)H_p(x(t),0)\rangle. \end{aligned} \tag{4.2.1}$$

We obtain the following result, giving an inequality that implies (EII)-α if there is only one attracting stationary point.

Lemma 4.2.3. *Let H and S satisfy Assumption 4.1.1. Let $\{x(t)\}_{t\geq 0}$ be a solution to the McKean-Vlasov equation $\dot{x} = H_p(x,0)$. Then the following two statements are equivalent.*

(a) For all $x \in E$, we have

$$\alpha I(x) \leq \langle DS(x), H_{px}(x,0)H_p(x,0)\rangle + \langle D^2S(x)H_p(x,0), H_p(x,0)\rangle. \tag{4.2.2}$$

(b) For all solutions $\{x(t)\}_{t\geq 0}$ of the McKean-Vlasov equation, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}S(x(t)) \geq e^{-\alpha t}\frac{\mathrm{d}}{\mathrm{d}t}S(x(t))|_{t=0}, \quad \textit{and} \quad I(x(t)) \leq e^{-\alpha t}I(x(0))$$

Suppose that S is bounded from below and that (4.2.2) is satisfied. Let $S_\infty := \lim_{t\to\infty} S(x(t))$, (which exists as S is decreasing along solutions of the McKean-Vlasov equation), then

$$S(x(t)) - S_\infty \leq e^{-\alpha t}\left(S(x(0)) - S_\infty\right). \tag{4.2.3}$$

Remark 4.2.4. If $S(x)$ is convex, $D^2S(x)$ is a positive operator. Thus, a weaker criterion for the exponential decay of entropy is given by

$$\alpha I(x) \leq \langle DS(x), H_{px}(x,0)H_p(x,0)\rangle. \tag{4.2.4}$$

Proof of Lemma 4.2.3. By (4.2.1), we note that (4.2.2) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}I(x(t)) \leq -\alpha I(x(t))$$

As in the proof of Lemma 4.2.3, we obtain the first claim of the lemma.

We proceed with the proof of (4.2.3). First, we integrate from t to T the inequality

$$\alpha\frac{\mathrm{d}}{\mathrm{d}t}S(x(t)) = -\alpha I(x(t)) \geq \frac{\mathrm{d}}{\mathrm{d}t}I(x(t)),$$

which yields

$$\alpha\left[S(x(T)) - S(x(t))\right] \geq \left[I(x(T)) - I(x(t))\right].$$

As $T \to \infty$, we find by the first claim of the Lemma that $I(x(T)) \to 0$. Additionally, as S is decreasing along the solutions of the McKean-Vlasov equation $S(x(T)) \to S_\infty$. We conclude that

$$\alpha \left[S(x(t)) - S_\infty\right] \leq I(x(t)).$$

The claim follows as in the proof of Lemma 4.2.2. □

We give a one-dimensional example where (4.2.2) is satisfied and a second multi-dimensional example where we verify (4.2.4). The first case is related to the large deviations of the trajectories of the magnetization of a Curie-Weiss model under a Glauber-dynamics evolution with potential $V(x) = -\frac{1}{2}\beta x^2$, see Example 3.1.9 in Chapter 3. The second is related to the large deviation behaviour of a d-dimensional Wright-Fisher model with mutations with vanishing diffusion coefficient.

Proposition 4.2.5 (Interacting jump processes on two states). *Let $\beta > 0$ and consider the Hamiltonian $H : [-1,1] \times \mathbb{R} \to \mathbb{R}$ defined by*

$$H(x,p) = \frac{1-x}{2} e^{\beta x}\left[e^{2p} - 1\right] + \frac{1+x}{2} e^{-\beta x}\left[e^{-2p} - 1\right].$$

Note that it has an entropy functional $S(x) = \frac{1-x}{2}\log(1-x) + \frac{1+x}{2}\log(1+x) - \frac{1}{2}\beta x^2$ with gradient $DS(x) = \frac{1}{2}\log\frac{1+x}{1-x} - \beta x$.

If $\beta < 1$, then (H,S) satisfy (EII) with constant $4(1-\beta)$.

Proof. First of all, note that $D^2 S(x) = \frac{1}{1-x^2} - \beta$. Because $\beta \leq 1$, we find that S is convex. We first consider (4.2.4). An elementary computation yields

$$\begin{aligned} H_p(x,0) &= 2\sinh(\beta x) - 2x\cosh(\beta x) \\ H_{px}(x,0) &= -2\left[1-\beta\right]\cosh(\beta x) - 2x\beta\sinh(\beta x). \end{aligned}$$

Because $H_{px}(x,0) \leq -2(1-\beta)$, we obtain

$$\langle DS(x), H_{px}(x,0) H_p(x,0)\rangle \geq 2(1-\beta) I(x). \tag{4.2.5}$$

We proceed with the second part or the right hand side of (4.2.2). By assumption (a), S is strictly convex. In particular, there is a unique stationary point $x_s = 0$ where $DS(0) = 0$. By the first part of the proof, we know that solutions to the McKean-Vlasov equation converge to this stationary point. Thus $I(x) = 0$ if and only if $x = 0$. In particular, this implies that $H_p(x, 0) < 0$ for $x > 0$ and $H_p(x, 0) > 0$ for $x < 0$, and thus

$$-xH_p(x, 0) \geq 0. \tag{4.2.6}$$

Without loss of generality, we consider $x \geq 0$ and prove that

$$-2(1-\beta)DS(x)H_p(x,0) \leq H_p(x,0)^2 D^2 S(x).$$

For $x \geq 0$, $H_p(x, 0) < 0$, so it suffices to prove that

$$2(1-\beta)DS(x) \leq -H_p(x,0)D^2S(x).$$

If $x = 0$, both sides of the inequality equal 0 as $DS(0) = H_p(0, 0) = 0$. Thus, it suffices to prove that the derivatives are ordered in the same way:

$$\begin{aligned} 2(1-\beta)D^2S(x) &\leq -H_{px}(x,0)D^2S(x) - H_p(x,0)D^3S(x) \\ &= -H_{px}(x,0)D^2S(x) - \frac{2xH_p(x,0)}{(1-x^2)^2}. \end{aligned}$$

Hence, the claim is proven as $-xH_p(x, 0) \geq 0$ by (4.2.6). We conclude that the second part yields

$$\langle DS(x), H_{px}(x,0)H_p(x,0)\rangle \geq 2[1-\beta]I(x). \tag{4.2.7}$$

We conclude from (4.2.5) and (4.2.7) that we have a entropy-information inequality with constant $4(1 - \beta)$. □

It should be noted that in principle the proof can be generalized to a more involved structure for the potential. More conditions need to be posed and the conclusions need to be changed appropriately. For example, we can consider a potential that includes an external magnetic field: $V_h(x) = -\frac{1}{2}\beta x^2 - hx$. The first part of the proof above can be carried out without any changes. However, the second part of the proof above breaks down as the stationary point for the dynamics is not equal to 0. So even though the first part yields a constant $2(1 - \beta)$, the second part yields a constant that is strictly less than $2(1 - \beta)$.

Example 4.2.6 (Wright Fisher model with mutation)**.** For $E = \{x \in \mathbb{R}^d \,|\, x_i \geq 0, \sum x_i = 1\}$. In Dawson and Feng [1998] the large deviations of the trajectories of the Wright-Fisher model are considered, and the Hamiltonian corresponding to this LDP is given by

$$H(x,p) = \frac{1}{2}\sum_{i,j} x_i(\delta_{ij} - x_j)p_i p_j + \sum_{i=1}^{d}\left(\sum_{j\neq i} x_j q_{ji} - x_i q_{ij}\right) p_i$$

where q_{ji} represents the mutation rate from j to i. In the case that the mutation rates are parent independent: $q_{ji} = \frac{1}{2}\mu_i > 0$, for $i \neq j$, the stationary measures of the associated Wright-Fisher processes with vanishing diffusion coefficient have entropy $S(x)$ given by

$$S(x) = \sum_{i=1}^{d} \mu_i \log \frac{\mu_i}{\mu x_i}.$$

Proposition 4.2.7. *Consider H and S from Example 4.2.6 in the setting that $q_{ji} = \frac{1}{2}\mu_i > 0$. Define $\mu = \sum_i \mu_i$. Then H and S satisfy* (4.2.2) *and the entropy-information inequality with constant $\frac{1}{2}\mu$.*

In the proof below, we will only verify (4.2.4) as opposed to the jump process example above. The constant obtained here, is thus, not optimal. In Theorem 4.4.3 below, we consider the setting of two species and improve the (EII) bound by a factor that depends on the difference between μ_1 and μ_2.

Proof. For the verification of (4.2.2) with constant $\frac{1}{2}\mu$ we observe that $x \mapsto S(x)$ is convex, so it suffices to verify (4.2.4). Thus, we calculate the vector $H_p(x, 0)$ and matrix $H_{px}(x, 0)$. We find

$$\begin{aligned} H_{p_j}(x, 0) &= \frac{1}{2} \sum_{l \neq j} x_l \mu_j - x_j \mu_l \\ &= \frac{1}{2} \left((1 - x_j)\mu_j - x_j(\mu - \mu_j) \right) = \frac{1}{2} \left(\mu_j - x_j \mu \right), \end{aligned}$$

$$H_{p_i, x_j}(x, 0) = \begin{cases} 0 & \text{for } i \neq j, \\ -\frac{1}{2}\mu & \text{for } i = j. \end{cases}$$

We find that

$$\sum_j H_{p_i, x_j}(x, 0) H_{p_j}(x, 0) = -\frac{1}{2}\mu H_{p_i}(x, 0),$$

and as a consequence

$$\langle DS(x), H_{px}(x, 0) H_p(x, 0) \rangle = -\frac{1}{2}\mu \langle DS(x), H_p(x, 0) \rangle = \frac{1}{2}\mu I(x).$$

□

4.3 ENTROPIC INTERPOLATIONS

In this section, we will consider entropic interpolations, which we define as the optimal trajectories of the Lagrangian system, conditioned on a starting and end point.

In the setting where the Hamiltonian corresponds to the large deviation behaviour of the trajectories of the empirical density of independent copies of a process, i.e. Chapter 6, this definition is formally equivalent to the one introduced in Léonard [2013]. Léonard defines an entropic interpolation between π and ν in time T in terms of an (f, g) transform. Using the connection of the (f, g) transform to solutions of the Schrödinger problem in [Léonard, 2014, Theorem 3.3], this transform corresponds to the trajectory of measures $\{\mu(t)\}_{0 \leq t \leq T}$, where $\mu(t) := \mathbb{Q}_t^*$ is the law of $X(t)$ under $\mathbb{Q}^*$, and where $\mathbb{Q}^*$ minimizes

$$\inf \{S(\mathbb{Q} \,|\, \mathbb{P}) \,|\, \mathbb{Q}_0 = \pi, \mathbb{Q}_T = \nu\},$$

where $S(\cdot \,|\, \cdot)$ is the relative entropy. This minimization problem can be re-expressed in terms of the path-space large deviation problem of the trajectory of the empirical distribution of independent copies. This reformulation of the minimization problem generalizes to interacting systems and motivates the following definition.

Definition 4.3.1. We say that an absolutely continuous trajectory $\gamma^* : [0, T] \to E$ is an entropic interpolation between x and y in time T if $\gamma^*(0) = x$, $\gamma^*(T) = y$ and

$$\int_0^T \mathcal{L}(\gamma^*(s), \dot{\gamma}^*(s)) \mathrm{d}s = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(0)=x, \gamma(T)=y}} \int_0^T \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s.$$

We will make the following assumption in this Section, which is necessary in the case that S is not differentiable on the boundary of E. In the setting of one-dimensional reversible processes, we will show that this assumption is always satisfied.

Assumption 4.3.2. For the results in this section, we only consider entropic interpolations $\gamma : [0, T] \to E$ such that for $t \in (0, T)$ the trajectory is in the interior of E.

Consider the Hamilton equations:

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} H_p(x, p) \\ -H_x(x, p) \end{bmatrix} \tag{4.3.1}$$

Lemma 4.3.3. *Let H and S satisfy Assumption 4.1.1 and let x be an entropic interpolation satisfying Assumption 4.3.2. For $t \in (0, T)$, set $p(t) = \mathcal{L}_v(x(t), \dot{x}(t))$. Then $(x(t), p(t))$ is twice continuously differentiable and solves the Hamilton equations for $t \in (0, T)$.*

Proof. This result follows as in Theorems 6.2.8 and 6.3.3 in Cannarsa and Sinestrari [2004] as we can work in the interior of E due to Assumption 4.3.2. □

The two components of the Hamilton equations take over the role of equations (14) and (15) in Léonard [2013]. The connection between the first component and the Hamilton equations is immediate, whereas for the second component, (15) in Léonard [2013] describes the evolution of f, whereas the second component of the Hamilton equations describes the evolution of $p(t) = \nabla f(x(t))$ along the trajectory of the first variable.

Note that the solution $x(t)$ to the McKean-Vlasov equation is always an entropic interpolation between $x(0)$ and $x(t)$ in time t for any time $t \geq 0$. This corresponds to a solution of the Hamilton equations in which $p(t) = 0$ for all $t \geq 0$.

To study the entropic interpolations, we introduce the adjoint Hamiltonian.

4.3.1 *The adjoint Hamiltonian*

Definition 4.3.4. Let H and S satisfy Assumption 4.1.1. We define the adjoint H^* of H with respect to S for $x \in E^\circ$ by

$$H^*(x, p) = H(x, DS(x) - p).$$

The motivation to call H^* the adjoint Hamiltonian comes from Lemma 4.3.9, where we relate H^* to the reversal of time. Note that $H^{**} = H$.

Definition 4.3.5. If H is a Hamiltonian with entropy S, we say that H is reversible with respect to S if $H^* = H$.

Note that this corresponds to the picture introduced in Lemma 4.1.4. If H is the Hamiltonian corresponding to a sequence of reversible processes, and S is the corresponding entropy of the stationary and reversible measures, then H will be reversible with respect to S.

Remark 4.3.6. Even tough it holds for most one-dimensional examples in this paper that $H = H^*$, a non reversible one-dimensional example is obtained by considering the large deviation behaviour of the average of n independent Levy processes on $\mathbb{R}$ with generator

$$Af(x) = \frac{1}{2}f''(x) - (x+1)f'(x) + f(x+1) - f(x),$$

which corresponds to a Hamiltonian of the form

$$H(x,p) = \frac{1}{2}p^2 - (x+1)p + e^p - 1.$$

For reversible one-dimensional Hamiltonians, we give conditions under which Assumption 4.3.2 is satisfied.

Proposition 4.3.7. *Suppose $E = [a,b]$ and H and S satisfy Assumption 4.1.1 and suppose that:*

(a) $H = H^*$.

(b) $H_p(a,0) > 0$ *and* $H_p(b,0) < 0$.

(c) The maps $x \mapsto \mathcal{L}(x,0)$ and $x \mapsto S(x)$ are decreasing on an open neighbourhood U_a of a and increasing on an open neighbourhood U_b of b.

(d) We have

$$\lim_{x \downarrow a} \frac{\mathcal{L}(x,0) - \mathcal{L}(a,0)}{S(x) - S(a)} = \infty, \quad \lim_{x \uparrow b} \frac{\mathcal{L}(x,0) - \mathcal{L}(b,0)}{S(x) - S(b)} = \infty.$$

Then all entropic interpolations $\{x(t)\}_{0 \leq t \leq T}$ satisfy $x(t) \in E^\circ$ for $t \in (0,T)$.

As the proof of this proposition is independent of the rest of the results, we postpone the proof until Section 4.5.

Lemma 4.3.8. *We have the following properties*

(a) $H^*_p(x,p) = -H_p(x, DS(x) - p)$,

(b) $H^*_{pp}(x,p) = H_p(x, DS(x) - p)$,

(c) $H^*_x(x,p) = H_x(x, DS(x) - p) + H_p(x, DS(x) - p)D^2S(x)$,

(d) $H^*_{px}(x,p) = -H_{px}(x, DS(x) - p) - H_{pp}(x, DS(x) - p)D^2S(x)$.

Proof. These properties are immediately verified using the definition of H^*. □

We now relate the adjoint Hamiltonian to the reversal of time.

Lemma 4.3.9. *Let H and S satisfy Assumption 4.1.1. Fix some time $T > 0$. The curve $(x(t), p(t))_{0<t<T}$ solves the Hamilton equations for H if and only if $(x^*(t), p^*(t))_{0<t<T} := (x(T-t), DS(x(T-t)) - p(T-t))_{0<t<T}$ solves the Hamilton equations for H^*.*

Proof. Let $(x(t), p(t))_{0<t<T}$ solve the Hamilton equations. First note that $H^*_p(x,p) = -H_p(x, DS(x) - p)$ by definition. We look at the derivative of $x^*(t)$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} x^*(t) &= \frac{\mathrm{d}}{\mathrm{d}t} x(T-t) \\
&= -\dot{x}(T-t) \\
&= -H_p(x(T-t), p(T-t)) \\
&= H^*_p(x(T-t), DS(x(T-t)) - p(T-t)) \\
&= H^*_p(x^*(t), p^*(t)).
\end{aligned}$$

Secondly, we consider the derivative of $p^*(t)$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p^*(t) &= \frac{\mathrm{d}}{\mathrm{d}t} p(T-t) \\
&= -\dot{p}(T-t) \\
&= H_x(x(T-t), p(T-t)) \\
&= H^*_x(x(T-t), DS(x(T-t)) - p(T-t)) \\
&= H^*_x(x^*(t), p^*(t)).
\end{aligned}$$

So indeed $(x^*(t), p^*(t))_{0<t<T}$ solve the Hamilton equations for H^*. The second implication of the lemma follows from the first one and the fact that $H^{**} = H$. □

4.3.2 *The evolution of entropy along an entropic interpolations*

Analogous to the definition of $\mathcal{L}$, we define $\mathcal{L}^*$ to be the Lagrangian corresponding to H^*, i.e. for $x \notin \partial E$, we set $\mathcal{L}^*(x,v) = \sup_p \langle p, v\rangle - H^*(x,p)$.

Lemma 4.3.10. *Let H and S satisfy Assumption 4.1.1 and let $\gamma : [0,T] \to E$ be an absolutely continuous trajectory and let t be a time at which γ is differentiable and $\gamma(t) \in E^\circ$. Define the time-backward trajectory $\gamma^*(s) := \gamma(T-s)$. Then we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} S(\gamma(t)) = \mathcal{L}(\gamma(t), \dot{\gamma}(t)) - \mathcal{L}^*(\gamma^*(T-t), \dot{\gamma}^*(T-t)).$$

In particular, if $H = H^*$, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t} S(\gamma(t)) = \mathcal{L}(\gamma(t), \dot{\gamma}(t)) - \mathcal{L}(\gamma(t), -\dot{\gamma}(t)).$$

Proof. Set $p = \mathcal{L}_v(\gamma(t), \dot{\gamma}(t))$ and $p^* = DS(\gamma(t)) - p$. We obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} S(\gamma(t)) = DS(\gamma(t))\dot{\gamma}(t) \\
&= pH_p(\gamma(t), p) - H(\gamma(t), p) \\
&\qquad - \left[(p - DS(\gamma(t)))\, H_p(\gamma(t), p) - H(\gamma(t), p)\right] \\
&= \mathcal{L}(\gamma(t), \dot{\gamma}(t)) - (DS(\gamma(t) - p)\, H_p^*(\gamma(t), DS(\gamma(t)) - p) \\
&\qquad + H^*(\gamma(t), DS(x(t)) - p(t)) \\
&= \mathcal{L}(\gamma(t), \dot{\gamma}(t)) - p^* H_p^*(\gamma^*(T-t), p^*) + H^*(\gamma^*(T-t), p^*) \\
&= \mathcal{L}(\gamma(t), \dot{\gamma}(t)) - \mathcal{L}^*(\gamma^*(T-t), \dot{\gamma^*}(T-t)),
\end{aligned}$$

where we have used in the last line that

$$\dot{\gamma}^*(T-t) = -\dot{\gamma}(t) = -H_p(\gamma(t), p) = H_p^*(\gamma^*(T-t), p^*).$$

□

Because an entropic interpolation $\{x(t)\}_{0 \le t \le T}$ gives rise to a twice continuously differentiable trajectory (x, p) that solves the Hamilton equations, we see that for this trajectory Lemma 4.3.10 holds for all times at which the trajectory is in the interior of E. We use this to study the behaviour of the entropy along the interpolation.

Because the entropy along an arbitrary entropic interpolation is not expected to decrease, we directly study the second derivative of the entropy S along an entropic interpolation $\{x(t)\}_{t \in [0,T]}$ satisfying Assumption 4.3.2. In Lemma 4.3.10, we saw that the first derivative of S contains a part involving $\mathcal{L}$ and a part involving $\mathcal{L}^*$. We first consider the part involving $\mathcal{L}$. Note that for an entropic interpolation $\frac{\mathrm{d}}{\mathrm{d}t} H(x(t), p(t)) = 0$ by the Hamilton equations. For $t \in (0, T)$, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{L}(x(t), \dot{x}(t)) &= \frac{\mathrm{d}}{\mathrm{d}t} \left(\langle p(t), H_p(x(t), p(t)) \rangle - H(x(t), p(t)) \right) \\
&= -\langle H_x(x(t), p(t)), H_p(x(t), p(t)) \rangle \\
&\quad + \langle p(t), H_{px}(x(t), p(t)) H_p(x(t), p(t)) \rangle \\
&\quad - \langle p(t), H_{pp}(x(t), p(t)) H_x(x(t), p(t)) \rangle
\end{aligned} \tag{4.3.2}$$

Set $x^*(t) := x(T-t)$ and $p^*(t) = DS(x(T-t)) - p(T-t)$. For the derivative of the second term, we obtain similarly that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(-\mathcal{L}^*(x^*(T-t), \dot{x}^*(T-t))\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}(T-t)}\mathcal{L}^*(x^*(T-t), \dot{x}^*(T-t)) \\
&= p^*(T-t) H^*_{px}(x^*(T-t), p^*(T-t)) \\
&\quad - p^*(T-t) H_{pp}(x^*(T-t), p^*(T-t)) H^*_x(x^*(T-t), p^*(T-t)) \\
&\quad - H^*_x(x^*(T-t), p^*(T-t)) H^*_p(x^*(T-t), p^*(T-t)).
\end{aligned} \tag{4.3.3}$$

Definition 4.3.11. Let H and S satisfy Assumption 4.1.1. We say that H and H^* satisfy the *α-entropy-convexity inequality* if for all $x \in E$ and $p \in \mathbb{R}^d$, we have

$$\begin{aligned}
&\alpha\left[\langle p, H_p(x,p)\rangle - H(x,p)\right] + \alpha\left[\langle p^*, H^*_p(x,p^*)\rangle - H^*(x,p^*)\right] \\
&\quad \leq \langle p, H_{px}(x,p) H_p(x,p)\rangle - \langle p, H_{pp}(x,p) H_x(x,p)\rangle \\
&\qquad - \langle H_x(x,p), H_p(x,p)\rangle + \langle p^*, H^*_{px}(x,p^*) H^*_p(x,p^*)\rangle \\
&\qquad - \langle p^*, H^*_{pp}(x,p^*) H^*_x(x,p^*)\rangle - \langle H^*_x(x,p^*), H^*_p(x,p^*)\rangle,
\end{aligned}$$

where $p^* = DS(x) - p$. If $H = H^*$, we say that H satisfies the *α-entropy-convexity inequality* if

$$\begin{aligned}
\alpha\left[\langle p, H_p(x,p)\rangle - H(x,p)\right] &\leq \langle p, H_{px}(x,p) H_p(x,p)\rangle \\
&\quad - \langle p, H_{pp}(x,p) H_x(x,p)\rangle - \langle H_x(x,p), H_p(x,p)\rangle,
\end{aligned} \tag{4.3.4}$$

for all $x \in E$ and $p \in \mathbb{R}^d$.

It is immediate that if $H = H^*$ (4.3.4) implies that H and H^* satisfy the *α-entropy-convexity inequality*. The following lemma connects the entropy convexity inequalities with the entropy-information inequality in the case that $H = H^*$.

Lemma 4.3.12. *Let H and S satisfy Assumption 4.1.1. Furthermore, suppose that $H = H^*$. If H satisfies the α-entropy convexity inequality, then H satisfies inequality* (4.2.2) *in Lemma 4.2.3 and if there is only one stationary point x_s for the McKean-Vlasov equation where $S(x_s) = 0$ then (EII)-α is satisfied.*

Proof. Because it holds that $H = H^*$, (4.3.4) is also satisfied for H^*. Taking $p = DS(x)$ and using the identities from Lemma 4.3.8 to rewrite all quantities involving H^* in terms of H yields (4.2.2). □

For $T > 0$ let

$$G_T(s,t) = \begin{cases} \frac{s(T-t)}{T} & \text{if } s \le t \\ \frac{t(T-s)}{T} & \text{if } s \ge t. \end{cases}$$

A direct computation for a function $\phi \in C([0,T])$ that is twice continuously differentiable on $(0,T)$ that

$$\phi(t) = \frac{T-t}{T}\phi(0) + \frac{t}{T}\phi(T) - \int_0^T \ddot{\phi}(s) G_T(s,t) \mathrm{d}s.$$

Combining (4.3.2) and (4.3.3) with the definition of the entropy convexity inequality, we have the following result.

Theorem 4.3.13. *Let H and S satisfy Assumption 4.1.1. Additionally, let H and H^* together with S satisfy the α-entropy convexity inequality. Consider an entropic interpolation $\{x(t)\}_{0 \le t \le T}$ satisfying Assumption 4.3.2. Then we have that*

$$\begin{aligned} S(x(t)) \le \frac{T-t}{T} S(x(0)) + \frac{t}{T} S(x(T)) \\ - \alpha \int_0^T \left[\mathcal{L}(x(s), \dot{x}(s)) + \mathcal{L}^*(x^*(T-s), \dot{x}^*(T-s))\right] G_T(s,t) \mathrm{d}s. \end{aligned}$$

In particular, if $\alpha \ge 0$, we have convexity of the entropy along entropic interpolations satisfying Assumption 4.3.2.

Remark 4.3.14. The integral term on the right hand side of the proposition is somewhat hard to interpret. It would be of interest to see whether this integral term can be bounded from below by some (non-symmetric) distance d. If so, this proposition can serve as starting point for the study of α-convexity of the entropy along entropic interpolations:

$$S(x(t)) \le \frac{T-t}{T} S(x(0)) + \frac{t}{T} S(x(T)) - \alpha \frac{t(T-t)}{2T^2} d^2(x(0), x(T)).$$

This is analogous to the setting of optimal transport, where under bounds on the curvature of the underlying space, the entropy is α convex with respect to the Wasserstein distance along displacement interpolations, see von Renesse and Sturm [2005] or Chapter 16 in Villani [2009]. Also see Erbar and Maas [2012] for related inequalities for the space of measures on a discrete space.

4.4 ENTROPIC INTERPOLATIONS: EXAMPLES

We verify the conditions for Assumptions 4.1.1 and 4.3.2 for three examples in which $H = H^*$ and prove the entropy-convexity inequality.

4.4.1 *Entropic interpolations corresponding to the Ornstein-Uhlenbeck process*

We verify the conditions for Assumptions 4.1.1 and 4.3.2 for three examples.

4.4.2 *The generalized Ornstein-Uhlenbeck process*

An example where we can easily verify an α entropy convexity inequality is for the Hamiltonian corresponding to the generalized Ornstein-Uhlenbeck process. Let $V : \mathbb{R}^d \to [0, \infty)$ be some twice continuously differentiable convex function. Consider the following sequence of processes:

$$\mathrm{d}X_n(t) = -\nabla V(X_n(t))\mathrm{d}t + \frac{1}{\sqrt{n}}\mathrm{d}W(t).$$

The Freidlin-Wentzell large deviation principle of the trajectories of these processes gives an operator

$$H(x,p) = \frac{1}{2}\sum_i p_i^2 - p_i V_i(x),$$

where V_i is the derivative of V in the i-th coordinate. The associated entropy S is given by $S(x) = 2V(x)$

Theorem 4.4.1. *Consider H and S introduced above. Then we have the entropy-convexity inequality with the largest constant $\alpha \in \mathbb{R}$ such that the matrix*

$$\nabla\nabla V - \alpha \mathbb{1}$$

is non-negative definite. Consequently, the conclusions of Theorem 4.3.13 hold for the entropy $S(x) = 2V(x)$.

Clearly, in this setting Assumption 4.3.2 is satisfied. Thus this result holds for all entropic interpolations.

Proof. It is immediate to verify that $H = H^*$, so we only check the entropy-convexity inequality for H. On one hand, we have $pH_p(x,p) - H(x,p) = \frac{1}{2}\sum_i p_i^2$, whereas on the other

$$pH_{px}(x,p)H_p(x,p) - pH_{pp}(x,p)H_x(x,p) - H_p(x,p)H_x(x,p) \\ = \sum_{i,j} p_i V_{i,j}(x) p_j.$$

□

4.4.3 *One-dimensional Wright-Fisher model*

We return to the setting of Example 4.2.6, but now we only consider the one-dimensional example. In particular, we choose our state-space to be equal to $E = [0,1]$ and H is given by

$$H(x,p) = \frac{1}{2}a(x)p^2 - b(x)p,$$

where

$$a(x) = x(1-x), \qquad b(x) = \frac{1}{2}\left(x\mu_1 - (1-x)\mu\right).$$

The entropy S reduces in this setting to

$$S(x) = \mu_1 \log \frac{\mu_1}{\mu(1-x)} + \mu_2 \log \frac{\mu_2}{\mu x}.$$

Lemma 4.4.2. *Let $\mu_1, \mu_2 > 0$ and $\mu = \mu_1 + \mu_2$. Then assumption 4.3.2 is satisfied:*

(a) $H = H^*$,

(b) $H_p(x,0) > 0$ *and* $H_p(x,1) < 0$,

(c) the maps $x \mapsto \mathcal{L}(x,0)$ and $x \mapsto S(x)$ are decreasing on an open neighbourhood U_0 of 0 and increasing on an open neighbourhood U_1 of 1,

(d) we have

$$\lim_{x\downarrow 0} \frac{\mathcal{L}(x,0) - \mathcal{L}(0,0)}{S(x) - S(0)} = \infty$$
$$\lim_{x\uparrow 1} \frac{\mathcal{L}(x,0) - \mathcal{L}(1,0)}{S(x) - S(1)} = \infty.$$

As a consequence, Assumption 4.3.2 is satisfied for this model.

Proof. Using that $DS(x) = \frac{2b(x)}{a(x)}$, it is straightforward to verify that $H = H^*$. We have $H_p(x,0) = -b(x)$, so that $H_p(0,0) > 0$ and $H_p(1,0) < 0$ by the positivity of μ_1 and μ_2.
As $DS(x) = \frac{2b(x)}{a(x)}$, we find that

$$\mathcal{L}(x,0) = -H(x, \frac{1}{2}DS(x)) = \frac{1}{2}\frac{b(x)^2}{a(x)}.$$

Differentiating this with respect to x yields

$$2\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(x,0) = \frac{2a(x)b(x)b'(x) - a'(x)b(x)^2}{a(x)^2}.$$

To verify the third claim for $\mathcal{L}$, we need to know the sign of this derivative. As the denominator is non-negative, we calculate the numerator(recall that $\mu = \mu_1 + \mu_2$):

$$2a(x)b(x)b'(x) - a'(x)b(x)^2 = \frac{1}{4}\left[x\mu - \mu_2\right]\left[(\mu - 2\mu_2)\, x + \mu\right].$$

Thus the claim in (c) for $\mathcal{L}$ follows as this quantity is negative for x close to 0 and positive for x close to 1. The statement for S is clear.

We verify (d) only for the left-hand boundary. The claim follows if we can show that $\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(x,0)$ diverges to $-\infty$ faster than $DS(x)$ diverges to $-\infty$. Note that

$$2\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(x,0) = DS(x)\frac{2a(x)b'(x) - a'(x)b(x)}{a(x)}$$

As $DS(x) < 0$ for x close to 0, we have to show that

$$\frac{2a(x)b'(x) - a'(x)b(x)}{a(x)} = \frac{x\mu_1 - x\mu_2 + \mu_2}{2x(1-x)}$$

diverges to ∞ as $x \downarrow 0$. This, however, is immediate from the $\frac{1}{x}$ term in the denominator and the positive μ_2 term in the numerator. □

In this one-dimensional setting, we improve the constant of the entropy-information inequality of Proposition 4.2.7 and extend it to the entropy-convexity inequality.

Theorem 4.4.3 (Wright-Fisher model with positive mutation rates for two species). *Let $\mu_1, \mu_2 > 0$ and let $\mu = \mu_1 + \mu_2$. Let H be the Hamiltonian given by*

$$H(x, p) = \frac{1}{2}a(x)p^2 - b(x)p,$$

where

$$a(x) = x(1 - x), \qquad b(x) = \frac{x\mu_1 - (1 - x)\mu_2}{2},$$

and where S is given by

$$S(x) = \mu_1 \log \frac{\mu_1}{\mu(1 - x)} + \mu_2 \log \frac{\mu_2}{\mu x}.$$

Then H satisfies the entropy-convexity inequality and the conclusions of Theorem 4.3.13 with respect to S with constant

$$\alpha = \frac{1}{2}\mu + \frac{1}{2}\sqrt{\mu^2 - (\mu_1 - \mu_2)^2} = \frac{1}{2}\mu + \sqrt{\mu_1\mu_2}$$

for all entropic interpolations. Additionally, this constant is optimal for the entropy-convexity inequality.

Proof. To start, we find

$$pH_p(x, p) - H(x, p) = \frac{1}{2}a(x)p^2.$$

A second tedious, but straightforward, calculation yields

$$\begin{aligned} pH_{px}(x, p)H_p(x, p) - pH_{pp}(x, p)H_x(x, p) - H_x(x, p)H_p(x, p) \\ = \left[a(x)b'(x) - \frac{1}{2}a'(x)b(x)\right] p^2. \end{aligned}$$

Using the definitions of a and b, we conclude that we need to find the largest α for which

$$2\alpha(x - x^2) \leq (\mu_1 - \mu_2)x + \mu_2$$

is satisfied for all $x \in [-1, 1]$. As $\mu_1, \mu_2 > 0$, there is at least some $\alpha > 0$ for which this inequality is satisfied. To find the largest $\alpha > 0$ for which this is the case, the minimum of

$$f_\alpha(x) := 2\alpha x^2 + (\mu_1 - \mu_2 - 2\alpha)x + \mu_2$$

for $x \in [0,1]$ should equal 0. As f_α is convex for $\alpha > 0$, the derivative in x of f_α is increasing. As $f_\alpha(0), f_\alpha(1) > 0$, α must be such that $f'_\alpha(0) < 0$ and $f'_\alpha(1) > 0$. We conclude that $2\alpha > |\mu_1 - \mu_2|$. The location of the minimum of f_α is found at

$$x_{min}(\alpha) = \frac{1}{2} - \frac{\mu_1 - \mu_2}{4\alpha}.$$

Evaluating the parabola in its minimum and putting this equal to 0 gives an equation for the value of α:

$$(\mu_1 - \mu_2 - 2\alpha)^2 - 8\alpha\mu_2 = 0$$

which is equivalent to solving

$$4\alpha^2 - 4\mu\alpha + (\mu_1 - \mu_2)^2 = 0.$$

Both zeros are non-negative, but an elementary computation shows that the smallest solution is smaller than $\frac{1}{2}|\mu_1 - \mu_2|$. We conclude that the largest suitable α equals

$$\alpha = \frac{1}{2}\mu + \frac{1}{2}\sqrt{\mu^2 - (\mu_1 - \mu_2)^2} = \frac{1}{2}\mu + \sqrt{\mu_1\mu_2}.$$

□

4.4.4 *Glauber dynamics for the Curie-Weiss model*

The final result of this chapter is the extension of Proposition 4.2.5 to the setting of entropy-convexity.

We introduce two auxiliary functions that turn up in the analysis at various points. Define

$$G_1(x) := \cosh(\beta x) - x\sinh(\beta x),$$
$$G_2(x) := \sinh(\beta x) - x\cosh(\beta x),$$

and note that the Hamiltonian can be rewritten in terms of G_1 and G_2 as

$$H(x,p) = [\cosh(2p) - 1]\, G_1(x) + \sinh(2p) G_2(x).$$

The following lemma follows from the definitions of G_1 and G_2.

Lemma 4.4.4. *For $\beta \in [0,1]$ the functions $G_1, G_2 : [-1,1] \to \mathbb{R}$ have the following properties* *(a)* G_1 *is even, positive, increasing for* $x \leq 0$ *and decreasing for* $x \geq 0$*,*

(b) G_2 *is odd, positive for* $x \leq 0$*, negative for* $x \geq 0$ *and decreasing.*

We start out by verifying Assumption 4.3.2.

Lemma 4.4.5. *Assumption 4.3.2 is satisfied:*

(a) $H = H^*$*,*

(b) $H_p(-1,0) > 0$*,* $H_p(1,0) > 0$*,*

(c) the maps $x \mapsto \mathcal{L}(x,0)$ *and* $x \mapsto S(x)$ *are decreasing on an open neighbourhood* U_{-1} *of* -1 *and increasing on an open neighbourhood* U_1 *of* 1*,*

(d) we have

$$\lim_{x \downarrow -1} \frac{\mathcal{L}(x,0) - \mathcal{L}(-1,0)}{S(x) - S(-1)} = \infty$$
$$\lim_{x \uparrow 1} \frac{\mathcal{L}(x,0) - \mathcal{L}(1,0)}{S(x) - S(1)} = \infty.$$

As a consequence Assumption 4.3.2 is satisfied for this model.

Proof. The (a) follows from a direct computation. For (b), note that $H_p(x,0) = 2G_2(x)$. By Lemma 4.4.4, we find $H_p(-1,0) = G_2(-1) > 0$ and $H_p(1,0) = G_2(1) > 0$.

Claim (c) for S is clear and for $\mathcal{L}$ it is immediate from

$$\begin{aligned}\mathcal{L}(x,0) &= -\inf_p H(x,p) = -H_p\left(x, \frac{1}{2}DS(x)\right) \\ &= -\sqrt{1-x^2} + G_1(x).\end{aligned}$$

The square root has diverging derivative for x close to the boundary, whereas the second term is continuously differentiable on $[-1,1]$, thus we obtain the result.

We only verify (d) for the left boundary. In particular, it is sufficient to show that $\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(x,0)$ diverges to $-\infty$ faster that $DS(x)$ diverges to $-\infty$ as $x \downarrow -1$. In particular, close to -1, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{L}(x,0) = -a(x)\frac{1}{\sqrt{1+x}} + c_1(x), \qquad DS(x) = \frac{1}{2}\log(1+x) + c_2(x),$$

where c_1, c_2 are functions that are bounded on a neighbourhood of -1 and where a is a function close to -1 for x close to -1. The result follows from the asymptotic behaviour of $\frac{-1}{\sqrt{1+x}}$ and $\log(1+x)$ close to -1. □

We proceed with the main result for this model.

Theorem 4.4.6 (Curie-Weiss jump process on two states)**.** *Consider the Hamiltonian* $H : [-1,1] \times \mathbb{R} \to \mathbb{R}$ *defined in Proposition 4.2.5 by*

$$H(x,p) = \frac{1-x}{2} e^{\beta x} \left[e^{2p} - 1\right] + \frac{1+x}{2} e^{-\beta x} \left[e^{-2p} - 1\right]$$

for $\beta \leq 1$*. Then* H *satisfies the entropy-convexity inequality with respect to the relative entropy* $S(x) = \frac{1-x}{2}\log(1-x) + \frac{1+x}{2}\log(1+x) - \frac{1}{2}\beta x^2$ *with constant* $4(1-\beta)$ *and thus the conclusions of Theorem 4.3.13 hold with constant* $4(1-\beta)$ *for all entropic interpolations.*

As noted above $H = H^*$ in this case, so we only have to consider the $4(1-\beta)$ entropy-convexity inequality for H. The proof is based on the basic inequality that $(1-\beta)p \leq (1-\beta)\sinh(p)$ for $p \geq 0$, and thus does not immediately generalize for $\beta > 1$.

Proof of Theorem 4.4.6. Based on the constant obtained in 4.2.5, we will prove

$$\begin{aligned} \alpha\left[\langle p, H_p(x,p)\rangle - H(x,p)\right] &\leq \langle p, H_{px}(x,p)H_p(x,p)\rangle \\ &\quad - \langle p, H_{pp}(x,p)H_x(x,p)\rangle - \langle H_x(x,p), H_p(x,p)\rangle, \end{aligned}$$

for $\alpha = 4(1-\beta)$. Note that for $p = 0$ all terms equal 0. Because the state-space for p is one-dimensional and the problem is symmetric under flipping (x,p) to $(-x,-p)$, it suffices to prove that the derivatives in p, for $p \geq 0$ for every fixed x of the functions on the left and right hand side are ordered with the same constant $\alpha = 4(1-\beta)$:

$$\begin{aligned} \alpha p H_{pp}(x,p) &\leq p H_{pxp}(x,p) H_p(x,p) \\ &\quad - p H_{ppp}(x,p) H_x(x,p) - 2 H_{pp}(x,p) H_x(x,p). \end{aligned}$$

Our argument will be based on the basic inequality that $2p \leq \sinh(2p)$ for $p \geq 0$. In particular, as $H_{pp}(x,p) > 0$ by the strict convexity of H in the momentum variable this implies that

$$4(1-\beta) p H_{pp}(x,p) \leq 2(1-\beta)\sinh(2p) H_{pp}(x,p).$$

Thus, it suffices to prove for $p \geq 0$ and all x that

$$\begin{aligned} 0 &\leq p H_{pxp}(x,p) H_p(x,p) - p H_{ppp}(x,p) H_x(x,p) \\ &\quad - 2 H_{pp}(x,p)\left(H_x(x,p) + (1-\beta)\sinh(2p)\right). \qquad (4.4.1) \end{aligned}$$

To do this, we study the Hamiltonian in terms of G_1 and G_2 as

$$H(x,p) = [\cosh(2p) - 1]\, G_1(x) + \sinh(2p) G_2(x).$$

This representation immediately yields that

$$H_{pp}(x,p) = 4\cosh(2p) G_1(x) + 4\sinh(2p) G_2(x),$$

which in turn implies that

$$H_{pxp}(x,p) = 4H_x(x,p) + 4G_1'(x).$$

Because $H_{ppp}(x,p) = 4H_p(x,p)$, we conclude that the first two terms of the right hand side of (4.4.1) equal

$$\begin{aligned} &pH_{pxp}(x,p)H_p(x,p) - pH_{ppp}(x,p)H_x(x,p) \\ &\qquad\qquad\qquad\qquad = 4pG_1'(x)H_p(x,p). \end{aligned}$$

The last term of (4.4.1) can be rewritten as

$$\begin{aligned} &-2H_{pp}(x,p)\left((\cosh(2p)-1)G_1'(x) + \sinh(2p)\left(G_2'(x) + 1 - \beta\right)\right) \\ &= -2H_{pp}(x,p)(\cosh(2p)-1)G_1'(x) \\ &\qquad + \sinh(2p)H_{pp}(x,p)\left[(1-\beta)(\cosh(\beta x) - 1) + \beta x \sinh(\beta x)\right]. \end{aligned}$$

Rewriting these last two equations, we have to prove for all x and $p \geq 0$ that

$$\begin{aligned} 0 \leq\, & 8\left[p\sinh(2p) - \cosh^2(2p) + \cosh(2p)\right] G_1(x)G_1'(x) \\ &+ 8\left[p\cosh(2p) + \sinh(2p)\cosh(2p) - \sinh(2p)\right] G_1'(x)G_2(x) \\ &+ 2H_{pp}(x,p)\sinh(2p)\left[(1-\beta)(\cosh(\beta x) - 1) + \beta x\sinh(\beta x)\right]. \end{aligned} \tag{4.4.2}$$

This will be proven in two steps, first we prove this inequality for $x \geq 0$ and all $p \geq 0$, and afterwards we consider the case that $x \leq 0$ and $p \geq 0$.

Case 1: $x \geq 0$. It can immediately be seen that the third line in (4.4.2) is bounded below by 0. For the first line, we show that

$$p \mapsto p\sinh(2p) - \cosh^2(2p) + \cosh(2p)$$

is non-positive for $p \geq 0$. First note that $2p \leq \sinh(2p)$, and thus

$$\begin{aligned}
&p\sinh(2p) - \cosh^2(2p) + \cosh(2p)\\
&\quad\leq \frac{1}{2}\sinh^2(2p) - \cosh^2(2p) + \cosh(2p)\\
&\quad= -\frac{1}{2} - \frac{1}{2}\cosh^2(2p) + \cosh(2p)\\
&\quad= -\frac{1}{2}(\cosh(2p) - 1)^2\\
&\quad\leq 0.
\end{aligned}$$

$G_1(x)G_1'(x) \leq 0$ for $x \geq 0$ by Lemma 4.4.4, which implies that also the first term of (4.4.2) is non-negative.

We proceed with the second term. The map

$$p\cosh(2p) + \sinh(2p)\cosh(2p) - \sinh(2p)$$

is non-negative for $p \geq 0$ as $\cosh(2p) \geq 1$. Additionally, by Lemma 4.4.4, the product $G_1'(x)G_2(x)$ is non-negative.

We conclude that (4.4.2) holds for $p \geq 0$ and $x \geq 0$.

Case 2: $x \leq 0$. The non-negativity for lines 2 and 3 of the right-hand side in (4.4.2) still hold, but we need to show that these lines compensate line 1, that is now negative due to the positivity of the product $G_1(x)G_1'(x)$. In particular, we will show that line three of the right hand side of (4.4.2) compensates the first term. Note that

$$0 \geq (1-\beta)(\cosh(\beta x)-1)+\beta x\sinh(\beta x) = -G_2'(x)-(1-\beta), \quad (4.4.3)$$

so that the third term of (4.4.2) equals

$$\begin{aligned}
&2H_{pp}(x,p)\sinh(2p)\left[(1-\beta)(\cosh(\beta x)-1)+\beta x\sinh(\beta x)\right]\\
&\quad= -2H_{pp}(x,p)\sinh(2p)\left[G_2'(x)+(1-\beta)\right]\\
&\quad= -8\cosh(2p)\sinh(2p)G_1(x)\left[G_2'(x)+(1-\beta)\right]\\
&\qquad - 8\sinh^2(2p)G_2(x)\left[G_2'(x)+(1-\beta)\right].
\end{aligned}$$

By equation (4.4.3) and Lemma 4.4.4 the term in the last line is non-negative if $x \leq 0$. Thus, we can use the term in line three to compensate the first term in (4.4.2). In particular, we have to show that

$$\begin{aligned}
0 \leq &-8\cosh(2p)\sinh(2p)G_1(x)\left[G_2'(x)+(1-\beta)\right]\\
&\quad + 8\left[p\sinh(2p)-\cosh^2(2p)+\cosh(2p)\right]G_1(x)G_1'(x).
\end{aligned}$$

for $x \leq 0$ and $p \geq 0$. We divide by $8G_1(x) > 0$ and show

$$\begin{aligned} 0 \leq &- \cosh(2p)\sinh(2p)\left[G_2'(x) + (1-\beta)\right] \\ &\qquad + \left[p\sinh(2p) - \cosh^2(2p) + \cosh(2p)\right] G_1'(x). \end{aligned}$$

Below, we will prove that $G_1'(x) + G_2'(x) + (1-\beta) \leq 0$ for $x \leq 0$. Using this inequality, we find

$$\begin{aligned} &- \cosh(2p)\sinh(2p)\left[G_2'(x) + (1-\beta)\right] \\ &\qquad + \left[p\sinh(2p) - \cosh^2(2p) + \cosh(2p)\right] G_1'(x) \\ &\geq \left[\cosh(2p)\sinh(2p) + p\sinh(2p) - \cosh(2p)(\cosh(2p) - 1)\right] G_1'(x). \end{aligned}$$

Because $G_1'(x) \geq 0$ for $x \leq 0$ and $p\sinh(2p) \geq 0$ and $\sinh(2p) \geq \cosh(2p) - 1$, we find that this term is non-negative.

We are left to prove that $G_1'(x) + G_2'(x) + (1-\beta) \leq 0$ for $x \leq 0$. First, we calculate

$$\begin{aligned} G_1'(x) &= (\beta - 1)\sinh(\beta x) - \beta x\cosh(\beta x), \\ G_2'(x) &= (\beta - 1)\cosh(\beta x) - \beta x\sinh(\beta x). \end{aligned}$$

We conclude that

$$\begin{aligned} &G_1'(x) + G_2'(x) + (1-\beta) \\ &\ = (\beta - 1)\left[\sinh(\beta x) + \cosh(\beta x) - 1\right] \\ &\qquad - \beta x\left[\cosh(\beta x) + \sinh(\beta x)\right], \end{aligned}$$

which yields that $G_1'(x) + G_2'(x) + (1-\beta) \leq 0$ for $x \leq 0$.

We conclude that (4.4.2) holds for all $x \in [-1, 1]$ and $p \geq 0$. This implies (4.4.1) and thus the entropy-convexity inequality with constant $4(1-\beta)$. $\square$

4.5 ENTROPIC INTERPOLATIONS REMAIN IN THE INTERIOR

To conclude this chapter, we prove Proposition 4.3.7. We need some additional results.

To prove that an interpolation $\{x(s)\}_{s\in[0,t]}$ from a to b remains in the interior, we argue by contradiction. Suppose that x that hits the boundary for some $s \in (0, t)$, then we find a cheaper trajectory that also connects a to b. To do this, we use the evolution of the entropy S along the interpolation.

We start out with a technical regularity result.

Lemma 4.5.1. *Let* $\gamma : [0, t] \to [-1, 1]$ *be absolutely continuous and such that*

$$\int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s < \infty.$$

Then $s \mapsto S(\gamma(s))$ *is absolutely continuous.*

Note that this result is non-trivial. A result of Fichtenholz, see Exercise 5.8.61 in Bogachev [2007], shows that if $DS(x) \to \infty$ or $DS(x) \to \infty$ for x close to the boundary, there exists an absolutely continuous trajectory γ taking values in $[a, b]$ such that $s \mapsto S(\gamma(s))$ is not absolutely continuous.

Proof. The proof is somewhat technical and needs the definition of Lusin's property (N). We say that a function $F : (X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ between to measure spaces satisfies (N) if $\nu(F(A)) = 0$ for all $A \in \mathcal{A}$ with $\mu(A) = 0$.

Pick γ that satisfies the assumptions of the lemma. Because γ and S are continuous, $s \mapsto S(\gamma(t))$ is continuous. γ is absolutely continuous, so it satisfies property (N). As S is continuously differentiable on (a, b) it is absolutely continuous on (a, b). Because S is decreasing in a neighbourhood of a and increasing in a neighbourhood of b, the absolute continuity of S on $[a, b]$ follows by the monotone convergence theorem. We conclude that S satisfies (N). Clearly the composition $s \mapsto S(\gamma(s))$ of functions that satisfy (N) also satisfies (N).

To prove that $s \mapsto S(\gamma(s))$ is absolutely continuous, we use Exercise 5.8.57 of Bogachev [2007] that states that a continuous function $f : [\alpha, \beta] \to \mathbb{R}$ with property (N) is absolutely continuous if there exists a Lebesgue integrable function g such that $f'(x) \leq g(x)$ at almost every point where $f'(x)$ exists.

We show that we can find such a function g for $f(x) := S(\gamma(s))$, using the assumption that the Lagrangian cost of the trajectory is finite.

First of all, γ is differentiable at almost every time. Thus, for almost every time s for which $\gamma(s) \in (a, b)$, the map f is differentiable. For such s, we have by Lemma 4.3.10 that

$$\frac{\mathrm{d}}{\mathrm{d}s} S(\gamma(s)) \leq \mathcal{L}(\gamma(s), \dot{\gamma}(s)).$$

Because S has its maxima at the boundary, a time s for which $\gamma(x) \in \{-1, 1\}$ and f is differentiable, must satisfy $f'(s) = 0 \leq \mathcal{L}(\gamma(s), \dot{\gamma}(s))$.

Thus, for almost every time s for which $s \mapsto S(\gamma(s))$, we have $\frac{\mathrm{d}}{\mathrm{d}s} S(\gamma(s)) \leq \mathcal{L}(\gamma(s), \dot{\gamma}(s))$. By the assumption of the lemma and Exercise 5.8.57 of Bogachev [2007], we conclude that $s \mapsto S(\gamma(s))$ is absolutely continuous. □

Our second auxiliary result is a decomposition for $\mathcal{L}$, which is a result also obtained in Mielke et al. [2014]. The decomposition there is given in terms of Ψ, Ψ^* and the decomposition is used to interpret the solution of the McKean-Vlasov equation as the flow that optimizes an entropy-dissipation inequality. Here we give a different interpretation of this decomposition. We first introduce a tilted Hamiltonian $H[x] : E^\circ \times \mathbb{R}^d \to \mathbb{R}$ by

$$H[x](y,p) = H(y, p + \frac{1}{2} DS(x)) - H(y, \frac{1}{2} DS(x)). \tag{4.5.1}$$

It follows that x is a stationary point of the McKean-Vlasov dynamics associated to $H[x]$, i.e. $H[x]_p(x,0) = 0$.

It can then be shown that $\mathcal{L}(x,v)$ can be decomposed into a cost for making x the stationary point, the cost for having speed v under the tilted dynamics and a correction term: the increase of entropy along the flow.

Lemma 4.5.2. *For $x \in (a,b)$ and $v \in \mathbb{R}$, we have the decomposition*

$$\mathcal{L}(x,v) = \mathcal{L}(x,0) + \mathcal{L}[x](x,v) + \frac{1}{2}\langle DS(x), v\rangle,$$

where $\mathcal{L}[x](y,v)$ is defined as the Legendre transform of $p \mapsto H[x](y,p)$, as defined in (4.5.1).

Furthermore, for $x \in (a,b)$, we have $\mathcal{L}[x](x,v) = \mathcal{L}[x](x,-v)$. Finally, for any absolutely continuous trajectory $\gamma : [0,t] \to [a,b]$ that has finite Lagrangian cost, we have

$$\int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) - \mathcal{L}(\gamma(s), -\dot{\gamma}(s)) \mathrm{d}s = S(\gamma(t)) - S(\gamma(0)).$$

Proof. Because H is smooth and has super-linear growth in p for $x \in (-1,1)$, we find that $\mathcal{L}(x,v) = \sup_p \{pv - H(x,p) = p^* v - H(x,p^*)\}$, where $p^* = L_v(x,v)$. Thus, rewriting, we find

$$\begin{aligned}
\mathcal{L}(x,v) &= \langle p^*, v\rangle - H(x,p^*) \\
&= \langle p^* - \frac{1}{2}DS(x), v\rangle - H[x]\left(x, p^* - \frac{1}{2}DS(x)\right) \\
&\quad - H(x, \frac{1}{2}DS(x)) + \frac{1}{2}\langle DS(x), v\rangle.
\end{aligned}$$

Now note that the supremum over p in $\mathcal{L}[x](x,v) = \sup_p \{pv - H[x](x,p)\}$ is attained at $p^* - \frac{1}{2}DS(x)$. Also, note that $\mathcal{L}(x,0) = H(x, \frac{1}{2}DS(x))$. Hence, the first claim of the lemma is proven.

For the second claim, note that $s \mapsto S(\gamma(s))$ is absolutely continuous by Lemma 4.5.1. For times s that $\gamma(s) \in (a,b)$, the derivative of $s \mapsto S(\gamma(s))$ is given by Lemma 4.3.10, using that $H = H^*$. For almost all times s such that $\gamma(s) \notin (a,b)$, $s \mapsto S(\gamma(s))$ is differentiable as the map is absolutely continuous. For these times the derivative must be 0 as S has its (strict) maxima on the boundary. Clearly $\mathcal{L}(\gamma(s),0) = \mathcal{L}(\gamma(s),-0)$. Thus, the second claim follows by integration. □

We conclude this section by proving that all entropic interpolations remain in the interior of E.

Proof of Proposition 4.3.7. Fix $t > 0$ and $\alpha, \beta \in E$. Let γ be an optimal trajectory such that $\gamma(0) = \alpha$ to $\gamma(t) = \beta$.

The strategy of the proof is as follows. We argue by contradiction. First we assume that there exists an interval $[t_0, t_1] \subseteq [0,t]$ on which the trajectory is on the boundary of E. Then, we construct a new trajectory, which is on the boundary for the times t_0 and t_1, but not for $s \in (t_0, t_1)$, which has lower cost. This contradicts the assumption that our trajectory was optimal. As a second step, we assume there is an isolated time $t^* \in (0,t)$ for which the trajectory is on the boundary. In this setting, we construct a compatible trajectory that remains in the interior for an interval $(t_-, t_+) \ni t^*$ with lower cost, again contradicting the assumption that our trajectory was optimal.

These two contradictions show that an optimal trajectory can not be on the boundary for a time $s \in (0,t)$.

First assume that there exists an interval $[t_0, t_1]$, $t_0 \neq t_1$ such that the optimal trajectory γ satisfies $\gamma(s) = a$ for $s \in [t_0, t_1]$. The argument for the boundary b is similar. We construct γ^* that has a lower cost to obtain a contradiction. Fix some $\varepsilon > 0$ small enough such that $\varepsilon < \frac{1}{2}(t_1 - t_0)$ and such that the solution of $\dot{x} = H_p(x,0)$ started at $x_0 = a$ does not leave U_a. Note any solution $\{x(t)\}_{t \geq 0}$ of the McKean-Vlasov equation $\dot{x} = H_p(x,0)$ satisfies $x(t) \in (a,b)$ for $t > 0$ by assumption (b) of the Proposition.

Define $\gamma_\varepsilon : [t_0, t_1] \to E$ as $\gamma_\varepsilon(0) = a$, $\dot{\gamma}_\varepsilon(s) = H_p(\gamma_\varepsilon(s), 0)$ for $s \leq \varepsilon$ and $\gamma_\varepsilon(s) = \gamma_\varepsilon(\varepsilon) =: z(\varepsilon)$ for $s \in [\varepsilon, \frac{1}{2}]$. Additionally, we set γ_ε to be the time-reversed trajectory on the second half of the interval: $\gamma_\varepsilon(t_0+s) = \gamma_\varepsilon(t_1-s)$.

Splitting $[0,t]$ into the two symmetric parts, applying the final part of Lemma 4.5.2 on the non-stationary part of γ, we find

$$\int_0^t \mathcal{L}(\gamma_\varepsilon(s),\dot\gamma_\varepsilon(s))\mathrm{d}s = 2\int_0^\varepsilon \mathcal{L}(\gamma_\varepsilon(s),\dot\gamma_\varepsilon(s))\mathrm{d}s \\ + (t_2-t_1-2\varepsilon)\mathcal{L}(z(\varepsilon),0) + (S(-1)-S(z(\varepsilon)))\,.$$

Now the first term on the right-hand side is 0 as $\dot\gamma_\varepsilon(s) = H_p(\gamma_\varepsilon(s),0)$ for $s\le\varepsilon$, thus

$$\int_0^t \mathcal{L}(\gamma_\varepsilon(s),\dot\gamma_\varepsilon(s)) - \mathcal{L}(\gamma(s),\dot\gamma(s))\mathrm{d}s \\ = (t_2-t_1)\left(\mathcal{L}(z(\varepsilon),0)-\mathcal{L}(a,0)\right) - 2\varepsilon\mathcal{L}(z(\varepsilon,0)) + S(a) - S(z(\varepsilon)).$$

The middle term on the right hand is non-negative. That the first and the third term combined are non-negative for small ε follows from assumption (d) of the proposition.

Thus, we have contradicted the assumption that there exists an interval $[t_0,t_1]$, $t_0\neq t_1$ such that γ satisfies $\gamma(s)=a$ for $s\in[t_0,t_1]$.

Now suppose there exists $t^*\in(0,t)$ such that $\gamma(t^*)=a$. We show that this leads to a contradiction. Fix $z>a$. Then the set $B_z := \gamma^{-1}([a,z))$ is open in $[0,T]$. Because an open set in $\mathbb{R}$ is the countable disjoint union of open intervals by the Lindelöf lemma, there are three possibilities:

(a) $t^*\in(t_-,t_+)$, $(t_-,t_+)\subseteq B_z$, $t_-,t_+\notin B_z$,
(b) $t^*\in(t_-,t]$, $(t_-,t]\subseteq B_z$, $t_-\notin B_z$,
(c) $t^*\in[0,t_+)$, $[0,t_+)\subseteq B_z$, $t_+\notin B_z$.

Clearly, if (b) happens for all $z>a$, then $[t^*,t]\subseteq\gamma^{-1}(a)$ which contradicts the conclusion of the first part of the proof. A similar contradiction occurs for (c). Note that in case of (a), we have that $\gamma(t_-)=\gamma(t_+)=z$ by the continuity of γ.

Thus, we can choose $z>a$ close enough to a, such that (a) is satisfied and such that $[a,z)\subseteq U_a$. Again we construct a cheaper trajectory γ^*. As noted above, there are $0<t_-<t^*<t_+<t$ such that we have

$\gamma(t_-) = \gamma(t_+) = z$ and $a \le \gamma(s) \le z$ for $t_- \le s \le t_+$. Consider the trajectory

$$\gamma_z(s) = \begin{cases} \gamma(s) & \text{for } s \notin [t_-, t_+], \\ z & \text{for } s \in [t_-, t_+]. \end{cases}$$

Using Lemma 4.5.2, integrating over time in $[t_-, t_+]$, we find

$$\int_{t_-}^{t_+} \mathcal{L}(\gamma_z(s), \dot{\gamma}_z(s))\mathrm{d}s = \int_{t_-}^{t_+} \mathcal{L}(z, 0)\mathrm{d}s \tag{4.5.2}$$

and

$$\int_{t_-}^{t_+} \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s = \int_{t_-}^{t_+} \mathcal{L}[\gamma(s)](\gamma(s), \dot{\gamma}(s)) + \mathcal{L}(\gamma(s), 0)\mathrm{d}s.$$

Because the first term of the integrand on the right is non-negative, and second term in the integrand is bounded from below by the integrand on the right in (4.5.2) by condition (c) of the proposition, we find that γ_z has a lower cost than γ, contradicting the assumption that γ was optimal.

We conclude that an optimal trajectory can only attain a boundary point at its initial or final time. □

5 GIBBS-NON-GIBBS TRANSITIONS

The results in this chapter are work in progress jointly with Christof Külske and Frank Redig.

5.1 LARGE DEVIATIONS FOR INTERACTING DIFFUSION PROCESSES

In this chapter, we consider n mean-field interacting diffusion processes $(B_1(t), \dots, B_n(t))$ on $\mathbb{R}$. Define the mean $X_n(t) = \frac{1}{n}\sum_{i\leq n} B^i(t)$ and consider a potential $V_a : \mathbb{R} \to \mathbb{R}$ defined by $V_a(x) = \frac{1}{4}x^4 - \frac{1}{2}ax^2 + C_a$, where C_a is a constant such that the minimum of V_a is 0, with gradient $F_a : \mathbb{R} \to \mathbb{R}$ given by $F_a(x) = x^3 - ax$. Then the evolution of the processes is given by

$$\mathrm{d}B_i(t) = -\frac{1}{2}F_a(X_n(t)) + \mathrm{d}W^i(t),$$

where W^i are independent standard Brownian motions. As a consequence, the evolution of the mean is given by

$$\mathrm{d}X_n(t) = -\frac{1}{2}F_a(X_n(t)) + \frac{1}{\sqrt{n}}\mathrm{d}W(t),$$

where W is a standard Brownian motion. Note that the solution to this stochastic differential equation exists due to Theorem 3.21 and Corollary 3.39 in Pardoux and Răşcanu [2014]. By varying the constant a, the potential V_a changes from a single-well for $a \leq 0$ to a double-well for $a > 0$. Motivated by the analogies of the Hamiltonian flow corresponding to the large deviations of this problem and the Hamiltonian flow for Glauber dynamics, we can view the $a \leq 0$ as a high temperature case and $a > 0$ as a low temperature case.

The non-Lipschitz character of F_a poses problems when directly applying the large deviation results in the literature. The drift $-\frac{1}{2}F_a(x)$, however, is one-sided Lipschitz:

$$(x-y)\left(-\frac{1}{2}F_a(x) - \frac{1}{2}F_a(y)\right) \leq (a \vee 0)(x-y)^2.$$

Thus, we can prove the following theorem as in Deng et al. [2011].

Theorem 5.1.1. *If $X_n(0)$ satisfies the large deviation principle with rate function I_0, then the sequence of the trajectories $\{X_n(t)\}_{t\geq 0}$ satisfies the large deviation principle on $D_{\mathbb{R}}(\mathbb{R}^+)$ with good rate function*

$$I(\gamma) = \begin{cases} I_0(\gamma) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{L}(x, v) = \frac{1}{2}|v + \frac{1}{2}F_a(x)|^2$.

Sketch of proof. The processes $X_n(t)$ have generators

$$A_n f(x) = \frac{1}{2n} f''(x) - \frac{1}{2} F_a(x) f'(x),$$

with a domain that includes the compactly supported twice continuously differentiable functions $C_c^2(\mathbb{R})$. Define $H_n f = n^{-1} e^{-nf} A_n e^{nf}$, and note that $H_n f(x) = \frac{1}{2n} f''(x) + \frac{1}{2}(f'(x))^2 - \frac{1}{2}F_a(x) f'(x)$ for $f \in C_c^2(\mathbb{R}^d)$. We define $Hf(x) = \frac{1}{2}(f'(x))^2 - \frac{1}{2}F_a(x)f'(x)$ and immediately obtain that for $f \in C_c^2(\mathbb{R}^d)$ we have $\lim_{n\to\infty} \|H_n f - Hf\| = 0$.

Based on Proposition 3.3.2 in Chapter 3, our goal is to prove the comparison principle for viscosity subsolutions u and supersolutions v of $f - \lambda Hf = h$ for fixed $\lambda > 0$ and $h \in C_b(\mathbb{R})$.

Consider the good distance function $\Psi(x, y) = \frac{1}{2}|x - y|^2$ and suppose there exist x_α, y_α that satisfy

$$u(x_\alpha) - v(y_\alpha) - \alpha\Psi(x_\alpha, y_\alpha) = \sup_{x,y\in E} \{u(x) - v(y) - \alpha\Psi(x, y)\}.$$

Writing as before $Hf(x) = H(x, f'(x))$, where $H(x, p) = \frac{1}{2}p^2 - \frac{1}{2}pF_a(x)$, we would need to prove that

$$\liminf_{\alpha\to\infty} H\left(x_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) - H\left(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) \leq 0.$$

But this is immediate as

$$\begin{aligned} &H\left(x_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) - H\left(y_\alpha, \alpha(\nabla\Psi(\cdot, y_\alpha))(x_\alpha)\right) \\ &\qquad \leq (x_\alpha - y_\alpha)(-\frac{1}{2}F_a(x_\alpha) + \frac{1}{2}F_a(y_\alpha)) \leq (a \vee 0)\alpha\Psi(x_\alpha, y_\alpha) \end{aligned}$$

which converges to 0 by Lemma 3.3.1.

The problem with the argument above is that due to non-compactness of $\mathbb{R}$ it is not clear that the points x_α, y_α exist. Thus, we need to make an adjustment to the test-function Ψ that makes sure the optima are attained. For sub-solutions, we need to add a function that grows to infinity for $|x| \to \infty$ and for super-solutions, we need to add a function that grows to $-\infty$. This can be carried out as in Deng et al. [2011]. Verification of the conditions in Chapter 8 of Feng and Kurtz [2006] follows as in Deng et al. [2011] or as in Chapter 3. □

5.2 OPTIMAL TRAJECTORIES

In the study of Gibbs-non-Gibbs transitions, it was shown in Ermolaev and Külske [2010], den Hollander et al. [2015] that a bad magnetization $\alpha \in \mathbb{R}$ corresponds to the non-uniqueness of optimal trajectories for

$$I_t(\alpha) = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(t)=\alpha}} \left\{ I_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s \right\}, \tag{5.2.1}$$

which in turn is equivalent to non-differentiability of I_t at α. In this chapter, we will not focus on the probabilistic aspects of the Gibbs-non-Gibbs transitions, but we will use the Hamilton equations to obtain concrete information on the optimal solutions of (5.2.1), and as a consequence on the occurrence of bad magnetizations.

Define the semigroup $T_t : C_b(\mathbb{R}) \to C_b(\mathbb{R})$ by

$$T_t f(x) = \inf_{\substack{\xi \in \mathcal{AC} \\ \xi(t)=x}} \left\{ f(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s))\mathrm{d}s \right\}.$$

Fix some continuously differentiable function u_0. We say that $\xi \in (CV)_{t,x}$ if the curve is an optimiser for the variational problem:

$$u(t,x) := T_t u_0(x) = u_0(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s))\mathrm{d}s. \tag{5.2.2}$$

In particular, note that if u_0 is the rate function of $X_n(0)$, then $u(t,\cdot)$ is the rate function of $X_n(t)$ by the contraction principle.

We say that a absolutely continuous curve ξ is an extremal to $(CV)_{t,x}$ if for any absolutely continuous perturbation ρ that satisfies $\rho(t) = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} |_{\varepsilon=0}\, u_0(\xi(0) + \varepsilon\rho(0)) + \int_0^t \mathcal{L}(\xi(s) + \varepsilon\rho(s), \dot{\xi}(s) + \varepsilon\dot{\rho}(s))\mathrm{d}s = 0.$$

By Theorems 6.2.4 and 6.2.8 in Cannarsa and Sinestrari [2004], an extremal ξ for $(CV)_{t,x}$ is twice continuously differentiable and solves the Euler-Lagrange equation. In other words, $t \mapsto \mathcal{L}_v(\xi(t), \dot{\xi}(t))$ is absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{L}_v(\xi(t), \dot{\xi}(t)) = \mathcal{L}_x(\xi(t), \dot{\xi}(t)).$$

Finally, ξ also satisfies the transversality condition

$$\mathcal{L}_v(\xi(0), \dot{\xi}(0)) = Du_0(\xi(0)).$$

Remark 5.2.1. Note that a function θ as in Cannarsa and Sinestrari [2004] that works globally can not be found due to the unboundedness of F_a. Note however, that we can find such a function locally. The global property, however, is only necessary to construct optimizers in Theorem 6.1.2. In our context, however, optimizers exist due to the goodness of the rate function, which we obtained via different methods. Functions θ that satisfy all bounds locally suffice for all other theorems from Cannarsa and Sinestrari [2004] by Remark 6.2.7.

Consider the Hamilton equations for the Hamiltonian $H(x,p) = \frac{1}{2}p^2 - \frac{1}{2}pF_a(x)$:

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} H_p(x,p) \\ -H_x(x,p) \end{bmatrix} = \begin{bmatrix} p - \frac{1}{2}F_a(x) \\ \frac{1}{2}pF_a'(x) \end{bmatrix}. \tag{5.2.3}$$

The Euler-Lagrange equations can be recast in terms of Hamilton's equations.

Theorem 5.2.2 (Theorem 6.3.3 Cannarsa and Sinestrari [2004])**.** *Let $u_0 \in C^1(\mathbb{R})$ and let $\xi \in C^2$ be an extremal of $(CV)_{t,x}$ and set*

$$\eta(s) := \mathcal{L}_v(\xi(s), \dot{\xi}(s)), \quad s \in [0,t].$$

Then $\eta(0) = Du_0(\xi(0))$ and the pair (ξ, η) satisfies the Hamilton equations (5.2.3).

Conversely, any C^2 solution (ξ, η) of the Hamilton equations that satisfy $\xi(t) = x$ and $\eta(0) = Du_0(\xi(0))$, yields an extremal $\xi(t)$ for $(CV)_{t,x}$.

To rigorously study the solutions of the Hamilton equations and the connection to the gradient of $x \mapsto u(t,x)$, we introduce some definitions. These definitions and results are taken from Chapters 2 and 3 in Cannarsa and Sinestrari [2004].

5.2.1 *Semi-concavity and generalized differentials*

Let Ω be some convex subset of $\mathbb{R}^d$. We say that a function $f \in C(\Omega)$ is *semi-concave* if there exists $c \geq 0$ such that the function $f(x) - \frac{c}{2}|x|^2$ is concave. f is *locally semi-concave* if f is semi-concave on every bounded, closed and convex subset of Ω. A locally semi-concave function is locally Lipschitz continuous by Theorem 2.1.7 in Cannarsa and Sinestrari [2004].

For any $x \in \Omega$, denote by

$$D^+u(x) = \left\{ p \in \mathbb{R}^d \,\middle|\, \limsup_{y \to x} \frac{u(y) - u(x) - p(y-x)}{|y-x|} \leq 0 \right\},$$
$$D^-u(x) = \left\{ p \in \mathbb{R}^d \,\middle|\, \liminf_{y \to x} \frac{u(y) - u(x) - p(y-x)}{|y-x|} \geq 0 \right\},$$

the (Fréchet) *superdifferential* and *subdifferential.* Both sets are closed and convex, but are possibly empty. u is differentiable at x if and only if $D^+u(x)$ and $D^-u(x)$ are both non-empty, and in this case we have $\{Du(x)\} = D^+u(x) = D^-u(x)$, see Proposition 3.1.5 in Cannarsa and Sinestrari [2004]

Let f be locally Lipschitz. A vector $p \in \mathbb{R}^d$ is called a *reachable gradient* of f at x if there exists a sequence $x_n \subseteq \mathbb{R} \setminus \{x\}$ such that f is differentiable at x_n for all n and $x_n \to x$ and $Df(x_n) \to p$. The set of all reachable gradients is denoted by $D^*u(x)$.

Let f be a locally semi-concave function. Then $D^+f(x)$ is a singleton if and only if f is differentiable at x by Proposition 3.3.4 in Cannarsa and Sinestrari [2004]. For a locally semi-concave function we have $D^+f(x) = coD^*f(x)$, where coA denotes the closed convex hull of $A \subseteq \mathbb{R}$, see Theorem 3.3.6 in Cannarsa and Sinestrari [2004].

Theorem 5.2.3 (Theorem 6.4.3 and Corollary 6.4.4 Cannarsa and Sinestrari [2004])**.** *Suppose $u_0 \in C(\mathbb{R})$. Then for any $t > 0$, the function $x \mapsto u(t,x)$ defined in* (5.2.2) *is locally semi-concave on $\mathbb{R}$. Also, the function $(t,x) \mapsto u(t,x)$ is locally semi-concave on $(0,\infty) \times \mathbb{R}$*

The optimization problem in the value function can be restricted to solutions of the Hamilton equations.

Theorem 5.2.4 (Theorem 6.4.6 Cannarsa and Sinestrari [2004])**.** *Let $u_0 \in C^1(\mathbb{R})$. For $\xi \in \mathcal{AC}$ set $\eta(s) = \mathcal{L}_v(\xi(s), \dot{\xi}(s))$ and denote*

$\mathcal{H}_{t,x} := \{\xi \in \mathcal{AC} \mid \xi(t) = x,\ (\xi(s), \eta(s))$ *solves* (5.2.3), $\eta(0) \in Du_0(\xi(0))\}$

The function $x \mapsto u(t,x)$ is given by the minimal selection

$$u(t,x) = \inf_{\xi \in \mathcal{H}_{t,x}} \left\{ u_0(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s)) \mathrm{d}s \right\}.$$

Theorem 5.2.5 (Theorem 6.4.8 Cannarsa and Sinestrari [2004]). *Consider (x,t). Let ξ be a minimizer for $(CV)_{t,x}$ and denote $\eta(s) = L_v(\xi(s), \dot{\xi}(s))$ for all s. Then we have*

$$\begin{aligned} &\eta(t) \in D^+ u(t, \xi(t)) \\ &\eta(s) = Du(s, \xi(s)) \quad s \in (0,t). \end{aligned}$$

The gradients are taken in the space variable only.

Theorem 5.2.6 (Theorem 6.4.9 Cannarsa and Sinestrari [2004]). *Let $t > 0$ and $x \in \mathbb{R}$. The map that associates with any $(p_t, p_x) \in D^* u(t,x)$ the arc ξ obtained by solving* (5.2.3) *with the terminal conditions*

$$\begin{cases} \xi(t) = x \\ \eta(t) = p_x, \end{cases}$$

provides a one-to-one correspondence between $D^* u(t,x)$ and the set of minimizers of $(CV)_{t,x}$.

Note that this implies by Theorem 3.3.6 and Proposition 3.3.4 in Cannarsa and Sinestrari [2004] that there is a unique minimizer for $(CV_{t,x})$ if and only if u is differentiable in (t,x).

For a semi-concave function f, denote by Γ_f the graph of the reachable super-gradient of f:

$$\Gamma_f := \{(x,p) \in \mathbb{R} \times \mathbb{R} \,|\, p \in D^* f(x)\}.$$

Note that for a continuously differentiable function f it holds that $\Gamma_f = \{(x, Df(x)) \,|\, x \in \mathbb{R}\}$.

Corollary 5.2.7. *Denote by Φ_s the diffeomorphism that maps each (x,p) to its image $(x(s), p(s))$ under the Hamiltonian flow. If $u_0 \in C^1(\mathbb{R})$ then $\Gamma_{u(s)} \subseteq \Phi_s(\Gamma_{u_0})$.*

Proof. Consider $(x,p) \in \Gamma_{u(s)}$. It follows that $(x,p) \in D^*(u(s))(x)$, where we take the reachable gradient only in the x coordinate. By the definition of the reachable gradient, it follows that $D^*(u(s))(x) \subseteq D^*u(s,x)$, where the latter is the reachable gradient in time and space.

Thus, it follows by Theorem 5.2.6, that we can find a trajectory $(\xi(r), \eta(r))_{0 \le r \le s}$ that solves the Hamilton equations with terminal conditions $\xi(s) = x$ and $\eta(s) = p$. Additionally, we know that this trajectory is a minimizer of $(CV_{s,x})$. By Theorem 4.2.2, this trajectory must satisfy the initial conditions $\eta(0) = Du_0(\xi(0))$. We conclude that $(x,p) \in \Phi_s(\Gamma_{u_0})$. □

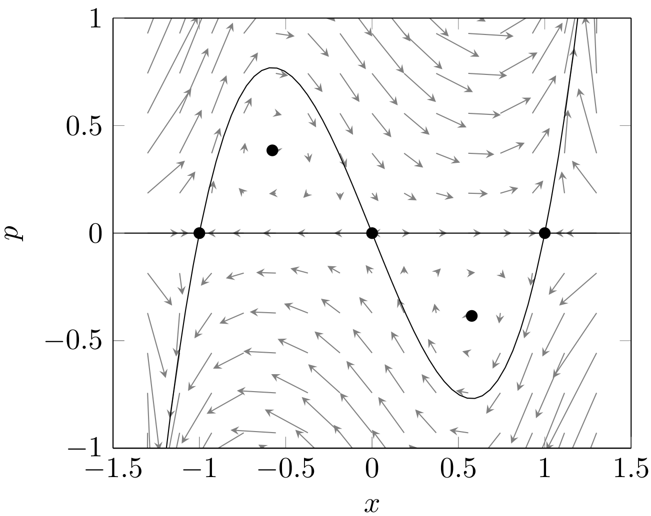
Therefore, we will study the Hamiltonian flow applied to Γ_{u_0}, which will yield information on the gradient of $u(t)$.

Lemma 5.2.8. *The graph of $DV_a(x) = F_a(x)$ is stationary for the Hamiltonian flow.*

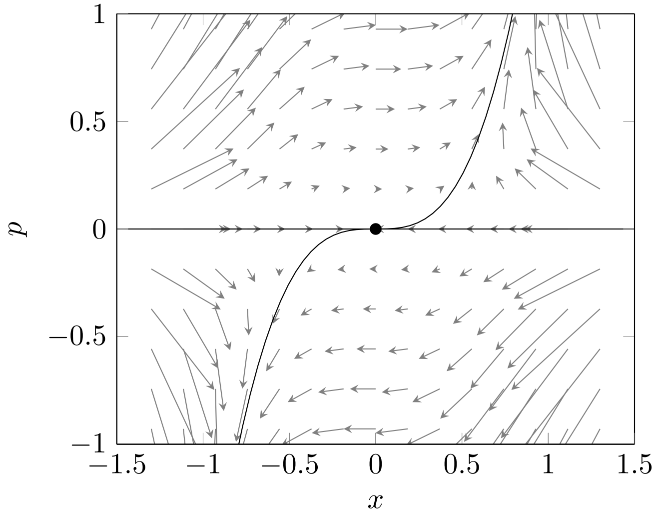
Proof. Note that $H(x, F_a(x)) = 0$ and that the Hamiltonian trajectories have constant energy. □

Below, we give two examples of the Hamiltonian vector-field. The first example is a low-temperature flow, the second example is a high-temperature flow.

Hamiltonian flow for $a = 1$, with stationary curve, and stationary points.



Hamiltonian flow for $a = 0$, with stationary curve, and stationary point.



5.3 UNIQUENESS OF OPTIMAL TRAJECTORIES FOR HIGH TEMPERATURE STARTING POINT

In this section we consider a general Hamiltonian $\hat{H}$ and quadrants in the position momentum plain that satisfy $x \geq x_s$ and $p \geq 0$ for some stationary point x_s of the McKean-Vlasov equation $\dot{x} = \hat{H}_p(x, 0)$.

Assumption 5.3.1. $\hat{H} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and convex in p for every x. Let x_s be a stationary point for the McKean-Vlasov equation $\dot{x} = \hat{H}_p(x, 0)$. We assume that for any fixed $p \geq 0$, the map $x \mapsto -\hat{H}_x(x, p)$ is non-decreasing for $x \geq x_s$.

We start with an auxiliary lemma.

Lemma 5.3.2. *Let $\hat{H}$ be a Hamiltonian and let x_s be a stationary point for $\dot{x} = \hat{H}_p(x, p)$ satisfying Assumption 5.3.1. Then the quadrant $x \geq x_s, p \geq 0$ is preserved under the Hamiltonian flow.*

Remark 5.3.3. By symmetry, we can prove that the quadrant $x \leq x_s, p \leq 0$ is preserved under the Hamiltonian flow by flipping the state-space $(x, p) \mapsto (-x, -p)$.

The appropriate assumption for this case is that

$$x \mapsto -\hat{H}_x(x, p)$$

is increasing for $x \leq x_s$ and $p \leq 0$.

Proof of Lemma 5.3.2. Because x_s is a stationary point for the Hamiltonian flow: $x_s = \hat{H}_p(x_s, 0)$ it follows by the convexity of $\hat{H}$ in the p coordinate that $\hat{H}_p(x_s, p) \geq 0$ for $p \geq 0$. Thus a solution $(x(t), p(t)$ of the Hamilton equations starting at (x_s, p) for some $p \geq 0$ satisfies $\dot{x}(0) \geq 0$.

Additionally, solutions of the Hamilton equations that start with $p(0) = 0$ will have $p(t) = 0$ for all $t \geq 0$ as $\hat{H}(x, 0) = 0$ for all $x \in \mathbb{R}$. This implies that solutions can never cross the $p = 0$ axis. These two statements yield that $\hat{H}$ is quadrant preserving. □

Lemma 5.3.4. *Consider the Hamiltonian $H(x,p) = \frac{1}{2}p^2 - \frac{1}{2}pF_a(x)$. The map $x \mapsto H_x(x,p)$ is non-decreasing for $x \geq 0$ all $p \geq 0$ and for $x \leq 0$ and $p \leq 0$.*

Proof. This follows immediately from

$$-H_x(x,p) = \frac{1}{2}p(3x^2 - a).$$

□

Lemma 5.3.5. *Let $\hat{H}$ and x_s satisfy Assumption 5.3.1.*

Consider $x_s < z_1 < z_2$ such that $0 \leq Du_0(z_1) \leq Du_0(z_2)$. Then we have for all $t > 0$ that $X(t, z_1) < X(t, z_2)$ and $P(t, z_1) < P(t, z_2)$.

Proof. Consider two solutions $(X(t, z_1), P(t, z_1)), (X(t, z_2), P(t, z_2))$ to the Hamilton equations for $\hat{H}$ that satisfy the assumptions of the lemma.

We prove the result by contradiction. Suppose for some time $T > 0$, we have $X(T, z_1) > X(T, z_2)$. Then let $t_1 = \inf\{t > 0 \,|\, X(t, z_1) \geq X(t, z_2)\}$. Note that $t_1 > 0$. It follows that $X(t_1, z_1) = X(t_1, z_2)$ and $\dot{X}(t_1, z_1) \geq \dot{X}(t_1, z_2)$. This implies that

$$\hat{H}_p(X(t_1, z_1), P(t_1, z_1)) \geq \hat{H}_p(X(t_1, z_2), P(t_1, z_2))$$

which, in turn implies that $P(t_1, z_1) \geq P(t_1, z_2)$. Because the image under the flow is a diffeomorphism, we obtain that $P(t_1, z_1) > P(t_1, z_2)$. Hence, there must have been a time $0 < t_0 < t_1$ such that $P(t_0, z_1) = P(t_0, z_2)$ and $\dot{P}(t_0, z_1) \geq \dot{P}(t_0, z_2)$.

Via Hamilton's equations, we find $\hat{H}_x(X(t_0, z_1), P(t_0, z_1)) \leq \hat{H}_x(X(t_0, z_2), P(t_0, z_2))$. Because $P(t_0, z_1) = P(t_0, z_2)$ we find by Assumption 5.3.1 that $X(t_0, z_1)) \geq X(t_0, z_2)$. This contradicts the fact that $t_0 < t_1$. □

Corollary 5.3.6. *Consider the Hamiltonian* $H(x,p) = \frac{1}{2}p^2 - \frac{1}{2}pF_a(x)$*. Suppose our starting rate function* I_0 *is given by* V_b *for* $b \leq 0$ *(high-temperature), then* $I_t = T_t V_b$ *is continuously differentiable for all* $t \geq 0$*. Additionally, this is equivalent to the uniqueness of the minimizer of*

$$I_t(x) := T_t V_b(x) = \inf_{\substack{\xi \in \mathcal{AC} \\ \xi(t)=x}} \left\{ V_b(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s)) \mathrm{d}s \right\}$$

for every fixed $x \in \mathbb{R}$ *and* $t > 0$*.*

Proof. Fix $t > 0$. Note that the graph of the curve $x \mapsto u_0(x) := F_b(x) = V_b'(x)$ is contained in the quadrants $x \geq 0, p \geq 0$ and $x \leq 0, p \leq 0$. Additionally, $x_s = 0$ is a stationary point for $\dot{x} = H_p(x, 0)$.

By Lemma 5.3.2 these two quadrants are preserved under the Hamiltonian flow. By Corollary 5.2.7, we have $\Gamma_{u(t)} \subseteq \Phi_t \Gamma_{u(0)}$.

Arguing separately for these two quadrants, it follows by Lemma 5.3.5, $\{p \,|\, (x,p) \in \Phi_t \Gamma_{u(0)}\}$ is a singleton for all $x \in \mathbb{R}$. We conclude that $u(t)$ is continuously differentiable by Proposition 3.3.4 in Cannarsa and Sinestrari [2004]. Additionally, this means that minimizers are unique by Theorem 5.2.6. □

Even though the methods in this chapter are not immediately applicable to Glauber dynamics for the Curie-Weiss model due to the fact that the results in Cannarsa and Sinestrari [2004] work *only for open sets*, preliminary results based on Proposition 4.3.7 and Lemma 4.4.5 show that issues due to the boundary can be resolved in an ad-hoc manner in the one-dimensional case. We state an analogue of Lemma 5.3.4 for this setting.

Lemma 5.3.7. *We consider* $H : [-1,1] \times \mathbb{R}$ *of the Glauber type:*

$$H(x,p) = \frac{1-x}{2} e^{\beta x} \left[e^{2p} - 1 \right] + \frac{1+x}{2} e^{-\beta x} \left[e^{-2p} - 1 \right].$$

Suppose that $\beta < 1$*, then* $x \mapsto -H_x(x,p)$ *is non-decreasing for* $x \geq 0$ *if* $p \geq 0$ *is fixed and for non-decreasing for* $x \leq 0$ *if* $p \leq 0$*.*

Proof. We only prove the first claim as the second one is proven analogously. We rewrite $H(x,p)$ as

$$H(x,p) = 2\sinh(\beta x + p)\sinh(p) - 2x\cosh(\beta x + p)\sinh(p).$$

We conclude that

$$\begin{aligned}-H_x(x,p) &= 2\,[1-\beta]\cosh(\beta x+p)\sinh(p)\\ &\quad - 2x\beta\sinh(\beta x+p)\sinh(p).\end{aligned}$$

Thus, it follows that both terms of $x \mapsto -H_x(x,p)$ are non-decreasing individually. □

5.4 MAXWELL CONSTRUCTION OF NON-GIBBSIAN POINTS

We have seen above that in the case of a starting rate function with a minimum for $x = 0$ and which is convex, we obtain differentiability in a straightforward way. In the more involved case that the starting rate function is not convex, we do not expect the push-forward $C_t = \Phi_t C_0$ under the Hamiltonian flow

$$\begin{bmatrix}\dot{x}\\ \dot{p}\end{bmatrix} = \begin{bmatrix}H_p(x,p)\\ -H_x(x,p)\end{bmatrix} = \begin{bmatrix}p - \frac{1}{2}F_a(x)\\ \frac{1}{2}pF_a'(x)\end{bmatrix},$$

to correspond to the graph of a function. In fact, this can be proven by combining the identification in Theorem 5.2.6 between optimal trajectories to elements in the reachable supergradient of u, with the results obtained in Ermolaev and Külske [2010], den Hollander et al. [2015].

Thus, a more elaborate approach is needed in this more general setting. Because the Hamilton equations have unique solutions, the graph C_t can not intersect itself. This means that C_t consists of finitely many pieces which can be represented as graphs of functions. This allows to give a Maxwell-construction for those points at which the time-evolved rate function is not C^1. We will show that the function u_t can be reconstructed as follows.

(a) For any branch of the derivative, expressible as a graph of a function Du_t^i, $i = 1, \dots, k$, we construct corresponding branches of the function u_t^i, up to constants C^i.

(b) We, adjust the constants in such a way that the branches of these functions when put together form a continuous curve.

(c) Take the lower envelope of that curve.

(d) Add a constant such that the minimal value of the resulting function is 0 to obtain the time-evolved rate-function.

From this construction we can identify the discontinuity points of the resulting time-evolved rate function by a graphical construction which only looks at sizes of overhangs which result from time-evolution. Indeed, the points where different branches come together in a non-differentiable way, are given in terms of an "equal-area under the overhang of Du"-requirement.

5.4.1 *The construction*

We carry out the construction introduced above, based on the method of characteristics. Consider the set of Hamilton equations

$$\begin{cases} \dot{X} = H_p(X, P) \\ \dot{P} = -H_x(X, P) \end{cases}$$

with starting conditions

$$\begin{cases} X(0) = z \\ P(0) = \nabla u_0(z) \end{cases}$$

and denote the z dependent family of solutions by $(X(t,z), P(t,z))$. Finally, we solve

$$\dot{U} = -H(X, P) + PH_p(X, P) \qquad U(0, z) = u_0(z) \tag{5.4.1}$$

and set $U(t, z)$ to be the solution based on $(X(t,z), P(t,z))$. Note that $\dot{U}$ equals

$$\begin{aligned} \mathcal{L}(X, \dot{X}) &= \sup_p \left\{ p\dot{X} - H(X, p) \right\} \\ &= \sup_p \{pH_p(X, P) - H(X, p)\} = PH_p(X, P) - H(X, P), \end{aligned}$$

in other words, U measures the Lagrangian cost along the solution of the Hamiltonian flow. By Theorem 5.2.4, we have

$$u(x, t) = \inf_{z, X(t,z)=x} U(t, z). \tag{5.4.2}$$

The branches of the derivative that we will be using for the Maxwell construction are exactly branches of $P(t, z)$, and their integrals will be branches of $U(t, z)$. Recall that $z \mapsto (X(t,z), P(t,z))$ are $z \mapsto$

$(X(t,z), U(t,z))$ are two parametrized curves from $\mathbb{R}$ to $\mathbb{R}^2$. Thus, using the inverse function in appropriate domains, we can re-express these curves implicitly as branches of functions $x \mapsto P(t, z(x))$ and $x \mapsto U(t, z(x))$, where $x \mapsto z(x)$ is an appropriate branch of inverse function of $z \mapsto X(t,z)$.

The next technical lemma is crucial to be able to apply the inverse function theorem.

Lemma 5.4.1. *For any $t \geq 0$ and $z \in \mathbb{R}$, we have*

$$U_z(t,z) = P(t,z)X_z(t,z).$$

Note in particular, that this means that if the x variable does not change under varying z, then the value function remains the same.

For parts where x does vary; i.e. where we want to interpret it as one of the candidate functions for the Maxwell construction; we have $\frac{\mathrm{d}}{\mathrm{d}x}U(t,z) = \frac{U_z(t,z)}{X_z(t,z)} = P(t,z)$ which is the desired property. Indeed, the graph of P is the derivative of the candidate value function.

Proof. In the proof, the dependence of X, P, U on (t,z) will be suppressed in the notation. As above, $\dot{U}$ means $\frac{\mathrm{d}}{\mathrm{d}t}U$. Starting from (5.4.1), differentiating with respect to z, we see

$$\begin{aligned}
\dot{U}_z &= -H_x(X,P)X_z - H_p(X,P)P_z \\
&\qquad + P_z H_p(X,P) + PH_{xp}(X,P)X_z + PH_{pp}(X,P)P_z \\
&= -H_x(X,P)X_z + PH_{xp}(X,P)X_z + PH_{pp}(X,P)P_z \\
&= -H_x(X,P)X_z + P\dot{X}_z \\
&= \dot{P}X_z + P\dot{X}_z.
\end{aligned}$$

This indeed equals the time derivative of $t \mapsto P(t,z)X_z(t,z)$. Because $U_z(0,z) = P(0,z)$ and $X_z(0,z) = 1$, the lemma is proven. □

Consider the parametrized curve $\gamma_t(z) := \Phi_t(z, Du_0(z)) = (X(t,z), P(t,z))$ in $\mathbb{R}^2$, so that

$$C_t = \Phi_t C_0 = \{\gamma_t(z) \,|\, z \in \mathbb{R}\}$$

To split the set $(X(t,z), P(t,z))$ into branches that can be interpreted as functions, we need to cut apart the set at the points where the tangent is vertical.

Let $V := \{(0,v) \mid v \in \mathbb{R}\}$. Define $z_1 = \inf\{z \mid D\gamma_t(z) \in V\}$, which is the first moment the curve $\gamma_t(z)$ has a vertical gradient. This is the first point z where an overhang can be created, and where $X_z(t,z) = 0$. Put $z_1^* = \inf\{z \geq z_1 \mid D\gamma_t(z) \notin V\}$. Note that $z_1^* \neq z_1$ only if the curve γ_t has constant first coordinate for $z \in [z_1, z_1^*]$, i.e. γ_t has a vertical slope on the interval $[z_1, z_1^*]$.

For $n \geq 2$, iteratively define points of vertical gradient by

$$\begin{aligned} z_n &= \inf\left\{z > z_{n-1}^* \,\middle|\, D\gamma_t(z) \in V\right\} \\ z_n^* &= \inf\left\{z > z_n^* \mid D\gamma_t(z) \notin V\right\}. \end{aligned}$$

Define for all n the projections of $\gamma_t(z_n)$ on the horizontal axis: $x_n := \pi_x(\gamma_t(z_n))$. We then define a collection of intervals $I_0 := (-\infty, x_1]$,

$$I_n := \begin{cases} [x_n, x_{n+1}] & \text{if } x_n < x_{n+1} \\ [x_{n+1}, x_n] & \text{if } x_n > x_{n+1}. \end{cases}$$

On the intervals I_n, we first define the continuous branches of the set C_t with the properties

(a) $w_n : I_n \to \mathbb{R}$,

(b) $w_n(X(t,z)) = P(t,z)$ for $z \in [z_n^*, z_{n+1}]$.

Then construct a collection of functions v_n such that

(a) $v_n : I_n \to \mathbb{R}$,

(b) $v_n(X(t,z)) = U(t,z)$ for $z \in [z_n^*, z_{n+1}]$.

The next two results show that the functions v_n and w_n have the desired properties.

Proposition 5.4.2. *For every n, the function v_n is differentiable on I_n and has derivative*

$$\frac{\mathrm{d}}{\mathrm{d}x} v_n(x) = w_n(x).$$

In other words, we have

$$\frac{\mathrm{d}}{\mathrm{d}x} v_n(X(t,z)) = P(t,z) = w_n(X(t,z)), \qquad z \in [z_n^*, z_{n+1}].$$

Proof. We prove the second statement, as it implies the first. On the interval (z_n^*, z_{n+1}), the function $z \mapsto X_z(t,z)$ is non-zero and has fixed sign, so by the chain rule and the inverse function theorem, we obtain

$$D_x v_n(X(t,z)) = \frac{D_z v_n(X(t,z))}{X_z(t,z)} = \frac{U_z(t,z)}{X_z(t,z)}$$
$$= P(t,z) = w_n(X(t,z)).$$

The third equality is a consequence of Lemma 5.4.1. □

The next result is a second consequence of Lemma 5.4.1.

Lemma 5.4.3. *We have for all* n *that* $v_n(x_{n+1}) = v_{n+1}(x_{n+1})$.

Proof. First recall that $x_k = \pi_x(\gamma_t(z_k)) = \pi_x(\gamma_t(z_k^*))$. For $z \in [z_k, z_{k+1}]$, we have $D\gamma_t(z) \in V$ which is equivalent to saying that $X_z(t,z) = 0$. By Lemma 5.4.1 it follows that $U_z(t,z) = 0$.
It follows that $v_n(x_n) = U(t,z_n) = U(t,z_n^*) = v_{n+1}(x_n)$. □

Combining (5.4.2) and the definition of the functions v_n, we obtain the following result.

Theorem 5.4.4 (Maxwell-construction). *For* $t \geq 0$, *we find that*

$$u(t,x) = \inf_{n : x \in I_n} v_n(x). \tag{5.4.3}$$

5.5 THE LIMIT OF THE RATE FUNCTION WITH TIME GOING TO INFINITY

The Maxwell-construction tells us that the rate function is obtained by integrating the area under the 'graph' of $(X(t,z), P(t,z))$. Even though the approach is work in progress, graphical analysis of the Maxwell construction allows one to reproduce results like in Ermolaev and Külske [2010]. The use of this construction to rigorously prove similar results is work in progress. Below we consider the simpler limiting behaviour of the rate function when time goes to infinity. In Section 5.6, we comment on the rate function for a finite time.

Recall that $V_c(x) = \frac{1}{4}x^4 - \frac{1}{2}cx^2 + C_c$, where $C_c = \sqrt{1}4c^2$ is the constant such that the minimum of V_c equals 0. Note that $F_c(x) = x^3 - cx$ and that V_c has either one minimum at 0 if $c \leq 0$, or two minima $x_{c,-} = -\sqrt{c}$ and $x_{c,+} = \sqrt{c}$ if $c \geq 0$.

The dynamics of our problem was defined in terms of the Hamiltonian $H(x,p) = \frac{1}{2}p^2 - \frac{1}{2}pF_a(x)$, and we start with rate function $V_b(x)$ at time 0.

We obtain the following result on the limiting rate function.

Theorem 5.5.1. *Let a be the parameter of the dynamics, and b parameter of the rate function. Suppose that it does not hold that [$b > 0$ and $a > b$]. Then if $u_0 = V_b$, we find that*

$$\lim_{t\to\infty} u(t,x) = V_\infty(x), \qquad x \in \mathbb{R},$$

where $V_\infty = V_a$ if $a \leq 0$ and

$$V_\infty(x) = \begin{cases} V_a(x) & \text{for } x \notin [-\sqrt{a}, \sqrt{a}] \\ V_a(x) \wedge V_b(0) & \text{for } x \in [-\sqrt{a}, \sqrt{a}], \end{cases}$$

if $a \geq 0$. If $b > 0$ and $a > b$, then we conjecture this result to be true.

In the case that $b > 0$ and $a > b$, note that the solutions of $V_a(x) = V_b(0)$ in the region $(-\sqrt{a}, \sqrt{a})$ are given by $\pm\sqrt{a-b}$.

If $b = a$, the curve $F_b = F_a$ which is stationary under the Hamiltonian flow and there is nothing to prove: $V_\infty = V_a = V_b$. The argument is divided into four cases:

(a) Cooling down from high temperature: $b \leq 0$ and $a > b$.

(b) Heating up a high temperature starting profile: $b \leq 0$ and $a < b$.

(c) Heating up a low temperature profile: $b > 0$ and $b > a$.

(d) Cooling down a low temperature profile with a low temperature dynamics: $b > 0$ and $a > b$.

We will prove (a)-(c) and argue why we expect the stated result to be true for (d).

Proof of (a), $b \leq 0$, $a > b$. Fix some $y \in \mathbb{R}$. We show that $u(t,y) \to V_\infty(x)$. Fix some $t > 0$, and denote by $C_t = \Phi_t F_b$, the image of the graph $(x, F_b(x))$ under the Hamiltonian flow. Because we have a high temperature starting profile, C_t is the graph of a function by Corollary 5.3.6, we denote this function by $v^t : \mathbb{R} \to \mathbb{R}$. In particular, we obtain that $Du(t,\cdot) = v^t(\cdot)$.

We will show that v^t converges uniformly on compacts to the function

$$v^\infty(x) := \begin{cases} F_a(x) \vee 0 & \text{for } x \geq 0 \\ F_a(x) \wedge 0 & \text{for } x \leq 0, \end{cases}$$

so that the result of Theorem 5.5.1 for this setting follows by integration.

We consider the branch $y \mapsto v^t(y)$ for $y \geq 0$. The other part follows by symmetry. Fix some $y_{max} > 0$ and consider $y \in [0, y_{max}]$. Let $\{x(s), p(s)\}_{0 \leq s \leq t}$, $x(t) = y$ be the unique optimal trajectory for

$$u(t,y) = \inf_{\substack{\xi \in \mathcal{AC} \\ \xi(t)=y}} \left\{ V_b(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s)) \mathrm{d}s \right\}.$$

By Theorem 5.2.2, there exists $y_0 \geq 0$ such that $(x(0), p(0)) = (y_0, F_b(y_0))$. As a consequence of $b \leq 0$, $a > b$, we have that the energy of this curve equals $E = H(y_0, V_b(y_0)) \geq 0$. Re-expressing the Hamilton equations in space-energy coordinates, we find that

$$\dot{x}(s) = \frac{1}{2}\sqrt{F_a(x(s))^2 + 8E} \geq \sqrt{2E}.$$

This implies that

$$\frac{y}{t} \geq \frac{x(t) - x(0)}{t} \geq \sqrt{2E},$$

and as $H(y, p(t)) = E$ by the conservation of energy along the Hamiltonian flow:

$$0 \leq \frac{1}{2} p(t) \left(p(t) - F_a(y)\right) \leq \frac{y^2}{2t^2} \leq \frac{y_{max}^2}{2t^2}. \tag{5.5.1}$$

Because $p(t) \geq 0 \vee F_a(x)$ by the condition that $E \geq 0$, $p(t)$ must be close to the largest zero of the function $z \mapsto H(y,z) = \frac{1}{2}z(z - F_a(y))$. As a function of $y \geq 0$, this largest zero is given by the function v^∞.

Thus equation (5.5.1) gives us a uniform bound on the difference between v^t and v^∞ in the interval $[0, y_{max}]$. We conclude that $v^t \to v^\infty$ uniformly on compacts and thus $u(t, \cdot) \to V_\infty(\cdot)$ point-wise and uniformly on compacts. □

Proof of (b), $b \leq 0$, $a < b$. In this case, both curves $x \mapsto F_a(x), F_b(x)$ are in the upper right and lower left quadrants. In contrast to the proof above, we have that $H(x, F_b(x)) \leq 0$ for all x. However, we still have that the set $C_t = \Phi_t F_b$ is the graph of a function by Corollary 5.3.6 As before, we denote this function by $v^t : \mathbb{R} \to \mathbb{R}$. Additionally, we have the equality $Du(t, \cdot) = v^t(\cdot)$.

We will show that $v^t \to F_a$ point-wise, as this proves the claim in this particular setting. Again we argue only for the part of the curve where $y \geq 0$, as the two regions are symmetric. Fix $y > 0$ and $t > 0$. Then there is a unique trajectory $\{x_t(s), p_t(s)\}_{0 \leq s \leq t}$, $x_t(t) = y$ that solves the Hamilton equations and such that $(x_t(0), p_t(0)) = (y_t, F_b(y_t))$. Using the Hamilton equation for the momentum, we find $\dot{p}_t(s) = -H_x(x_t(s), p_t(s)) = \frac{1}{2}p_t(s)(x_t(s)^2 - a) \geq -\frac{1}{2}ap_t(s)$. Grönwall's lemma yields $p_t(t) \geq e^{-\frac{1}{2}at}p_t(0)$. Because the energy of the curve is negative, we find that $p_t(t) < F_a(y)$ and thus that $F_b(y_t) = p_t(0) < e^{\frac{1}{2}at}F_a(y)$.

F_b is a strictly increasing function, which gives the upper bound $y_t < F_b^{-1}(e^{\frac{1}{2}at}F_a(y)) =: C_{t,y}$. And as such a lower bound on the energy of the curve:

$$0 > H(y_t, p_t(0)) \geq H(y_t, \frac{1}{2}F_b(y_t)) = -\frac{1}{8}F_a(y_t)^2 \geq -\frac{1}{8}F_a(C_{t,y})^2.$$

Note that as $t \to \infty$, we have $C_{t,y} \downarrow 0$, so the energy of the optimizing trajectory gets pushed to 0 from below. We obtain that any limit point of $p_t(t)$ as $t \to \infty$ must equal 0 or $F_a(y)$. The first however, is not possible as we now explain.

The evolution of the Hamiltonian flow for a fixed time T is a diffeomorphism, and on the horizontal axis the McKean-Vlasov dynamics converges to the equilibrium point 0, i.e. $\dot{x} = H_p(x, 0) < 0$ for $x \geq 0$. This implies that if $p_t(t)$ is very close to 0, the starting point $(x_t(0), p_t(0))$ must satisfy $x_t(0) > y$ which contradicts the fact that $(x_t(0), p_t(0)) = (y_t, F_b(y_0))$. □

We proceed with the preparations of the proof of (c), where we start with a low temperature starting profile. In this case, we are not able to use Corollary 5.3.6. Instead, we work with the Maxwell-construction.

We sketch the strategy of the proof. First note that the starting curve $x \mapsto F_b(x)$ satisfies $F_b(x) > 0$ for $x \in (-\sqrt{b}, 0)$ and $F_b(x) < 0$ for $x \in (0, \sqrt{b})$. We first consider the part of the graph $(x, F_b(x))$ for $x \in (-\sqrt{b}, 0)$. As in the proof of (a), we re-parametrize the Hamilton equations in terms of space and energy. Thus, we re-parametrize (x, p) as

$$x(x,p) = x$$
$$E(x,p) = H(x,p) = \frac{p^2}{2} - \frac{1}{2}pF_a(x)$$

We solve for p to re-express the Hamilton equations in terms of x and E. We can only do in a one-to-one manner in restricted regions. Note

$$p = -\frac{1}{2}F_a(x) \pm \sqrt{\frac{1}{4}F_a(x)^2 + 2E}.$$

For the proof of (c), we consider the setting where $E \geq 0$ and $p \geq 0$, thus we have to choose

$$p = \frac{1}{2}F_a(x) + \sqrt{\frac{1}{4}F_a(x)^2 + 2E}.$$

We conclude that the Hamilton equations get transformed to

$$\dot{x} = \sqrt{\frac{1}{4}F_a(x)^2 + 2E}, \qquad \dot{E} = 0.$$

Again let $z \in (-\sqrt{b}, 0)$ and let $(X(t,z), P(t,z))$ be the solution to the Hamilton equations with $X(0,z) = z$ and $P(0,z) = F_b(z)$. Because the corresponding energies are non-negative, we see that $\dot{X}(t,z) \geq \sqrt{2E}$. In particular, for large enough times, we have $X(t,z) \geq 0$. The same thing happens for trajectories starting with $z \in (0, -\sqrt{b})$, so we see that an overhang is created. We conclude that we need multiple branches of the Maxwell construction for our analysis.

We refine our analysis. First, we transform our curve $F_b(x)$ for $x \in [-\sqrt{b}, 0]$ to the corresponding energies $E : [-\sqrt{b}, 0] \to [0, \infty)$:

$$E(x) = \frac{1}{2}(x^3 - bx)^2 - \frac{1}{2}(x^3 - bx)(x^3 - ax) = \frac{(a-b)}{2}(x^4 - bx^2) \quad (5.5.2)$$

which takes its maximum E_{max} at some point $x_{max} \in [-\sqrt{b}, 0]$. Note that the function is increasing for $x \leq x_{max}$ and decreasing for $x \geq x_{max}$.

Fix some $y_{max} > 0$, we study $u(t,x)$ for $x \in [0, y_{max}]$ and large times. In general, it is hard to study the exact structure of the push forward of the curve $(x, F_b(x))$ under the time evolution of the Hamiltonian vector field. However, combining the fact that $\dot{X}(t,z) > \sqrt{2E}$ and the form of the energy curve in (5.5.2), it is clear that if we choose t large enough, the only Hamiltonian trajectories that contribute to the Maxwell construction for $x \in [0, y_{max}]$ must have started in either the region close to 0, or the region close to $-\sqrt{b}$.

First, we prove a lemma that we will use to study the evolution of the Hamiltonian flow for curves that have started close to $-\sqrt{b}$.

Lemma 5.5.2. *For $z \in \mathbb{R}$ let $(X(t,z), P(t,z))$ be the solution to the Hamilton equations with $X(0,z) = z$ and assume that . Suppose that*

(a) $z_1 < z_2$,

(b) $P(0, z_1), P(0, z_2) \geq 0$,

(c) $0 \leq H(z_1, F_b(z_1)) < H(z_2, F_b(z_2))$.
Then $X(t, z_1) < X(t, z_2)$ *for all* $t \geq 0$.

Proof. Set $E_i = H(z_i, F_b(z_i))$, $i \in \{1, 2\}$.
Suppose that the claim is false. Then there exists t_0 such that $x = X(t_0, z_1) = X(t_0, z_2)$ and $\dot{X}(t_0, z_1) \geq \dot{X}(t_0, z_2)$. But this implies, using the Hamilton equations in terms of energy(using the correct representation implied by (b)), that $\sqrt{\frac{1}{4}F_a(x)^2 + 2E_1} > \sqrt{\frac{1}{4}F_a(x)^2 + 2E_2}$ which is in contradiction with $E_1 < E_2$. □

By (5.5.2), we see that $x \mapsto E(x)$ is increasing for $-\sqrt{b} \leq x \leq x_{max}$, so this lemma tells us that for large times, we get exactly one contributing function to the Maxwell construction on the interval $[0, y_{max}] \subseteq [X(t, -\sqrt{b}), y_{max}]$, that originates from points that start close to $-\sqrt{b}$.

We proceed by considering curves that start close to 0. A priori we cannot immediately use Lemma 5.5.2, as the energies $x \mapsto E(x)$ are decreasing for $x_{max} \leq x \leq 0$.
By linearising the curve $F_b(x)$ close to 0 and linearising the Hamiltonian vector field around 0, we see that for large times, the evolution tilts the starting curve from having a negative slope first to one with a vertical slope and then to one that has a positive slope. Using Lemma 5.5.2 from this specific moment onward, we see that curves starting close to 0 contribute can be expressed as a unique branch of the Maxwell-construction.

To make this argument rigorous, we need that the linearisation of the Hamiltonian flow can be achieved in a C^1 manner. Otherwise, we are not able to establish the conditions for Lemma 5.5.2.

First, we start by linearising the Hamiltonian vector field around $(0, 0)$. Consider the Hamilton equations:

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} H_p(x, p) \\ -H_x(x, p) \end{bmatrix} = \begin{bmatrix} p - \frac{1}{2}\left(x^3 - ax\right) \\ \frac{p}{2}\left(3x^2 - a\right) \end{bmatrix}.$$

We linearise around $(0, 0)$ and obtain a linear ordinary differential equation

$$\begin{bmatrix} \dot{x} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}a & 1 \\ 0 & -\frac{1}{2}a \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix} =: A \begin{bmatrix} x \\ p \end{bmatrix}.$$

This linearised local system is solved by

$$\begin{bmatrix} x_l(t) \\ p_l(t) \end{bmatrix} = \begin{bmatrix} e^{\frac{1}{2}at} & t \\ 0 & e^{-\frac{1}{2}at} \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix}.$$

Locally, the curve $x \mapsto F_b(x)$ looks like a line with slope $-b$. The time t_0 at which the line with this slope becomes vertical is given by the solution of $e^{\frac{1}{2}at_0} = t_0 b$.

To use this idea in a rigorous proof of Lemma 5.5.4 below, we need C^1 regularity which is provided by Theorem 5 in Sell [1985].

Lemma 5.5.3. *Fix $t^* > t_0$. Denote by $F := \Phi_{t^*}$ the diffeomorphism generated by the Hamiltonian flow corresponding to the image of the flow at time t^*. Denote by $G = e^{t^* A}$ the image under the flow of the linearised system.*

Then there are open neighbourhoods U_1, U_2 of $(0,0)$ and a C^1 diffeomorphism $\mathcal{H} : U_1 \to U_2$, $\mathcal{H}(0) = 0$, such that $F = \mathcal{H}^{-1} G \mathcal{H}$

Proof. In the terminology of Sell [1985], the matrix A satisfies the strong Sternberg condition of order 2 and the 2-smoothness of A is 1. Thus the lemma follows from Theorem 5 in Sell [1985]. □

Using this C^1 diffeomorphism between the image of the Hamiltonian flow and the linearised system, we obtain the following lemma.

Lemma 5.5.4. *For $z \in \mathbb{R}$ let $(X(t,z), P(t,z))$ be the solution to the Hamilton equations with $X(0,z) = z$ and $P(0,z) = F_b(z)$.*

There exists x_0 satisfying $x_{max} < x_0 < 0$ such that for all $t > t_0$, where t_0 is the solution of $e^{\frac{1}{2}at} = tb$, it holds that for all $x_0 < z_1 < z_2 \leq 0$, we have $X(t, z_1) > X(t, z_2)$.

Proof. By the analysis preceding Lemma 5.5.3, we saw that the line tangent to $x \mapsto F_b(x)$ becomes vertical at the solution t_0 of $e^{\frac{1}{2}at} = tb$ under the linearised equation $(\dot{x}, \dot{p}) = A(x,p)$. Because the map $\mathcal{H}$ is a C^1 diffeomorphism by Lemma 5.5.3, it follows that for a fixed $t^* > t_0$, the slope of the image of $(x, F_b(x))$ under the Hamiltonian flow at the point $(0,0)$ is positive.

Because $F_b(x)$ is continuously differentiable in x, we obtain by the C^1 property of $\mathcal{H}$, the existence of a $x_0 < 0$ (and $x_0 > x_{max}$), such that for $x_0 \leq z_1 < z_2 < 0$, we have that $X(t^*, z_1) > X(t^*, z_2)$. By Lemma 5.5.2, this ordering remains true for all times $t \geq t^*$. □

Proof of Theorem 5.5.1 (c): $b > 0$, $b > a$. We study the behaviour of the curve $x \mapsto F_b(x)$ under the Hamiltonian flow for $x \geq -\sqrt{b}$, which is sufficient by symmetry.

First of all, dynamics of the curve $x \mapsto F_b(x)$ for $x \geq \sqrt{b}$ can be treated completely as in the proof of (b). To summarize: over the course of time this curve converges point-wise to the curve $x \mapsto F_a(x)$, for $x \geq \sqrt{a \vee 0}$.

Next, we study the part of the graph $(z, F_b(z))$ where $z \in [-\sqrt{b}, 0]$, and thus $F_b(x) \geq 0$. The part $z \in [0, \sqrt{b}]$ has similar behaviour. To study the trajectories starting with $z \in [-\sqrt{b}, 0]$, we use Lemma's 5.5.4 and 5.5.2 and the methods from the proof of (a).

As before, denote for $z \in \mathbb{R}$ let $(X(t,z), P(t,z))$ be the solution to the Hamilton equations with $X(0,z) = z$ and $P(0,z) = F_b(z)$.

Fix some $y_{max} > 0$. We show that the time evolved graph of $(z, F_b(z))$ for $z \in [-\sqrt{b}, 0]$, gives rise to two branches above the interval $[-\sqrt{b}, y_{max}]$ in the Maxwell construction for sufficiently large times t.

Fix $\varepsilon > 0$ and fix a time t such that $t\sqrt{2\varepsilon} > -\sqrt{b} + y_{max}$ and $t \geq t_0$, where t_0 was introduced in Lemma 5.5.4.

The solutions to the Hamilton equations satisfy $\dot{X} \geq \sqrt{2E}$. Thus, for any z such that $H(z, F_b(z)) \geq \varepsilon$, we obtain that $X(t,z)$ satisfies

$$X(t,z) \geq z + t\sqrt{2H(z, F_b(z))} \geq z + t\sqrt{2\varepsilon} > z + y_{max} + \sqrt{b} \geq y_{max}.$$

In particular, these z do not contribute to a branch of the Maxwell construction above the interval $[-\sqrt{b}, y_{max}]$. That leaves two sub-intervals of $[-\sqrt{b}, 0]$ where z is such that $H(z, F_b(z)) \leq \varepsilon$. Because $t \geq t_0$, the time evolution of the graphs of $z \mapsto F_b(z)$ over these two sub-intervals are again graphs by Lemma's 5.5.2 and 5.5.4. In particular, the part where the energy increases for increasing z becomes a graph over the interval $[X(t, -\sqrt{b}), y_{max}]$ and the part where the energy decreases becomes a graph over the interval $[0, y_{max}]$. Note that as $t \to \infty$, we have that $X(t, -\sqrt{b}) \to \sqrt{a \vee 0}$.

We conclude by connecting the analysis of the time-evolved graph with the Maxwell-construction to conclude the proof. In particular, for t large as above, the argument above gives two contributions to the Maxwell construction on the interval $[0, y_{max}]$. Because the problem is symmetric around 0, there is also one contribution with curves starting in the $[0, \sqrt{b}]$ region. We conclude that there are three contributions to the push-forward of the graph of $(z, F_b(z))$ that are of interest for the Maxwell construction on the interval $[0, y_{max}]$. For the first and third contribution , we consider it on an extended interval, as this will be of use for the integration of this curve.

(1) $w_1 : [X(t, -\sqrt{b}), y_{max}] \to [0, \infty)$, $w_1(X(t, -\sqrt{b})) = 0$ and $w_1(x)$ uniformly close to the function obtained by taking the largest zero of $p \mapsto H(x, p)$, i.e. the function $h_+(x) := 0 \vee F_a(x)$.

(2) $w_2 : [0, y_{max}] \to [0, \infty)$, $w_2(0) = 0$, and $w_2(x)$ uniformly close to $h_1(x) = 0 \vee F_a(x)$

(3) $w_3 : [0, y_{max} \vee X(t, \sqrt{b})] \to (-\infty, 0]$, $w_3(X(t, \sqrt{b})) = 0$ and $w_3(x)$ is uniformly close to the smallest zero of $p \mapsto H(x, p)$, i.e. the function $h_-(x) := 0 \wedge F_a(x)$.

Because we are not using all values of $z \in \mathbb{R}$ to construct these three branches of the push-forward, we can-not immediately use the Maxwell-construction. However, each of the three branches contains a point where we know the value of $z \mapsto U(t, z)$. In particular, we know that $U(t, -\sqrt{b}) = U(t, \sqrt{b}) = 0$, as these points start with $U(0, -\sqrt{b}) = U(0, \sqrt{b}) = 0$ and follow the zero cost trajectory. Additionally, we know that $U(t, 0) = V_b(0)$, as this curves start at $U(0, 0) = V_b(0)$ and is stationary with zero Lagrangian cost.

We conclude, by integration that v_1, v_2, v_3 are uniformly close to the following three functions.

(a) For $x \in [X(t, -\sqrt{b}), y_{max}]$ the function v_1 is uniformly close to

$$\begin{cases} V_a(x) & \text{if } x \leq 0 \\ V_a(0) & \text{if } 0 \leq x \leq \sqrt{a \vee 0} \\ V_a(0) + \int_{\sqrt{a \vee 0}}^{x} F_a(q) \mathrm{d}q & \text{if } x \geq -\sqrt{a \vee 0}. \end{cases}$$

(b) For $x \in [0, y_{max}]$ the function v_2 is uniformly close to

$$\begin{cases} V_a(0) & \text{if } 0 \leq x \leq \sqrt{a \vee 0} \\ V_a(0) + \int_{\sqrt{a \vee 0}}^{x} F_a(q) \mathrm{d}q & \text{if } x \geq -\sqrt{a \vee 0}. \end{cases}$$

(c) For $x \in [0, y_{max} \vee X(t, \sqrt{b})]$ the function v_3 is uniformly close to $x \mapsto V_a(x)$.

In particular, by Theorem 5.4.4, the function $x \mapsto u(t, x)$ is given by $u(t, x) = v_1(x) \wedge v_2(x) \wedge v_3(x)$. If $a \leq 0$ all three functions are uniformly close to $V_a(x)$ on the interval $[0, y_{max}]$, whereas if $a < 0$ and t is sufficiently large v_3 is the smallest. We conclude that $u(t,\cdot) \to V_a(\cdot)$ as $t \to \infty$ uniformly on compact sets. □

The final case of the proof of Theorem 5.5.1 is the case where we cool down a low-temperature starting profile.

Argument for the conjecture (d): $b > 0$, $a > b$. We show that $u(t,\cdot)$ converges to the function

$$V_\infty(x) = \begin{cases} V_a(x) & \text{for } x \notin [-\sqrt{a}, \sqrt{a}] \\ V_a(x) \wedge I_b(0) & \text{for } x \in [-\sqrt{a}, \sqrt{a}], \end{cases} \tag{5.5.3}$$

uniformly on compact sets. As before, we only consider the part $x \geq 0$. In this setting, the part of the curve that starts out in the quadrant $x \geq 0, p \geq 0$ remains there under the Hamiltonian flow. The part where $x \geq 0$ and $p \leq 0$ remains in the bounded region where $H(x,p) \leq 0$ as the energy is preserved.

The evolution under the Hamiltonian flow of the curve $(z, F_b(z))$, for z such that $F_b(z) \geq 0$, behaves as in the proof of part (a). In particular, the curve converges to $x \mapsto F_a(x)$, which takes care of the region $x \notin [0, \sqrt{a}]$ of (5.5.3).

Thus, we are left with the region where $z \geq 0$ and $F_b(z) \leq 0$. This part of the curve lies in a region where the Hamiltonian flow is rotating around a stationary point, see the graphs on page 127. Suppose that E_{min} is the minimal energy in this region. Then the area where $x > 0$, $p \leq 0$ and $E_{min} < H(x,p) < 0$ allows for global action-angle coordinates by Theorem 2.2 in Duistermaat [1980]. In these coordinates, the Hamiltonian vector field is non-zero only in the angle coordinates.

Thus, for very large times, we find a branch $w_1 : [0, x_{1,max}]$, $x_{1,max} < \sqrt{a}$ of the push-forward of the graph $(z, F_b(z))$ that is uniformly close to $x \mapsto 0$. Additionally, we find a branch $w_2 : [x_{2,min}, X(t\sqrt{b})]$, $x_{2,min} > 0$ uniformly close to $x \mapsto F_a(x)$. Note that $X(t, \sqrt{b}) \to \sqrt{a}$ as $t \to \infty$. As in the proof of (c), we know that $U(t,0) = I_0(b)$ and $U(t,\sqrt{b}) = 0$ as these are zero cost trajectories. We conclude that the integrated curves v_1 and v_2 satisfy

(a) For $x \in [0, x_{1,max}]$ the function v_1 is uniformly close to $x \mapsto I_b(0)$.

(b) For $x \in [x_{2,min}, X(t,\sqrt{b})]$ the function v_2 is uniformly close to $V_a(x)$.

We conjecture that the other branches of the push-forward of the graph $(z, F_b(z))$ are sub-optimal. Namely; the trajectories corresponding to these branches have oscillated at least once around the x value of the stationary point of the Hamiltonian vector field, whereas the trajectories corresponding to the branches w_1, w_2 start close to an equilibrium point and move in a 'straight' line towards their end point.

If this argument is accepted, we find that $u(t, \cdot) = v_1(x) \wedge v_2(x)$, which was our conjecture. □

5.6 THE RATE FUNCTION FOR A FINITE TIME

The methods in the proof of Theorem 5.5.1 can be heuristically used to obtain information on the differentiability of the rate function for a finite time. Note that by Corollary 5.3.6, we have the differentiability of the rate function at time t if we start with a starting curve F_b, $b \leq 0$.

Thus, we restrict ourselves to the analogues of (c) and (d) of Theorem 5.5.1 above.

Even though we do not spell out these results here in full detail. We conjecture on their form. Recall that we say that x is a bad point at time t if there are at least two distinct optimal trajectories of

$$u(t,x) = \inf_{\substack{\xi \in \mathcal{AC} \\ \xi(t)=x}} \left\{ V_b(\xi(0)) + \int_0^t \mathcal{L}(\xi(s), \dot{\xi}(s)) \mathrm{d}s \right\}.$$

Regarding the case where we heat up a low temperature profile, we have the following conjecture.

Conjecture 5.6.1. Let a be the parameter of the dynamics, and b parameter of the initial rate function: $u_0 = V_b$. Suppose that $b > 0$ and $b > a$.

There are times $0 < t_0 \leq t^*$, where t^* is the solution of $e^{\frac{1}{2}at} = bt$, such that:

(a) For $t \in [0, t_0)$ there is no bad point.

(b) For $t \in [t_0, t^*)$ there are exactly two bad points $x_{bad}(t) > 0$ and $-x_{bad}(t)$.

(c) For $t \geq t^*$ there is a unique bad point $x_{bad} = 0$.

As $t \uparrow t^*$, we have $x_{bad}(t) \to 0$.

Note that (a) follows immediately from the fact the Hamiltonian flow is smooth that the starting curve F_b has a locally bounded derivative. Also (c) of the conjecture can be proven with the methods of the proof of Theorem

5.5.1 (c). The remaining question is whether t_0 and t^* are equal. This seems to be a model dependent issue. It should be noted that for Glauber dynamics on the Curie-Weiss model, the results in Ermolaev and Külske [2010] imply $t_0 \neq t^*$.

In the setting where we cool down a low temperature profile, we have the following conjecture, based on the argument for (d) of Theorem 5.5.1.

Conjecture 5.6.2. Let a be the parameter of the dynamics, and b parameter of the initial rate function: $u_0 = V_b$. Suppose that $b > 0$ and $a > b$.

There is a time $0 < t^*$ such that:

(a) For $t < t^*$ there is no bad point.

(b) For $t \geq t^*$ there are exactly two bad points $x_{bad}(t) \in (0, \sqrt{a})$ and $-x_{bad}(t) \in (-\sqrt{a}, 0)$.

As $t \to \infty$, we have $x_{bad}(t) \to x_{bad}(\infty)$, where $x_{bad}(\infty) = \sqrt{a-b}$ is the solution of $V_a(x) = V_b(0)$ in the interval $(0, \sqrt{a})$.

We give a short summary of the (partially proven) conjectures. Let b the parameter of the starting position and let a be the parameter of the dynamics. We say that a magnetization x is Gibbs at time t if $u(t, \cdot)$ is differentiable at x and we say that x is bad if this is not the case. We find

(a) $b \leq 0, a \in \mathbb{R}$ (high temperature starting configuration): All x are Gibbs for all times $t \geq 0$.

(b) $b < 0, a \leq b$ (low temperature starting configuration that is heated up) There are two times $t_0 \leq t^*$ such that all x are Gibbs for $t < t_0$. For $t \in [t_0, t^*)$ there are two bad magnetization $x_{bad}(t)$ and $-x_{bad}(t)$. For $t \geq t^*$, 0 is a pad point, and additionally, $x_{bad}(t) \to 0$ as $t \uparrow t^*$.

(c) $b < 0$, $a > b$ (low temperature starting configuration that is cooled down) There is a time $0 < t^*$, such that all x are Gibbs for $t < t_0$, and there are two bad magnetizations $x_{bad}(t)$ and $-x_{bad}(t)$ for $t \geq t^*$. Furthermore, we have $x_{bad}(t) \to \sqrt{a-b}$ as $t \to \infty$.

Note that the general picture coincides with the results in Ermolaev and Külske [2010].

6 THE LARGE DEVIATION PRINCIPLE FOR THE TRAJECTORY OF THE EMPIRICAL DISTRIBUTION OF A FELLER PROCESS

In this chapter, we reproduce results proved in:

Richard Kraaij. Large deviations of the trajectory of empirical distributions of Feller processes on locally compact spaces. *preprint; ArXiv:1401.2802*, 2014.

In particular, we give a proof of the large deviation principle for trajectories of empirical averages of independent copies of Feller processes on some space E without explicitly specifying the structure of the underlying process. Additionally, we express the rate function in terms of a Lagrangian.

The independence assumption implies that the large deviation principle can be proven via Sanov's theorem and the contraction principle. Also, we can explicitly give the limiting non-linear semigroup $V(t)$ on E as $\log S(t)e^f$ where $S(t)$ is the semigroup of conditional expectations of the Markov process. This approach avoids the difficult problem of constructing a semigroup which we encountered in Chapter 3.

To obtain a Lagrangian form of the rate function, the main technical challenge is to show that $V(t)$ equals a Nisio semigroup $\mathbf{V}(t)$. The definition of the Nisio-semigroup as in Section 2.4.2 poses us with two problems. First, we need a context-independent way to define absolutely continuous trajectories of measures, and secondly, we need a way to define a Lagrangian. To this end, we assume the existence of a suitable topology on a core of the generator $(A, \mathcal{D}(A))$ of the Feller process. The equality of $V(t)$ and $\mathbf{V}(t)$ is then proven using resolvent approximation arguments and Doob-h transform techniques.

The rest of the chapter is organised as follows. We start out in Section 6.1 with the preliminaries and state the two main theorems. Theorem 6.1.1 gives, under the condition that the processes solves the martingale problem, the large deviation principle. Under the condition that there exists a suitable core for the generator of the process, Theorem 6.1.8 gives the decomposition of the rate function.

In Section 6.2, we prove Theorem 6.1.1 using Sanov's theorem for large deviations on the Skorokhod space and the contraction principle. We show that the rate function is given by a rate for the initial law, and a second term that is given as the supremum over sums of conditional large deviation rate functions. The Legendre transforms of such conditional rate functions is given in terms of the non-linear semigroup $V(t)$. Additionally, we give a short introduction to the Doob-h transform, which we will use to study the non-linear semigroup

In Sections 6.3 and 6.4, we prove Theorem 6.1.8. In the first section, we study the Hamiltonian, Lagrangian and a family of 'controlled' generators. Finally, in Section 6.4, we introduce the Nisio semigroup $\mathbf{V}(t)$ in terms of absolutely continuous trajectories and the Lagrangian, and show that it equals the non-linear semigroup $V(t)$.

In Section 6.5, we give three examples where Theorem 6.1.8 applies. We start with a Markov jump process. After that, we check the conditions for spatially extended interacting particle systems of the type that are found in Liggett [1985]. Lastly, we check the conditions for a class of diffusion processes and show that, at least if the process is time-homogeneous and the diffusion and drift coefficients are sufficiently smooth, we recover the result for averages of independent and time-homogeneous processes by Dawson and Gärtner [1987].

6.1 PRELIMINARIES AND MAIN RESULTS

We follow the notation as in Chapter 2. As before, (E, d) is a complete separable metric space and on E we have a time-homogeneous Markov process $\{X(t)\}_{t \geq 0}$ given by a path space measure $\mathbb{P}$ on $D_E(\mathbb{R}^+)$. Let $X^1, X^2, \ldots$ be independent copies of X and let P the measure that governs these processes. We look at behaviour of the sequence $L_n := \left\{ L_n^{X(t)} \right\}_{t \geq 0}$,

$$L_n^{X(t)} := \frac{1}{n} \sum_{i=1}^{n} \delta_{\{X^i(t)\}},$$

under the law P. L_n takes values in $D_{\mathcal{P}(E)}(\mathbb{R}^+)$, the Skorokhod space of paths taking values in $\mathcal{P}(E)$. We also consider $C_{\mathcal{P}(E)}(\mathbb{R}^+)$ the space of continuous paths on $\mathcal{P}(E)$ with the topology inherited from $D_{\mathcal{P}(E)}(\mathbb{R}^+)$.

By $S(\nu \mid \mu)$ we denote the relative entropy of ν with respect to ν:

$$S(\nu \mid \mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu & \text{if } \nu << \mu, \\ \infty & \text{otherwise.} \end{cases}$$

In Section 6.2, we obtain the following preliminary result.

Theorem 6.1.1. *Let X, represented by the measure $\mathbb{P}$ on $D_E(\mathbb{R}^+)$ solve the martingale problem for $(A, \mathcal{D}(A))$ with starting measure $\mathbb{P}_0$. Then, the sequence L_n satisfies the large deviation principle with good rate function I, which is given for $\nu = \{\nu(t)\}_{t \geq 0} \in D_{\mathcal{P}(E)}(\mathbb{R}^+)$ by*

$$I(\nu) = \begin{cases} S(\nu(0) \mid \mathbb{P}_0) + \sup_{\{t_i\}} \sum_{i=1}^{k} I_{t_i - t_{i-1}}(\nu(t_i) \mid \nu(t_{i-1})) & \\ & \text{if } \nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+), \\ \infty & \text{otherwise,} \end{cases}$$

where $\{t_i\}$ is a finite sequence of times: $0 = t_0 < t_1 < \cdots < t_k$. For $s \leq t$, we have

$$I_t(\nu_2 \mid \nu_1) = \sup_{f \in C_b(E)} \left\{ \langle f, \nu_2 \rangle - \langle V(t) f, \nu_1 \rangle \right\}, \tag{6.1.1}$$

where $V(t)f(x) = \log \mathbb{E}\left[e^{f(X(t))} \,\middle|\, X(0) = x \right]$.

For further results, we introduce some additional notation. For a locally convex space $(\mathcal{X}, \tau)$, we write $\mathcal{X}'$ for its continuous dual space and $\mathcal{L}(\mathcal{X}, \tau)$ for the space of linear and continuous maps from $(\mathcal{X}, \tau)$ to itself. Also, for $x \in \mathcal{X}$ and $x' \in \mathcal{X}'$, we write $\langle x, x' \rangle := x'(x) \in \mathbb{R}$ for the natural pairing between x and x'. For two locally convex spaces $\mathcal{X}, \mathcal{Y}$ and a continuous linear operator $T : \mathcal{X} \to \mathcal{Y}$, we write $T' : \mathcal{Y}' \to \mathcal{X}'$ for the adjoint of T, which is uniquely defined by $\langle x, T'(y') \rangle = \langle Tx, y' \rangle$, see for example Treves [Treves, 1967, Chapter 19]. For a neighbourhood $\mathcal{N}$ of 0 in $\mathcal{X}$, we define the polar of $\mathcal{N}^\circ \subset \mathcal{X}'$ by

$$\mathcal{N}^\circ := \left\{ u \in \mathcal{X}' \,\middle|\, |\langle x, u \rangle| \leq 1 \text{ for every } x \in \mathcal{N} \right\}. \tag{6.1.2}$$

We say that a locally convex space $\mathcal{X}$ is barrelled if every barrel is a neighbourhood of 0. A set S is a barrel if it is convex, balanced, absorbing and closed. S is balanced if we have the following: if $x \in S$ and $\alpha \in \mathbb{R}$, $|\alpha| \leq 1$ then $\alpha x \in S$. S is absorbing if for every $x \in \mathcal{X}$ there exists a $r \geq 0$ such that if $|\alpha| \geq r$ then $x \in \alpha S$. We give some background and basic results on barrelled spaces in Appendix 6.7.

To rewrite the rate function obtained in Theorem 6.1.1, we restrict to locally compact metric spaces (E, d) and we consider the situation where $S(t)f(x) = E[f(X(t)) \,|\, X(0) = x]$ is a strongly continuous semigroup on the space $(C_0(E), \|\cdot\|)$ of continuous functions that vanish at infinity equipped with the supremum norm. To be precise: for every $t \geq 0$ the map $S(t) : (C_0(E), \|\cdot\|) \to (C_0(E), \|\cdot\|)$ is continuous, and for every $f \in C_0(E)$, the trajectory $t \mapsto S(t)f$ is continuous in $(C_0(E), \|\cdot\|)$.

Let $(A, \mathcal{D}(A))$ be the generator of the semigroup $S(t)$. It is a well known result that X solves the martingale problem for $(A, \mathcal{D}(A))$ [Ethier and Kurtz, 1986, Proposition 4.1.7], so the above result holds for the process $\{X(t)\}_{t\geq 0}$.

Our goal is to rewrite I as

$$I(\nu) = S(\nu(0) \,|\, \mathbb{P}_0) + \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s$$

for a trajectory ν of probability measures that is absolutely continuous in some sense. Thus our first problem is to define differentiation in a context for which no suitable structure on E or $\mathcal{P}(E)$ is known. Therefore, we will have to tailor the definition of differentiation to the process itself. Suppose that $\mu(t)$ is the law of $X(t)$ under $\mathbb{P}$. Then we know that $t \mapsto \mu(t) = S(t)'\mu(0)$ is a weakly continuous trajectory in $\mathcal{P}(E)$, so can ask whether for $f \in \mathcal{D}(A)$ the trajectory $t \mapsto \langle f, \mu(t)\rangle$ is differentiable as a function from $\mathbb{R}^+ \to \mathbb{R}$:

$$\frac{\partial}{\partial t}\langle f, \mu(t)\rangle = \frac{\partial}{\partial t}\langle S(t)f, \mu(0)\rangle = \langle S(t)Af, \mu(0)\rangle = \langle Af, \mu(t)\rangle. \quad (6.1.3)$$

So our candidate for $\dot{\mu}(t)$ would be $A'\mu(t)$, which is a problematic because $(A, \mathcal{D}(A))$ could be unbounded. To overcome this, and other problems, we introduce two sets of conditions on $(A, \mathcal{D}(A))$.

Recall that D is a core for $(A, \mathcal{D}(A))$ if for every $f \in \mathcal{D}(A)$, we can find a sequence $f_n \in D$ such that $f_n \to f$ and $Af_n \to Af$.

Condition 6.1.2. There exists a core $D \subseteq \mathcal{D}(A)$ dense in $(C(E), \|\cdot\|)$ that satisfies

(a) D is an algebra, i.e. if $f, g \in D$ then $fg \in D$,

(b) if $f \in D$ and $\phi : \mathbb{R} \to \mathbb{R}$ a smooth function on the closure of range of f, then $\phi \circ f - \phi(0) \in D$, In the case that E is compact, $C_0(E) = C(E)$, then (b) can be replaced by

(b') if $f \in D$ and $\phi : \mathbb{R} \to \mathbb{R}$ a smooth function on the range of f, then $\phi \circ f \in D$.

Note if a dense subspace $D \subseteq \mathcal{D}(A)$ satisfies $S(t)D \subseteq D$, then D is a core for $(A, \mathcal{D}(A))$ [Ethier and Kurtz, 1986, Proposition 1.3.3].

Under Condition 6.1.2, we define the operator $H : D \to C_0(E)$ and for every $g \in D$ the operator $A^g : D \to C_0(E)$ by

$$\begin{aligned} Hf &= e^{-f} A e^{f}, \\ A^g f &= e^{-g} A(f e^{g}) - (e^{-g} A e^{g}) f. \end{aligned}$$

If E is non-compact, these definitions needs some care as $e^f \notin C_0(E)$. This can be solved by looking at the one-point compactification of E, see Section 6.3.1. In Section 6.3, we will show that $\{V(t)\}_{t \geq 0}$ turns out to be a non-linear semigroup on $C_0(E)$ which has H as its generator. The operators A^g are generators of Markov processes with law $\mathbb{Q}^g$ on $D_E([0,t])$ that are obtained from $\mathbb{P}$ by

$$\frac{\mathrm{d}\mathbb{Q}_t^g}{\mathrm{d}\mathbb{P}_t}(X) = \exp\left\{ g(X(t)) - g(X(0)) - \int_0^t Hg(X(s))\mathrm{d}s \right\}, \quad (6.1.4)$$

where $\mathbb{P}_t$ and $\mathbb{Q}_t^g$ are the measures $\mathbb{P}$ and $\mathbb{Q}^g$ restricted to times up to t, see Theorem 4.2 in Palmowski and Rolski [2002].

Condition 6.1.3 (Conditions on the core)**.** D satisfies Condition 6.1.2 and there exists a topology τ_D on D such that

(a) (D, τ_D) is a separable barrelled locally convex Hausdorff space.

(b) The topology τ_D is finer than the sup norm topology restricted to D.

(c) The maps $\exp -1 : (D, \tau_D) \to (D, \tau_D)$ and $\times : (D, \tau_D) \times (D, \tau_D) \to (D, \tau_D)$, defined by $f \mapsto e^f - 1$, respectively $(f, g) \mapsto fg$ are continuous.

(d) We have $S(t)D \subseteq D$, $V(t)D \subseteq D$ and the semigroups $\{S(t)\}_{t \geq 0}$ and $\{V(t)\}_{t \geq 0}$ restricted to D are strongly continuous for (D, τ_D).

(e) The map $A : (D, \tau_D) \to (C_0(E), \|\cdot\|)$ is continuous.

(f) There exists a barrel $\mathcal{N} \subseteq D$ such that

$$\sup_{f \in \mathcal{N}} \|Hf\| \leq 1,$$

and for every $c > 0$

$$\sup_{f \in c\mathcal{N}} \|Hf\| < \infty.$$

Conditions (a) to (e) are related to the differentiation of the trajectories of measures that will turn up in our large deviation problem. Condition (a) implies that (D, τ_D) is well behaved as a locally convex space and, among other things, makes sure that we are able to define the Gelfand integral, see Appendix 6.7. Condition (b) implies that $(\mathcal{M}(E), wk)$ is continuously embedded in (D', wk^*), so that every weakly continuous trajectory of measures can in fact be seen as a weak* continuous trajectory in D'. Important is this light is that the conditions in (d) on $\{S(t)\}_{t\geq 0}$ imply that $t \mapsto S(t)'\mu$ is weak* continuous in D' for all measures $\mu \in \mathcal{P}(E)$. (e) implies that we take the adjoint of $A : (D, \tau_D) \to (C_0(E), \|\cdot\|)$, so that we obtain a weak to weak* continuous map $A' : \mathcal{M}(E) \to D'$. Returning to (6.1.3), we now have a good definition for $\dot{\mu}(t) := A'\mu(t) \in D'$. Furthermore, we can also differentiate trajectories of measures that are obtained from X via a tilting procedure, e.g. (6.1.4), by Lemma 6.1.5 below. Condition (f) is the main technical condition on H which implies for example the compactness of the level sets of $\mathcal{L}$. Using these compactness arguments, we are able to rewrite I.

Remark 6.1.4. Condition (d) is removed in the latest version of Kraaij [2014]. The condition is used to prove one of the inequalities that imply that $V(t)$ equals $\mathbf{V}(t)$, see Proposition 6.4.10. In Kraaij [2014], a new proof is given in which this inequality is obtained via an approximation of the Doob-h transform by Markov processes that have generators that are piecewise constant in time.

The removal of this condition increases the set of examples to which the main result can be applied. For example, we can relax the conditions on the diffusion process in Section 6.5.3. Additionally, it can be used for diffusion processes on compact manifolds or Lévy processes on $\mathbb{R}^d$.

The following lemma is a consequence of Condition 6.1.3 (c) and (e) and the proof is elementary.

Lemma 6.1.5. *Let (D, τ_D) satisfy Condition 6.1.3, then the maps $\mathcal{A}$: $(D, \tau_D) \times (D, \tau_D) \to (C_0(E), \|\cdot\|)$ given by $\Phi(g, f) = A^g f$ and the operator $H : (D, \tau_D) \to (C_0(E), \|\cdot\|)$ are continuous.*

Remark 6.1.6. The results of this chapter also hold in the case that Condition 6.1.3 (c) fails as long as the conclusions of Lemma 6.1.5 hold. In all examples that we consider in Section 6.5 (c) is satisfied.

For the next definition we will need the Gelfand or weak* integral, which is introduced in Appendix 6.7, but the rigorous construction of this integral can be skipped on the first reading.

Definition 6.1.7. Define $D - \mathcal{AC}$, or if there is no chance of confusion, $\mathcal{AC}$, the space of absolutely continuous paths in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. A path $\nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+)$ is called absolutely continuous if there exists a (D', wk^*) measurable curve $s \mapsto u(s)$ in D' with the following properties:

(i) for every $f \in D$ and $t \geq 0$ $\int_0^t |\langle f, u(s)\rangle| \mathrm{d}s < \infty$,

(ii) for every $t \geq 0$, $\nu(t) - \nu(0) = \int_0^t u(s)\mathrm{d}s$ as a D' Gelfand integral.

We denote $\dot{\nu}(s) := u(s)$. Furthermore, we will denote $\mathcal{AC}_\mu$ for the space of absolutely continuous trajectories starting at μ_0, and $\mathcal{AC}^T$ for trajectories that are only considered up to time T. Similarly, we define $\mathcal{AC}_\mu^T$.

A direct consequence of property (ii) is that for almost every time $t \geq 0$ and all $f \in D$ the limit

$$\lim_{h \to 0} \frac{\langle f, \nu(t+h)\rangle - \langle f, \nu(t)\rangle}{h}$$

exists and is equal to $\langle f, \dot{\nu(t)}\rangle$. This justifies the notation $u(s) = \dot{\nu}(s)$.

Using these definitions, we are able to sharpen Theorem 6.1.1. In Section 6.3, we study the semigroup $V(t)$ and its generator H. Also, we give a number of properties of the level sets of the Lagrangian $\mathcal{L}$, defined in the theorem below. The proof of the theorem is given in Section 6.4.

Theorem 6.1.8. *Let (E, d) be locally compact. Let $(A, \mathcal{D}(A))$ have a core D equipped with a topology τ_D such that (D, τ_D) satisfies Condition 6.1.3. Then, the rate function in Theorem 6.1.1 can be rewritten as*

$$I(\nu) = \begin{cases} S(\nu(0) \,|\, \mathbb{P}_0) + \int_0^\infty \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s & \text{if } \nu \in \mathcal{AC}_{\nu(0)} \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{L} : \mathcal{P}(E) \times D' \to [0, \infty]$ is given by

$$\mathcal{L}(\mu, u) := \sup_{f \in D} \left\{ \langle f, u\rangle - \langle Hf, \mu\rangle \right\}.$$

Remark 6.1.9. If we restrict ourselves to $[0, T]$ instead of $\mathbb{R}^+$, then we obtain

$$I^T(\{\nu(s)\}_{0 \leq s \leq T}) = \begin{cases} H(\nu(0) \,|\, \mathbb{P}_0) + \int_0^T \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s & \text{if } \nu \in \mathcal{AC}_{\nu(0)}^T \\ \infty & \text{otherwise,} \end{cases}$$

by applying the contraction principle.

6.2 THE LARGE DEVIATION PRINCIPLE VIA SANOV'S THEOREM AND OPTIMAL TRAJECTORIES

Let (E, d) is a complete separable metric space. We start by proving the large deviation principle for a general class of processes via Sanov's theorem and the contraction principle. This will lead to the proof of Theorem 6.1.1.

Define for every t the map $\pi_t : D_E(\mathbb{R}^+) \to E$ by $\pi_t(x) := x(t)$. By Proposition III.7.1 in Ethier and Kurtz, π_t is a measurable map. Complementary to π_t, we introduce the map π_{t-}. For every path $x \in D_E(\mathbb{R}^+)$, the value $x(t-) := \lim_{s\uparrow t} x(s)$ is well defined, which makes it possible to define $\pi_{t-} : D_E(\mathbb{R}^+) \to E$ by $\pi_{t-}(x) := x(t-)$. Because π_{t-} is the point-wise limit of the measurable maps π_{t_n}, for $t_n < t$, $t_n \uparrow t$, also π_{t-} is measurable.

Let $\mathbb{P}$ be a probability measure on $D_E(\mathbb{R}^+)$, and let $X = (X(t))_{t\geq 0}$ be the process with law $\mathbb{P}$. Define $\mu(t) = \mathbb{P} \circ \pi_t^{-1}$ and $\mu(t-) = \mathbb{P} \circ \pi_{t-}^{-1}$ the laws of $X(t)$ and $X(t-)$. Also define the map $\phi : \mathcal{P}(D_E(\mathbb{R}^+)) \to \mathcal{P}(E)^{\mathbb{R}^+}$ by setting $\phi(\mathbb{P}) = (\mu(t))_{t\geq 0}$ and finally define the maps $\phi_t : \mathcal{P}(D_E(\mathbb{R}^+)) \to \mathcal{P}(E)$ by setting $\phi_t(\mathbb{P}) = \mu(t)$.

Lemma 6.2.1. *ϕ is a map from $\mathcal{P}(D_E(\mathbb{R}^+))$ to $D_{\mathcal{P}(E)}(\mathbb{R}^+)$.*

Proof. First, we prove that if $s \downarrow t$ then $\mu(s) \to \mu(t)$ weakly. Because the paths of X are right-continuous, we have $X(s) \to X(t)$. Hence, we have a.s. convergence, which in turn implies that $\mu(s) \to \mu(t)$ weakly.

If $s \uparrow t$, then we need to show that $\lim_{s\uparrow t} \mu(t)$ exists, but as above $X(s) \to X(t-)$, hence, the weak limit $\lim_{s\uparrow t} \mu(s)$ is equal to $\mu(t-)$. □

We would like to prove that ϕ and $\{\phi_t\}_{t\geq 0}$ are continuous maps, but this is not always true as can be seen from the following example.

Example 6.2.2. Pick two distinct points e_1, e_2 in E. Define

$$x_n^+(t) = \begin{cases} e_1 & \text{for } t < 1 + 1/n \\ e_2 & \text{for } t \geq 1 + 1/n \end{cases}$$

$$x_n^-(t) = \begin{cases} e_1 & \text{for } t < 1 - 1/n \\ e_2 & \text{for } t \geq 1 - 1/n \end{cases}$$

and let $\mathbb{P}^n \in \mathcal{P}(D_E(\mathbb{R}^+))$ be defined by $\mathbb{P}^n = \frac{1}{2}\delta_{x_n^+} + \frac{1}{2}\delta_{x_n^-}$. Clearly, the sequence $\mathbb{P}^n$ converges weakly to $\tilde{\mathbb{P}} = \delta_{\tilde{x}}$ where $\tilde{x}(t)$ is equal to e_1 for $t < 1$ and e_2 for $t \geq 1$.

If we look at the images $\phi(\mathbb{P}^n) = (\mu^n(t))_{t\geq 0}$ and $\phi(\tilde{\mathbb{P}}) = (\tilde{\mu}(t))_{t\geq 0}$, then we obtain

$$\mu^n(t) = \begin{cases} \delta_{e_1} & \text{for } t < 1 - 1/n \\ \frac{1}{2}\delta_{e_1} + \frac{1}{2}\delta_{e_2} & \text{for } 1 - 1/n \leq t < 1 + 1/n \\ \delta_{e_2} & \text{for } 1 + 1/n \leq t, \end{cases}$$

$$\tilde{\mu}(t) = \begin{cases} \delta_{e_1} & \text{for } t < 1 \\ \delta_{e_2} & \text{for } t \geq 1. \end{cases}$$

Clearly, $\mu^n(1) \to \frac{1}{2}\delta_{e_1} + \frac{1}{2}\delta_{e_2}$, which is not equal to $\tilde{\mu}(1)$ or $\tilde{\mu}(1-)$. We obtain that both ϕ and ϕ_1 are not continuous. Obviously, it follows that ϕ_t for $t \geq 0$ are not continuous either.

So problems arise when the time marginals of the limiting measure $\mathbb{P}$ are discontinuous in time. However, this is the only thing that can happen.

Proposition 6.2.3. $\phi : \mathcal{P}(D_E(\mathbb{R}^+)) \to D_{\mathcal{P}(E)}(\mathbb{R}^+)$ *is continuous at measures* $\mathbb{P}$ *for which it holds that for every* $t > 0$*:* $\mathbb{P}[X(t) = X(t-)] = 1$.

A similar statement for the finite dimensional projections ϕ_t, can be found in Ethier and Kurtz [Ethier and Kurtz, 1986, Theorem 3.7.8].

Proof. Let $\mathbb{P}^n, \mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ such that $\mathbb{P}^n \to \mathbb{P}$ weakly and $\mathbb{P}$ such that for every t $\mathbb{P}[X(t) = X(t-)] = 1$. By the Skorokhod representation Theorem [Ethier and Kurtz, 1986, Theorem 3.1.9], we can find a probability space $(\Omega, \mathcal{F}, P)$ and $D_E(\mathbb{R}^+)$ valued random variables Y^n, Y distributed as X^n and X under $\mathbb{P}^n, \mathbb{P}$ such that $Y^n \to Y$ P a.s.

Let $\{t_n\}_{n\geq 0}$ be a sequence converging to $t > 0$. Define the sets

$$A := \{Y(t) = Y(t-)\},$$
$$B := \{d(Y^n(t_n), Y(t)) \wedge d(Y^n(t_n), Y(t-)) \to 0\}.$$

By the assumption that $\mathbb{P}[X(t) = X(t-)] = 1$, it follows that $P[A] = 1$. By Proposition 3.6.5 in Ethier and Kurtz [1986], and the fact that $Y^n \to Y$ P a.s. it follows that $P[B] = 1$. Combining these statements yields

$$P[Y^n(t_n) \to Y(t)] \geq P[A \cap B] = 1,$$

which implies that $\mu^n(t_n) \to \mu(t)$. Because $\mu(t) = \mu(t-)$ by assumption, Proposition 3.6.5 in Ethier and Kurtz yields the final result. □

6.2.1 *Large deviations for measures on the Skorokhod space*

Suppose that we have a process X on $D_E(\mathbb{R}^+)$ and a corresponding measure $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$. Then Sanov's theorem, Theorem 6.2.10 in Dembo and Zeitouni [1998], gives us the large deviation behaviour of the empirical distribution L_n^X of independent copies of the process X: $X^1, X^2, \ldots$:

$$L_n^X := \frac{1}{n}\sum_{i=1}^n \delta_{\{X^i\}} \in \mathcal{P}(D_E(\mathbb{R}^+)).$$

Theorem 6.2.4 (Sanov)**.** *The empirical measures L_n^X satisfy the large deviation principle on $\mathcal{P}(D_E(\mathbb{R}^+))$ with respect to the weak topology with the good and convex rate function*

$$I^*(\mathbb{Q}) = S(\mathbb{Q} \mid \mathbb{P}) := \int \frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

We are interested in obtaining a large deviation principle on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$. In Proposition 6.2.3, we saw that we have a map ϕ that is continuous on a part of its domain. Hence, we we are in the position to use the contraction principle.

Theorem 6.2.5. *Suppose that $\mathbb{P}$ satisfies $\mathbb{P}[X(t) = X(t-)] = 1$ for every $t \geq 0$, then the large deviation principle holds for*

$$\left(L_n^{X(t)}\right)_{t\geq 0} = \left(\frac{1}{n}\sum_{i=1}^n \delta_{X^i(t)}\right)_{t\geq 0}$$

on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with rate function

$$I((\nu_t)_{t\geq 0}) = \inf\{S(\mathbb{Q} \mid \mathbb{P}) \mid \mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)), \phi(\mathbb{Q}) = (\nu(t))_{t\geq 0}\}$$

and I is finite only on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. The measures $\mathbb{Q}$ for which it holds that $I(\mathbb{Q}) < \infty$ satisfy $\mathbb{Q} << \mathbb{P}$ hence it follows that for every t: $\mathbb{Q}[X(t) = X(t-)] = 1$. This yields that ϕ is continuous at $\mathbb{Q}$ by Proposition 6.2.3.

By the contraction principle, Theorem 4.2.1 and remark (c) after Theorem 4.2.1 in Dembo and Zeitouni Dembo and Zeitouni [1998], we obtain the large deviation principle on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with I as given in the theorem. □

6.2.2 *The large deviation principle for Markov processes*

Although Theorem 6.2.5 can be applied to a wide range of (time-inhomogeneous) processes, we explore its consequences for time-homogeneous Markov processes. Recall the definition of a solution to the martingale problem preceding Theorem 6.1.1.

Lemma 6.2.6. *Suppose that the process X with corresponding measure $\mathbb{P}$ on $D_E(\mathbb{R}^+)$ solves the martingale problem for $(A, \mathcal{D}(A))$ with starting measure $\mathbb{P}_0$. Then, it holds that for every $t \geq 0$ $\mathbb{P}[X(t) = X(t-)] = 1$. Hence, the large deviation principle holds for $\{L_n^{X(t)}\}_{t\geq 0}$ on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with rate function*

$$I((\nu_t)_{t\geq 0}) = \inf\{S(\mathbb{Q} \,|\, \mathbb{P}) \,|\, \mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)), \phi(\mathbb{Q}) = (\nu(t))_{t\geq 0}\}$$

and I is finite only on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. To apply Theorem 6.2.5, we need to check that $\mathbb{P}[X(t) = X(t-)] = 1$ for every $t \geq 0$, but this follows by Theorem 4.3.12 in Ethier and Kurtz [1986]. □

Using this result, Theorem 6.1.1 follows without much effort.

Proof of Theorem 6.1.1. The large deviation principle follows by Lemma 6.2.6. This lemma also gives that the rate function is ∞ on the complement of $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

To obtain the rate function as a supremum over rate functions for finite dimensional problems

$$I(\nu) = \begin{cases} \sup_{0,t_1,\ldots,t_k} I[0,t_1,\ldots,t_k](\nu(0),\nu(t_1)\ldots,\nu(t_k)) & \\ & \text{if } \nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+), \\ \infty & \text{otherwise,} \end{cases}$$

we use Theorem 4.13 and Theorem 4.30 in Feng and Kurtz [2006]. Proposition 6.6.3 gives us the final decomposition of the rate function. □

6.2.3 *The semigroup $V(t)$ and the Doob-h transform*

Before we turn to the proof of Theorem 6.1.8, we start with rewriting $V(t)f$ in terms of the Doob transform. We have the following useful variant of Lemma 2.19 in Seppäläinen [1993].

Lemma 6.2.7. *Let $h \in C(E)$ and let $t > 0$. Set*

$$S^{\mathbb{P}_0}(\mathbb{Q}) = \begin{cases} S(\mathbb{Q} \,|\, \mathbb{P}) & \textit{if } \mathbb{Q}_0 = \mathbb{P}_0, \\ \infty & \textit{otherwise.} \end{cases}$$

Then,

$$\langle V(t)h, \mathbb{P}_0 \rangle = \sup_{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))} \left\{ \langle h, \mathbb{Q}_t \rangle - S^{\mathbb{P}_0}(\mathbb{Q}) \right\},$$

where $\mathbb{Q}_t$ denotes the time t marginal of $\mathbb{Q}$. The supremum is attained by the measure $\mathbb{Q}^h$ defined by

$$\frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}}(X) = \frac{e^{h(X(t))}}{\langle e^h, \mathbb{P}_t \rangle} = e^{h(X(t)) - \langle V(t)h, \mathbb{P}_0 \rangle}$$

Proof. Let $\mathbb{P}_{0,t} \in \mathcal{P}(E^2)$ be the restriction of $\mathbb{P}$ to the time 0 and time t marginals. As before, we denote by $\mathbb{P}_0$ the time 0 marginal of $\mathbb{P}$ and for a measure $\nu \in \mathcal{P}(E^2)$ we denote by ν_0 respectively ν_1 the restriction to the first marginal and second marginal. Set

$$S_t^{\mathbb{P}_0}(\nu) = \begin{cases} S(\nu \,|\, \mathbb{P}_{0,t}) & \text{if } \nu_0 = \mathbb{P}_0, \\ \infty & \text{otherwise.} \end{cases}$$

By Lemma 2.19 in Seppäläinen [1993] and convex duality, we obtain

$$\langle V(t)h, \mathbb{P}_0 \rangle = \sup_{\nu \in \mathcal{P}(E^2)} \left\{ \langle h, \nu_2 \rangle - S_t^{\mathbb{P}_0}(\nu) \right\}.$$

By the contraction principle, we have

$$S(\nu \,|\, \mathbb{P}_{0,t}) = \inf \left\{ S(\mathbb{Q} \,|\, \mathbb{P}) \,|\, \mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \,:\, \mathbb{Q}_{0,t} = \nu \right\},$$

which implies that

$$\langle V(t)h, \mu \rangle = \sup_{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))} \left\{ \langle h, \mathbb{Q}_t \rangle - S^{\mathbb{P}_0}(\mathbb{Q}) \right\}.$$

Now we show that the supremum is achieved for $\mathbb{Q}^h$ defined by

$$\frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}}(X) = \frac{e^{h(X(t))}}{\langle e^h, \mathbb{P}_t \rangle} = e^{h(X(t)) - \langle V(t)h, \mathbb{P}_0 \rangle}.$$

Note that $\mathbb{Q}_0^h = \mathbb{P}_0$. Therefore, we obtain that

$$\begin{aligned}\langle h, \mathbb{Q}_t^h \rangle - S^{\mathbb{P}_0}(\mathbb{Q}^h) &= \langle h, \mathbb{Q}_t^h \rangle - \int \log \frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{Q}^h \\ &= \langle h, \mathbb{Q}_t^h \rangle - \langle h, \mathbb{Q}_t^h \rangle + \langle V(t)h, \mathbb{P}_0 \rangle = \langle V(t)h, \mathbb{P}_0 \rangle.\end{aligned}$$

□

The optimising measure $\mathbb{Q}^h$ defined in the lemma above has the form of a Doob-h transform, see Doob [Doob, 1984, page 566] or Jamison [1975], Föllmer and Gantert [1997]. We study the law of $\mathbb{Q}^h$. For $s \leq t$, define $h(s) = V(t-s)h$, or $e^{h(s)} = S(t-s)e^h$.

The transition probabilities of the Markov process described by $\mathbb{Q}^h$ up to time t can be written down as a semigroup of transition operators $\{S^{h[0,t]}(r,s)\}_{0 \leq r \leq s \leq t}$, where $S^{h[0,t]}(r,s) : C(E) \to C(E)$ is defined by $S^{h[0,t]}(r,s)f(x) := \mathbb{Q}^h[f(X(s)) \mid X(r) = x]$. The following result is obtained by a straightforward calculation.

Lemma 6.2.8. *The semigroup of transition probabilities of* $\mathbb{Q}^f$ *defined by*

$$\frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}}(X) = \frac{e^{h(X(t))}}{\langle e^h, \mathbb{P}_t \rangle} = e^{h(X(t)) - \langle V(t)h, \mathbb{P}_0 \rangle},$$

is given by

$$S^{h[0,t]}(r,s)f(x) = e^{-h(r)}(x) S(r,s) \left(f e^{h(s)} \right)(x).$$

To use this representation of $\mathbb{Q}^h$ to obtain a Lagrangian representation of the rate function, we first study the properties of the operators A^g, H and L.

6.3 A STUDY OF THE OPERATORS $V(t)$, H, L AND A^g.

6.3.1 *The semigroup* $V(t)$ *and the generator* H

We return to the situation that (E, d) is a locally compact metric space, so that we can use semigroup theory to rewrite the rate function.

First suppose that E is non-compact. Let $E^\Delta = E \cup \{\Delta\}$ be the one-point compactification. By Lemma 4.3.2 in Ethier and Kurtz [1986], $S(t)$

extends to a strongly continuous contraction semigroup on $(C(E^\Delta), \|\cdot\|)$ by setting $S^\Delta(t)f = f(\Delta) + S(t)(f - f(\Delta))$. Therefore, we can argue using the semigroup on the compact space E^Δ, and then obtain the result in Theorem 6.1.8 on E by Theorem 4.11 in Feng and Kurtz Feng and Kurtz [2006].

From this point onward, we assume that (E, d) is compact and that the transition semigroup $\{S(t)\}_{t\geq 0}$ is strongly continuous on $C(E)$. Let $A : \mathcal{D}(A) \subseteq C(E) \to C(E)$ be the associated infinitesimal generator.

We examine $V(t)f(x) = \log S(t)e^f(x) = \log \mathbb{E}\left[e^{f(X(t))} \,\middle|\, X(0) = x\right], f \in C(E)$, which was defined in Theorem 6.1.1. It is an elementary calculation to check that $\{V(t)\}_{t\geq 0}$ is a strongly continuous contraction semigroup on $C(E)$.

As in the linear case, we calculate the generator of V:

$$Hf = \lim_{t\downarrow 0} \frac{V(t)f - f}{t}$$

defined for $f \in \mathcal{D}(H)$, where

$$\mathcal{D}(H) := \left\{ f \in C(E) \,\middle|\, \exists g \in C(E) : \lim_{t\downarrow 0} \left\| \frac{V(t)f - f}{t} - g \right\| = 0 \right\}.$$

We start with an extension of the chain rule to Banach spaces. The proof is rather standard and is left to the reader.

Lemma 6.3.1. *Let $f \in \mathcal{D}(A)$ and let $\phi : f(E) \to \mathbb{R}$ be differentiable on $f(E)$ and let ϕ' be Lipschitz continuous. Then it holds that*

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(S(t)f)|_{t=0} = \phi'(f)Af,$$

which should be interpreted as

$$\lim_{t\to 0} \frac{\phi(S(t)f) - \phi(f)}{t} = \phi'(f)Af$$

with respect to the sup norm.

A direct consequence is that we can calculate the generator H of $V(t)$ on a subset of its domain.

Corollary 6.3.2. *For $f \in C(E)$ such that $e^f \in \mathcal{D}(A)$, we have $f \in \mathcal{D}(H)$ and*

$$Hf = e^{-f}A(e^f).$$

In order to proceed, we need Condition 6.1.2. We see that Corollary 6.3.2 gives us that if $f \in D$, then $f \in \mathcal{D}(H)$ and $Hf = e^{-f}Ae^f$.

We will use this operator (H, D), under Condition 6.1.3, to construct a new *Nisio* semigroup $\{\mathbf{V}(t)\}_{t\geq 0}$ on $C(\mathcal{P}(E))$. This semigroup will be introduced in Section 6.4.2, and there we will show that for $\mu \in \mathcal{P}(E)$ and $f \in C(E)$, we have $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu\rangle$, where $[f] \in C(\mathcal{P}(E))$ is the function defined by $[f](\mu) = \langle f, \mu\rangle$.

We start with some results on $V(t)f$ and H that will be useful for proving the equality $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu\rangle$. For $f \in C(E)$, let $J(\lambda)f := (\mathbb{1} - \lambda A)^{-1}f = \int_0^\infty \lambda^{-1}e^{-\lambda^{-1}t}S(t)f\mathrm{d}t$. Using $J(\lambda)$, we set $R(\lambda)f := \log J(\lambda)e^f$.

We constructed the semigroup $V(t)$ from the linear semigroup $S(t)$, and the operator $R(\lambda)$ from the linear resolvent $J(\lambda)$. One would therefore hope that $R(\lambda)$ equals $(\mathbb{1} - \lambda H)^{-1}$. This is not the case, but we do have the following two results, which we will need for the proof of Lemma 6.4.8 and Proposition 6.4.10.

Lemma 6.3.3. *For $f \in C(E)$, we have $R(\lambda)f \in \mathcal{D}(H)$ and $(\mathbb{1} - \lambda H)R(\lambda)f \geq f$.*

Proof. $J(\lambda)$ maps $C(E)$ bijectively on $\mathcal{D}(A)$, therefore, $e^{R(\lambda)f} = J(\lambda)e^f \in \mathcal{D}(A)$. Thus by Corollary 6.3.2, we have that $R(\lambda)f \in \mathcal{D}(H)$.

Let $x \in E$, we prove $(\mathbb{1} - \lambda H)\, R(\lambda)f(x) \geq f(x)$. We prove that the following quantity is larger than 0:

$$\begin{aligned}
&(\mathbb{1} - \lambda H)\, R(\lambda)f(x) - f(x) \\
&\qquad = R(\lambda)f(x) - f(x) - \lambda\frac{AJ(\lambda)e^f(x)}{J(\lambda)e^f(x)} \\
&\qquad = R(\lambda)f(x) - f(x) - \frac{J(\lambda)e^f(x) - e^{f(x)}}{J(\lambda)e^f(x)}.
\end{aligned}$$

This is equivalent to showing that

$$J(\lambda)e^f(x)\log\Big(J(\lambda)e^f(x)\Big) - f(x)J(\lambda)e^f(x) - J(\lambda)e^f(x) + e^{f(x)}$$

is positive, which follows from the fact that for every $c \in \mathbb{R}$, the function defined for non-negative y, given by $y \mapsto y\log y - (c+1)y + e^c$ is non-negative. □

Note that the fact that the function $y \mapsto y \log y - (c+1)y + e^c$ has a unique point where it hits 0. This means that $(\mathbb{1} - \lambda H)R(\lambda)f(x) = f(x)$ only if $\mathbb{E}[e^{f(X_\tau)} \mid X_0 = x] = e^{f(x)}$, where τ is an exponential random variable with mean λ independent of the process X. This can not be true in general.

Even though $R(\lambda)$ does not invert $(\mathbb{1} - \lambda H)$, it does approximate the semi-group in a way that the resolvents of H would as well.

Lemma 6.3.4. *For every $f \in C(E)$, we have that $\lim_{n\to\infty} R\left(n^{-1}\right)^{\lfloor nt \rfloor} f = V(t)f$.*

Proof. By definition, we have $R\left(n^{-1}\right)^{\lfloor nt \rfloor} f = \log J\left(n^{-1}\right)^{\lfloor nt \rfloor} e^f$. For linear semigroups, we know that the resolvents approximate the semigroup: $J\left(\frac{1}{n}\right)^{\lfloor nt \rfloor} e^f \to S(t)e^f$, see for example Corollary 1.6.8 in Ethier and Kurtz [1986]. Therefore, by uniform continuity of the logarithm on $[e^{-\|f\|}, e^{\|f\|}]$, we obtain the final result by applying the logarithm. □

6.3.2 *Operator duality for H*

Additionally to the operator H, we introduce operators A^g that serve as generators of tilted Markov processes obtained from $X(t)$ by the change of measure given in (6.1.4). We also introduce an operator L, that will serve as a precursor to our final Lagrangian $\mathcal{L}$.

Definition 6.3.5. Under Condition 6.1.2, define the following operators for $f, g \in D$:

$$
\begin{aligned}
Hf &= e^{-f} A e^f, \\
A^g f &= e^{-g} A(f e^g) - (e^{-g} A e^g) f, \\
Lg &= A^g g - Hg.
\end{aligned}
$$

H will be called the Hamiltonian and L the (pre-)Lagrangian in analogy to the Lagrangian and Hamiltonian of classical mechanics. A^g is a generator itself, see for example Palmowski and Rolski Palmowski and Rolski [2002]. This is also illustrated by the next two examples.

We calculate H and A^g in the case of a Markov jump process and a standard Brownian motion.

Example 6.3.6. Let E be a finite set and let $\{X(t)\}_{t\geq 0}$ be generated by

$$
Af(x) = \sum_y r(x,y)\left[f(y) - f(x)\right],
$$

where r is some transition kernel. A calculations shows that

$$
\begin{aligned}
Hf(x) &= \sum_y r(x,y)\left[e^{f(y)-f(x)} - 1\right], \\
A^g f(x) &= \sum_y r(x,y) e^{g(y)-g(x)} \left[f(y) - f(x)\right].
\end{aligned}
$$

Example 6.3.7. Let $E = \mathbb{R}$, and let $\{X(t)\}_{t\geq 0}$ be a standard Brownian motion, for which the generator A is given for $f \in C_c^\infty(\mathbb{R})$, i.e. smooth and compactly supported functions, by $Af(x) = \frac{1}{2}f''(x)$. H and A^g are given by

$$
\begin{aligned}
Hf(x) &= \frac{1}{2}f''(x) + \frac{1}{2}(f'(x))^2, \\
A^g f(x) &= \frac{1}{2}f''(x) + f'(x)g'(x).
\end{aligned}
$$

In both examples, it is seen that A^g is also a generator of a Markov process. More importantly, however, L and H are operator duals.

Lemma 6.3.8. *Under Condition 6.1.2, we have for $f \in D$ that*

$$
\langle Hf, \mu \rangle = \sup_{g \in D} \left\{ \langle A^g f, \mu \rangle - \langle Lg, \mu \rangle \right\}, \tag{6.3.1}
$$

and equality holds for $g = f$. Furthermore, for $g \in D$ and $\mu \in \mathcal{P}(E)$ it holds that

$$
\langle Lg, \mu \rangle = \sup_{f \in D} \left\{ \langle A^g f, \mu \rangle - \langle Hf, \mu \rangle \right\}, \tag{6.3.2}
$$

with equality for $f = g$.

Proof. For $\lambda > 0$, let $A_\lambda := \lambda^{-1}(J(\lambda) - \mathbb{1})$ be the Yosida approximant of A. It is well known that A_λ is bounded and is given by

$$
A_\lambda f(x) = \lambda^{-1} \int q_\lambda(x, \mathrm{d}y) \left[f(y) - f(x)\right],
$$

where $q_\lambda(x, \cdot)$ is the law of the process generated by A after an exponential random time with mean λ.

Now define H_λ, A_λ^g and L_λ in terms of A_λ. Because A_λ is bounded, it follows by Lemma 5.7 in Feng and Kurtz [2006] that

$$
\begin{aligned}
H_\lambda f(x) &\geq A_\lambda^g f(x) - L_\lambda g(x), \\
H_\lambda f(x) &= A_\lambda^f f(x) - L_\lambda f(x).
\end{aligned}
$$

Therefore, it follows by Yosida approximation [Ethier and Kurtz, 1986, Lemma 1.2.4] that

$$Hf(x) = \sup_{g \in D} \{A^g f(x) - Lg(x)\}.$$

The first statement now follows by integration. The variational statement for L is obtained similarly. □

6.3.3 *The Lagrangian and a variational expression for the Hamiltonian*

The Lagrangian in the previous section is still an operator acting on functions. Here we embed this object in a new Lagrangian $\mathcal{L}$ that is a function of place and speed. Also, we introduce a map ρ that transforms 'momentum' into speed.

Definition 6.3.9. Let (D, τ_D) satisfy Condition 6.1.3. Define the Lagrangian $\mathcal{L} : \mathcal{P}(E) \times D' \to [0, \infty]$ by

$$\mathcal{L}(\mu, u) = \sup_{f \in D} \{\langle f, u \rangle - \langle Hf, \mu \rangle\}.$$

Also, define the map $\rho : \mathcal{P}(E) \times D \to D'$ by $\rho(\mu, g) = (A^g)'(\mu)$.

$\mathcal{L}$ can be considered as an extension of L. Pick $\mu \in \mathcal{P}(E)$ and $g \in D$, then

$$\begin{aligned} \mathcal{L}(\mu, \rho(\mu, g)) &= \sup_{f \in D} \{\langle f, \rho(\mu, g) \rangle - \langle Hf, \mu \rangle\} \\ &= \sup_{f \in D} \{\langle A^g f, \mu \rangle - \langle Hf, \mu \rangle\} \\ &= \langle Lg, \mu \rangle, \end{aligned} \tag{6.3.3}$$

where the last equality follows by (6.3.2).

Lemma 6.3.10. $(\mu, u) \mapsto \mathcal{L}(\mu, u)$ *is convex and lower semi-continuous with respect to the weak and weak* topologies.*

Proof. $\mathcal{L}$ is lower semi-continuous, because it is the supremum over continuous functions. Convexity of $\mathcal{L}$ follows by the linearity of $u \mapsto \langle f, u \rangle$ and $\mu \mapsto \langle Hf, \mu \rangle$. □

It turns out that the space D' is to large for practical purposes. In particular, it is not immediately clear that D' with the weak topology is separable. In the proof of Proposition 6.4.2, we need to integrate over D' and because we want to employ an extended version of the Prohorov theorem that needs separability, we will construct a more regular subspace of D' that contains all relevant 'speeds'.

Recall the set $\mathcal{N}$ introduced in Condition 6.1.3 (f) and the definition of a polar in (6.1.2). Define $U \subseteq D'$ by

$$U := \bigcup_{n \in \mathbb{N}} n\mathcal{N}^\circ. \tag{6.3.4}$$

We equip U with the weak* topology inherited from D'. The importance of U follows from the following lemma, which shows that we can restrict the set of allowed 'speeds' to U.

Lemma 6.3.11. *Let $\mu \in \mathcal{P}(E)$. If $u \notin U$, then $\mathcal{L}(\mu, u) = \infty$. Furthermore, for $\mu \in \mathcal{P}(E)$ and $g \in \mathcal{N}$, we have $\rho(\mu, g) \in U$.*

Proof. For $u \notin U = \bigcup_n n\mathcal{N}^\circ$, we can find functions $f_n \in \mathcal{N}$, such that $|\langle f_n, u\rangle| \geq n$. The inequality $|\langle f_n, u\rangle| \leq \mathcal{L}(\mu, u) + \langle H f_n, \mu\rangle \vee \langle H(-f_n), \mu\rangle$, yields that $\mathcal{L}(\mu, u) \geq n - 1$ for every n, which implies that $\mathcal{L}(\mu, u) = \infty$.

The second statement follows from the first, equation (6.3.3), and the fact that Lg is bounded. □

As can be seen from (6.3.3), $\mathcal{L}$ is an extension of L. As expected, H can also be obtained by a Fenchel-Legendre transform of $\mathcal{L}$.

Lemma 6.3.12. *The variational expression for H in* (6.3.1) *extends to*

$$\begin{aligned}\langle Hf, \mu\rangle &= \sup_{u \in D'} \left\{\langle f, u\rangle - \mathcal{L}(\mu, u)\right\} \\ &= \sup_{u \in U} \left\{\langle f, u\rangle - \mathcal{L}(\mu, u)\right\}.\end{aligned}$$

Proof. First of all, note the equality of the two variational expressions on the right had side, as $\mathcal{L}(\mu, u) = \infty$ if $u \notin U$. We give two proofs of the first equality.

First of all, Hölders inequality tells us that $f \mapsto \langle V(t)f, \mu\rangle$ is convex. Therefore, Hf is the norm, and thus, point-wise limit of convex functions which implies that $f \mapsto \langle Hf, \mu\rangle$ is convex. The result follows directly from the fact that the double Fenchel-Legendre transform of the convex lower semi-continuous function $f \mapsto \langle Hf, \mu\rangle$ is $f \mapsto \langle Hf, \mu\rangle$ by the Fenchel-Moreau theorem[Dembo and Zeitouni, 1998, Lemma 4.5.8].

The second approach is more direct. By Definition 6.3.9 of $\mathcal{L}$, we obtain that for every $f \in D$, $\mu \in \mathcal{P}(E)$, $u \in D'$: $\langle Hf, \mu\rangle \geq \langle f, u\rangle - \mathcal{L}(\mu, u)$.

We now show that we in fact have equality. By (6.3.3), we know that $\mathcal{L}(\mu, \rho(\mu, g)) = \langle Lg, \mu\rangle$. Hence, by the second item in Lemma 6.3.8, we obtain

$$\begin{aligned} \langle Hf, \mu\rangle &= \langle A^f f, \mu\rangle - \langle Lf, \mu\rangle \\ &= \langle f, \rho(\mu, f)\rangle - \mathcal{L}(\mu, \rho(\mu, f)), \end{aligned} \tag{6.3.5}$$

which concludes the proof. □

The latter approach in the proof of Lemma 6.3.12, gives us even more information.

Proposition 6.3.13. *Let $\mu \in \mathcal{P}(E)$ and define Γ_μ to be the weak* closed convex hull of $\{\rho(\mu, g) \in U \mid g \in D\}$. If $u \notin \Gamma_\mu$, then $\mathcal{L}(\mu, u) = \infty$.*

Proof. Fix $\mu \in \mathcal{P}(E)$. Define $\hat{\mathcal{L}} = \mathcal{L}$ if $u \in \Gamma_\mu$ and set $\hat{\mathcal{L}} = \infty$ for $u \notin \Gamma_\mu$. It is clear that $\hat{\mathcal{L}}$ is also convex and lower semi-continuous. Because $\hat{\mathcal{L}} \geq \mathcal{L}$, it follows by Lemma 6.3.12 that $\langle Hf, \mu\rangle \geq \sup_u \left\{\langle f, u\rangle - \hat{\mathcal{L}}(\mu, u)\right\}$.

As in (6.3.5), we obtain

$$\begin{aligned} \langle Hf, \mu\rangle &= \langle A^f f, \mu\rangle - \langle Lf, \mu\rangle \\ &= \langle f, \rho(\mu, f)\rangle - \mathcal{L}(\mu, \rho(\mu, f)) \\ &= \langle f, \rho(\mu, f)\rangle - \hat{\mathcal{L}}(\mu, \rho(\mu, f)), \end{aligned}$$

which shows that $\langle Hf, \mu\rangle = \sup_u \left\{\langle f, u\rangle - \hat{\mathcal{L}}(\mu, u)\right\}$. In other words, the double Fenchel-Legendre transform of the convex and lower semi-continuous function $\hat{\mathcal{L}}$ is $\mathcal{L}$. This implies that they are equal. □

6.3.4 *The Doob-h transform in terms of tilted generators*

We connect the operators introduced in the last few sections to the discussion on the Doob-transform in Section 6.2.3. There, we considered a measure $\mathbb{Q}^h \in \mathcal{P}(D_E(\mathbb{R}^+))$, defined by

$$\frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}}(X) = \frac{e^{h(X(t))}}{\langle e^h, \mathbb{P}_t\rangle} = e^{h(X(t)) - \langle V(t)h, \mathbb{P}_0\rangle},$$

and in Lemma 6.2.8, we observed that if we define $h(s) = V(t-s)h$ for $s \leq t$, then the transition operators of $\mathbb{Q}^f$ for times $r \leq s \leq t$ are given by

$$S^{h[0,t]}(r,s)f(x) = e^{-h(r)}(x)S(r,s)\left(fe^{h(s)}\right)(x).$$

It is straightforward to check that $(r,s) \mapsto S^{h[0,t]}(r,s)f$ is continuous for all $f \in C(E)$ and $(r,s) \in \{(r',s') \mid 0 \leq r' \leq s' \leq t\}$. We can say more even. The next lemma shows that the tilted generators of the previous section turn up in the study of this semigroup. After that, we will show that $H(\mathbb{Q}^h \mid \mathbb{P})$ can be given in terms of an integral over the Lagrangian $\mathcal{L}$. We start with two definitions.

Let $C([0,t],D)$ be be space of trajectories $\{g(s)\}_{s\in[0,t]}$, $g(s) \in D$ such that $s \mapsto g(s)$ is continuous with respect to τ_D. Furthermore, let $C^1([0,t],D) \subseteq C([0,t],D)$ be those trajectories for which there exists a trajectory $\{\partial g(s)\}_{s\in[0,t]}$ in $C([0,t],C(E))$ such that for all $s \in [0,t]$, we have

$$\lim_{r\to 0}\left\| \frac{g(s+r)-g(s)}{r} - \partial g(s)\right\| = 0.$$

Now suppose that $h \in D$, then Condition 6.1.2 (b) and (c) imply that $h(s) = V(t-s)h \in D$ for all $s \in [0,t]$. In this case, we can find the trajectory of generators of the semigroup $S^{h[0,t]}$.

Proposition 6.3.14. *Fix $t > 0$ and suppose that $h \in D$. For every $s \in [0,t]$ and $f \in D$, we have*

$$\lim_{r\to 0}\frac{S^{h[0,t]}(s,s+r)f - f}{r} = A^{h(s)}f.$$

If $f \in C^1([0,t],D)$, then we have for every $s \in [0,t]$ that

$$\lim_{r\to 0}\frac{S^{h[0,t]}(s,s+r)f(s+r) - f(s)}{r} = A^{h(s)}f(s) + \partial f(s).$$

Proof. We start with the proof of the first statement. Let $f \in D$ and $s \in [0,t]$, we prove the result for $r > 0$, the proof of the other side is similar. Clearly,

$$\lim_{r\to 0}\left\|\frac{S^{h[0,t]}(s,s+r)f - f}{r} - A^{h(s)}f\right\| = 0$$

if and only if

$$\lim_{r\to 0}\left\| e^{h(s)}\left[\frac{S^{h[0,t]}(s,s+r)f-f}{r} - A^{h(s)}f\right]\right\| = 0.$$

Therefore, we will prove the latter. We see

$$\begin{aligned}
&e^{h(s)}\left[\frac{S^{h[0,t]}(s,s+r)f-f}{r} - A^{h(s)}f\right] \\
&\quad = e^{h(s)}\left[\frac{e^{-h(s)}S(r)\left(fe^{h(s+r)}\right)-f}{r} - A^{h(s)}f\right] \\
&\quad = \frac{S(r)\left(fe^{h(s+r)}\right)-e^{h(s)}f}{r} - A\left(fe^{h(s)}\right) + fAe^{h(s)} \\
&\quad = \frac{S(r)\left(fe^{h(s+r)}\right)-S(r)\left(fe^{h(s)}\right)}{r} + S(r)\left(fAe^{h(s)}\right) \\
&\qquad + \frac{S(r)\left(fe^{h(s)}\right)-fe^{h(s)}}{r} - A\left(fe^{h(s)}\right) \\
&\qquad + fAe^{h(s)} - S(r)\left(fAe^{h(s)}\right).
\end{aligned}$$

The last two lines converge to 0 as $r \downarrow 0$. We consider the term in line four. First note that $S(r)$ is a contraction, thus it suffices to look at

$$f\left[\frac{e^{h(s+r)}-e^{h(s)}}{r} + Ae^{h(s)}\right],$$

but by the definition of $h(s)$ and $h(s+r)$, this equals

$$-f\left[\frac{S(t-s-r)e^{h}-S(t-s)e^{h}}{-r} - AS(t-s)e^{h}\right]$$

which converges to 0 in norm as $r \downarrow 0$.

For the proof of the second statement, let $\{f(s')\}_{s'\leq t} \in C^1([0,t], D)$, then we have for $s \in [0,t]$ that

$$\begin{aligned}
&\frac{S^{h[0,t]}(s,s+r)f(s+r) - f(s)}{r} - \left(A^{h(s)}f(x) + \partial f(s)\right) \\
&= \frac{S^{h[0,t]}(s,s+r)f(s) - f(s)}{r} - A^{h(s)}f(s) \\
&\quad + \frac{S^{h[0,t]}(s,s+r)f(s+r) - S^{h[0,t]}(s,s+r)f(s)}{r} \\
&\qquad\qquad - S^{h[0,t]}(s,s+r)\partial f(s) \\
&\quad + S^{h[0,t]}(s,s+r)\partial f(s) - \partial f(s).
\end{aligned}$$

The first term converges to 0 as shown in the first part of the proof. The second term converges to 0 as $S^{h[0,t]}(s, s+r)$ is contractive for all $r > 0$ and the definition of $\partial f(s)$. The last term converges to 0 as $\{S^{h[0,t]}(r', s')\}_{0\leq r'\leq s'\leq t}$ is strongly continuous. □

The next corollary follows directly from the second statement of proposition.

Corollary 6.3.15. *Let $f \in C^1([0,t], D)$ and $s \in [0,t]$, then*

$$\begin{aligned}
M^f(s) := f(s)(X(s)) - f(0)(X(0)) \\
- \int_0^s A^{h(r)}f(r)(X(r)) + \partial f(r)(X(r))\mathrm{d}r
\end{aligned}$$

is a mean 0 martingale for $\mathbb{Q}^h$.

Proposition 6.3.16. *Fix $t > 0$ and suppose $h \in D$. Define $\mathbb{Q}^h \in \mathcal{P}(D_E(\mathbb{R}^+))$ by the change of measure*

$$\frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}}(X) = \frac{e^{h(X(t))}}{\langle e^h, \mathbb{P}_t\rangle} = e^{h(X(t)) - \langle V(t)h, \mathbb{P}_0\rangle},$$

Then, we have

$$S(\mathbb{Q}^h \,|\, \mathbb{P}) = \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s,$$

where for every $s \in \mathbb{R}^+$, $\gamma(s)$ is the law of $X(s)$ under $\mathbb{Q}^h$ and where $\dot{\gamma}(s) = (A^{h(s)})'(\gamma(s))$.

Proof. Because $S(\mathbb{Q}^h \mid \mathbb{P}) = \int \log \frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{Q}^h$, we study $h(t)(X(t)) - h(0)(X(0))$. Recall that $\partial h(s) = -Hh(s)$ by Corollary 6.3.2 and Lemma 6.3.1:

$$\begin{aligned} &h(t)(X(t)) - h(0)(X(0)) \\ &\quad = h(t)(X(t)) - h(0)(X(0)) \\ &\qquad - \int_0^t A^{h(s)}h(s)(X(s)) + \frac{\partial h(s)}{\partial s}(X(s))\mathrm{d}s \\ &\qquad + \int_0^t A^{h(s)}h(s)(X(s)) + \frac{\partial h(s)}{\partial s}(X(s))\mathrm{d}s \\ &\quad = M^h(t) + \int_0^t A^{h(s)}h(s)(X(s)) - Hh(s)(X(s))\mathrm{d}s \end{aligned}$$

where $s \mapsto M^h(s)$ is a mean 0 $\mathbb{Q}^h$ martingale by Corollary 6.3.15. Therefore, using Lemma 6.3.8 in line 3, we see that

$$\begin{aligned} S(\mathbb{Q}^h \mid \mathbb{P}) &= \int \log \frac{\mathrm{d}\mathbb{Q}^h}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{Q}^h \\ &= \int \int_0^t A^{h(s)}h(s)(X(s)) - Hh(s)(X(s))\mathrm{d}s\mathrm{d}\mathbb{Q}^h(X) \\ &= \int \int_0^t Lh(s)(X(s))\mathrm{d}s\mathrm{d}\mathbb{Q}^h(X). \end{aligned}$$

By Lemma 6.1.5, the operator $L : (D, \tau_D) \to (C(E), \|\cdot\|)$, given by $Lg = A^g g - Hg$ is continuous. Because $s \mapsto h(s) = V(t-s)h$ is continuous in (D, τ_D) by Condition 6.1.3 (d), we see that $s \mapsto Lh(s)$ is norm continuous. Therefore, Fubini's theorem gives us

$$\begin{aligned} S(\mathbb{Q}^h \mid \mathbb{P}) &= \int_0^t \int Lh(s)(X(s))\mathrm{d}\mathbb{Q}^h(X)\mathrm{d}s \\ &= \int_0^t \langle Lh(s), \gamma(s)\rangle \mathrm{d}s \\ &= \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s. \end{aligned}$$

□

6.4 PROOF OF THE MAIN THEOREM

We proceed with the proof of Theorem 6.1.8. We start with two crucial compactness results which are necessary for the Nisio semigroup, introduced in Section 6.4.2, to be well behaved.

6.4.1 *Compactness of the space of paths with bounded Lagrangian cost*

We start with proving the compactness of the level sets of $\mathcal{L}$.

Proposition 6.4.1. *For each $C \geq 0$, the set*

$$\{(\mu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\mu, u) \leq C\}$$

is compact with respect to the weak topology on $\mathcal{P}(E)$ and the weak topology on U.*

Proof. First of all, as $\mathcal{L}$ is lower semi-continuous $\{(\nu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\nu, u) \leq C\}$ is closed. We show that it is contained in a compact set.

Pick the neighbourhood of 0 $\mathcal{N}$ that was given in Condition 6.1.3 (f), so that $\sup_{f \in \mathcal{N}} \|Hf\| \leq 1$. Because $\langle f, u\rangle \leq \mathcal{L}(\mu, u) + \langle Hf, \mu\rangle$, we obtain

$$|\langle f, u\rangle| \leq \mathcal{L}(\nu, u) + \langle Hf, \nu\rangle \vee \langle H(-f), \nu\rangle.$$

As a consequence,

$$\{(\nu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\nu, u) \leq C\} \subseteq \mathcal{P}(E) \times |C+1|\mathcal{N}^\circ$$

Because (D', wk^*) is Hausdorff and a locally convex space, the closure of this set is compact in (D', wk^*) by the Bourbaki-Aloaglu theorem[Treves, 1967, Propositions 32.7 and 32.8], [Robertson and Robertson, 1973, Theorem III.6]. □

We now state an essential ingredient of the proof of Theorem 6.1.8.

Proposition 6.4.2. *For each $M > 0$, and time $T \geq 0$,*

$$\mathcal{K}_M^T := \left\{ \mu \in C_{\mathcal{P}(E)}([0,T]) \,\middle|\, \mu \in \mathcal{AC}, \int_0^T \mathcal{L}(\mu(s), \dot{\mu}(s))\mathrm{d}s \leq M \right\}$$

is a compact subset of $C_{\mathcal{P}(E)}([0,T])$.

We postpone the lengthy proof of the proposition to Sections 6.4.4 and 6.4.5 and focus on proving Theorem 6.1.8 first.

6.4.2 *The Nisio semigroup*

Definition 6.4.3. The Nisio semigroup $\mathbf{V}$ mapping upper semi-continuous functions on $\mathcal{P}(E)$ to upper semi-continuous functions on $\mathcal{P}(E)$ is defined by

$$\mathbf{V}(t)G(\mu) = \sup_{\nu \in \mathcal{AC}_\mu} \left\{ G(\nu(t)) - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) \mathrm{d}s \right\}.$$

For a function $f \in C(E)$, we denote with $[f]$ the weakly continuous function on $\mathcal{P}(E)$ defined by $[f](\mu) = \langle f, \mu \rangle$. Our goal in this section is to show that $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle$.

Note that as a direct consequence of Proposition 6.4.2, if G is a bounded continuous function, than the supremum is actually attained by a curve starting at μ in $\mathcal{K}^t_{3\|G\|}$. For example, this is the case if $G = [g]$, for $g \in C(E)$.

We need one small result, that is essential for the analysis. In particular, it is used for the proof of Lemma 6.4.8.

Lemma 6.4.4. *For each $\mu \in \mathcal{P}(E)$ and $f \in D$, there exists $\nu \in \mathcal{AC}_\mu$ such that for every $t \geq 0$*

$$\int_0^t \langle f, \dot{\nu}(s) \rangle \mathrm{d}s = \int_0^t \langle Hf, \nu(s) \rangle + \mathcal{L}(\nu(s), \dot{\nu}(s)) \mathrm{d}s.$$

In particular by taking $f = 0$, we find that there is a path with zero cost. This in turn yields $\mathbf{V}(t)\mathbf{0} = \mathbf{0}$, where $\mathbf{0}$ is the function defined by $\mathbf{0}(\mu) = 0$ for all $\mu \in \mathcal{P}(E)$.

Proof. Let $\nu(s)$ be the path obtained by the time projections of the Markov process started at μ generated by the operator A^f, see for example Theorem 4.2 in Palmowski and Rolski [2002]. This gives us a path such that $\dot{\nu}(s) = (A^f)'(\nu(s)) = \rho(\nu(s), f)$.

By (6.3.5) on page 166, it follows that

$$\langle Hf, \nu(s) \rangle = \langle f, \rho(\nu(s), f) \rangle - \mathcal{L}(\nu(s), \rho(\nu(s), f))$$

for every s, implying that

$$\int_0^t \langle Hf, \nu(s) \rangle \mathrm{d}s = \int_0^t \left(\langle f, \dot{\nu}(s) \rangle - \mathcal{L}(\nu(s), \dot{\nu}(s)) \right) \mathrm{d}s.$$

□

The semigroup $\{\mathbf{V}(t)\}_{t\geq 0}$ enjoys good continuity properties.

Lemma 6.4.5. *For every* $t \geq 0$, $\mathbf{V}(t)$ *is contractive, i.e. for bounded and upper semi-continuous functions* F, G, *we have*

$$\|\mathbf{V}(t)F - \mathbf{V}(t)G\| \leq \|F - G\|.$$

The proof of this lemma is straightforward. The next result can be proven using Proposition 6.4.2 as Lemma 8.16 in Feng and Kurtz [2006].

Lemma 6.4.6. *For every* $f \in C(E)$ *and* $\mu \in \mathcal{P}(E)$, *we have that* $t \mapsto \mathbf{V}(t)[f](\mu)$ *is continuous.*

We proceed with the preparations of Proposition 6.4.10 where we will prove that $\langle V(t)f, \mu\rangle = \mathbf{V}(t)[f](\mu)$ for $f \in C(E)$ and $\mu \in \mathcal{P}(E)$.

6.2.7, then Proposition 6.3.16

The inequality $\langle V(t)f, \mu\rangle \leq \mathbf{V}(t)[f](\mu)$ is based on the Doob-h transform method and in particular on Lemma 6.2.7 and Proposition 6.3.16. The other inequality is based on approximation arguments. In the next definition, we introduce the resolvent $\mathbf{R}(\lambda)$ of the Nisio semigroup. Based on Lemma 6.3.3, we show that $\mathbf{R}(\lambda)[f](\mu) \leq [R(\lambda)f](\mu)$ which by approximation yields $\mathbf{V}(t)[f](\mu) \leq \langle V(t)f, \mu\rangle$.

Definition 6.4.7. Let G be upper semi-continuous and bounded and let $\lambda > 0$. Define the resolvent $\mathbf{R}(\lambda)$ by

$$\mathbf{R}(\lambda)G(\mu) = \sup_{\nu \in \mathcal{AC}_\mu} \int_0^\infty \frac{1}{\lambda} e^{-\lambda^{-1}s} \left[G(\nu(s)) - \int_0^s \mathcal{L}(\nu(r), \dot{\nu}(r))\mathrm{d}r\right] \mathrm{d}s.$$

Lemma 6.4.8. *For* $g \in D$, *we have* $\mathbf{R}(\lambda)[(\mathbb{1} - \lambda H)g] = [g]$. *As a consequence, we have for* $f \in C(E)$ *and* $\mu \in \mathcal{P}(E)$ *that*

$$\mathbf{R}(\lambda)[f](\mu) \leq [R(\lambda)f](\mu). \tag{6.4.1}$$

Proof. The first statement follows along the lines of the proof of Lemma 8.19 in Feng and Kurtz [2006]. Summarising, the inequality $\mathbf{R}(\lambda)[(\mathbb{1} - \lambda H)g] \leq [g]$ follows by integration by parts and Young's inequality:

$$\langle g, u\rangle \leq \langle Hg, \mu\rangle + \mathcal{L}(\mu, u), \quad \mu \in \mathcal{P}(E),\, u \in D,\, g \in C(E).$$

The second inequality, $\mathbf{R}(\lambda)[(\mathbb{1} - \lambda H)g] \geq [g]$, follows by integration by parts and Lemma 6.4.4, which gives us a trajectory for which equality is attained for all times in Young's inequality.

For the second statement, first note that if $F \geq G$, then $\mathbf{R}(\lambda)F \geq \mathbf{R}(\lambda)G$. Therefore, we obtain by Lemma 6.3.3 that

$$\mathbf{R}(\lambda)[f](\mu) \leq \mathbf{R}(\lambda)[(\mathbb{1} - \lambda H)R(\lambda)f](\mu) = \langle R(\lambda)f, \mu \rangle.$$

□

The next lemma relies on Lemma 6.4.6 and follows exactly as Lemma 8.18 in Feng and Kurtz [2006].

Lemma 6.4.9. *For $t \geq 0$, $f \in D$ and $\mu \in \mathcal{P}(E)$, we have*

$$\lim_{n \to \infty} \mathbf{R}(n)^{\lfloor nt \rfloor}[f](\mu) = \mathbf{V}(t)[f](\mu).$$

We are now able to prove the important result that identifies the Nisio semigroup with $\{V(t)\}_{t \geq 0}$.

Proposition 6.4.10. *For $t \geq 0$, $f \in C(E)$ and $\mu \in \mathcal{P}(E)$, we have*

$$\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle.$$

Proof. By repeatedly using (6.4.1), we obtain

$$\mathbf{R}(n^{-1})^{\lfloor nt \rfloor}[f](\mu) \leq \langle R(n^{-1})^{\lfloor nt \rfloor} f, \mu \rangle,$$

which implies by Lemmas 6.3.4 and 6.4.9 that

$$\mathbf{V}(t)[f](\mu) \leq \langle V(t)f, \mu \rangle. \tag{6.4.2}$$

For the second inequality, we first pick $f \in D$. Let $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+)$ be the Markov measure started from $\mu \in \mathcal{P}(E)$ with transition semigroup $\{S(t)\}_{t \geq 0}$. By Lemma 6.2.7, we have

$$\langle V(t)f, \mu \rangle = \sup_{\mathbb{Q}} \langle f, \mathbb{Q} \rangle - S^{\mathbb{P}_0}(\mathbb{Q}) = \langle f, \mathbb{Q}_t^f \rangle - S(\mathbb{Q}^f \mid \mathbb{P}).$$

If we denote by $\gamma(s)$ the law of $\mathbb{Q}^f$ at time s, then Proposition 6.3.16 yields

$$\langle V(t)f, \mu \rangle = \langle f, \gamma(t) \rangle - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s \leq \mathbf{V}(t)[f](\mu).$$

This inequality, together with (6.4.2), yields $\langle V(t)f, \mu \rangle = \mathbf{V}(t)[f](\mu)$ for $f \in D$. The result for $f \in C(E)$ follows by the continuity of $f \mapsto V(t)f$ and the continuity of $f \mapsto \mathbf{V}(t)f$ given by Lemma 6.4.5. □

6.4.3 *The Lagrangian form of the rate function*

In this section, we show that I_t can be re-expressed using the Nisio semi-group.

Lemma 6.4.11. *Under the Condition 6.1.3, it holds that*

$$I_t(\mu_1 \,|\, \mu_0) = \inf_{\substack{\nu \in \mathcal{AC}_{\mu_0} \\ \nu(t)=\mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) \mathrm{d}s.$$

The proof is a classical proof using convex duality.

Proof. For a fixed measure $\mu_0 \in \mathcal{P}(E)$, consider the function $\mathbb{L}_{\mu_0} : \mathcal{P}(E) \to \infty$ defined by

$$\mathbb{L}_{\mu_0}(\mu_1) := \inf_{\substack{\nu \in \mathcal{AC}_{\mu_0} \\ \nu(t)=\mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) \mathrm{d}s$$

Our goal is to prove that $I_t(\mu_1 \,|\, \mu_0) = \mathbb{L}_{\mu_0}(\mu_1)$ by showing that both are the Fenchel-Legendre transform of $\langle V(t)g, \mu_1 \rangle$. First, we will prove that $\mathbb{L}_{\mu_0}$ is convex and has compact level sets. This last result implies the lower semi-continuity.

Step 1. The convexity of $\mathbb{L}_{\mu_0}$ follows directly from the convexity of $\mathcal{L}$ and the fact that $\mathcal{AC}$ is convex. So we are left to prove compactness of the level sets. Pick a sequence μ^n in the set $\{\mu \,|\, \mathbb{L}_{\mu_0}(\mu) \leq c\}$. We know by definition of $\mathbb{L}_{\mu_0}$ and Proposition 6.4.2 that there are $\nu^n \in \mathcal{K}^t_{c,\{\mu_0\}}$ such that $\nu^n(0) = \mu_0$, $\nu^n(t) = \mu^n$ and

$$\int_0^t \mathcal{L}(\nu^n(s), \dot{\nu}^n(s)) \mathrm{d}s \leq c.$$

Again by Proposition 6.4.2, we obtain that the sequence ν^n has a converging subsequence ν^{n_k} with limit ν^* such that

$$\int_0^t \mathcal{L}(\nu^*(s), \dot{\nu}^*(s)) \mathrm{d}s \leq c.$$

Denote with $\mu^* := \nu^*(t)$, then we know that $\nu^{n_k}(t) \to \mu^*$ and $\mathbb{L}_{\mu_0}(\mu^*) \leq c$, which implies that $\mathbb{L}_{\mu_0}(\cdot)$ has compact level sets and is lower semi-continuous.

Step 2. Now that we know that $\mathbb{L}_{\mu_0}$ is convex and lower semi-continuous, we are able to prove that $\mathbb{L}_{\mu_0}(\cdot) = I_t(\cdot \mid \mu_0)$.

$\mathbb{L}_{\mu_0}(\cdot)$ is lower semi-continuous on $\mathcal{P}(E)$ with respect to the weak topology, so extending its domain of definition to $\mathcal{M}(E)$ by setting it equal to ∞ outside $\mathcal{P}(E)$ does not change the fact that it is lower semi-continuous. Because the dual of $(\mathcal{M}(E), \text{weak})$ is $C(E)$ by the Riesz representation theorem and [Conway, 2007, Theorem V.1.3], we obtain by Lemma 4.5.8 in Dembo and Zeitouni [1998] that the Fenchel-Legendre transform of

$$\sup_{\mu_1} \{\langle g, \mu_1\rangle - \mathbb{L}_{\mu_0}(\mu_1)\} = \sup_{\nu\in\mathcal{AC}_{\mu_0}} \left\{\langle g, \nu(t)\rangle - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s\right\} = \mathbf{V}(t)[g](\mu_0)$$

satisfies $\mathbb{L}_{\mu_0}(\mu_1) = \sup_{g\in C(E)} \{\langle g, \mu_1\rangle - \mathbf{V}(t)[g](\mu_0)\}$. Therefore, by Proposition 6.4.10, we see

$$\mathbb{L}_{\mu_0}(\mu_1) = \sup_{g\in C_0(E)} \{\langle g, \mu_1\rangle - \langle V(t)g, \mu_0\rangle\}. \tag{6.4.3}$$

On the other hand, by Theorem 6.1.1,

$$I_t(\mu_1 \mid \mu_0) = \sup_{g\in C_0(E)} \{\langle g, \mu_1\rangle - \langle V(t)g, \mu_0\rangle\}. \tag{6.4.4}$$

The combination of equations (6.4.3) and (6.4.4), i.e. both are the Legendre-Fenchel transform of $\langle V(t)g, \mu_0\rangle$, yields that

$$I_t(\mu_1 \mid \mu_0) = \mathbb{L}_{\mu_0}(\mu_1) = \inf_{\substack{\nu\in\mathcal{AC}_{\mu_0}\\ \nu(t)=\mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s.$$

□

We proceed with the final lemma before the proof of Theorem 6.1.8.

Lemma 6.4.12. *The function* $J : C_{\mathcal{P}(E)}(\mathbb{R}^+) \to [0, \infty]$, *given by*

$$J(\mu) = \begin{cases} S(\mu(0) \mid \mathbb{P}_0) + \int_0^\infty \mathcal{L}(\mu(s), \dot{\mu}(s))\mathrm{d}s & \text{if } \mu \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

has compact level sets in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. Clearly, $\{J \leq M\} \subseteq \bigcap_T \mathcal{K}_M^T$. So, pick a sequence $\mu^n \in \{J \leq M\}$. For $n = 1$, we can construct a converging subsequence μ^{n_k} in $\mathcal{K}_M^1$ seen as a subset of $C_{\mathcal{P}(E)}([0,1])$. From this subsequence, we can extract yet another subsequence that has the same property on $[0,2]$. By a diagonal argument, this yields a converging subsequence in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. By the lower semi-continuity of $H(\cdot\,|\mathbb{P}_0)$ and $\mathcal{L}$ this yields that the limit is in $\{J \leq M\}$. □

Proof of Theorem 6.1.8. By using the contraction principle from the space

$$C_{\mathcal{P}(E)}(\mathbb{R}^+) \to \prod_{\mathbb{R}^+} \mathcal{P}(E)$$

using the identity map, we find that the rate function in Theorem 6.1.1 coincides with the rate function which would have been found via the Dawson-Gärtner theorem [Dembo and Zeitouni, 1998, Theorem 4.6.1] for the large deviation problem on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$.

In this context, we can apply Lemma 4.6.5 Dembo and Zeitouni [1998] to find that if we have a good rate function J on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$ that satisfies

$$\begin{aligned} I\,[0, t_1, \ldots, t_k]\,(\mu(0), \mu(t_1), \ldots, \mu(t_k)) \\ = \inf\left\{J(\nu) \,|\, \nu(0) = \mu(0), \nu(t_i) = \mu(t_i)\right\}, \end{aligned} \tag{6.4.5}$$

then it holds that $I = J$. The candidate

$$J(\mu) = \begin{cases} H(\mu(0)\,|\mathbb{P}_0) + \left\{\int_0^\infty \mathcal{L}(\mu(s), \dot{\mu}(s))\mathrm{d}s\right\} & \text{if } \mu \in \mathcal{AC}_{\mu_0}, \\ \infty & \text{otherwise,} \end{cases}$$

clearly satisfies (6.4.5). By Lemma 6.4.12, we know that J is a good rate function on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$ and therefore also on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$. □

6.4.4 *Preparations for the proof of Proposition 6.4.2*

We say that a topological space is Souslin if it is the continuous image of a complete separable metric space. For the proof of Proposition 6.4.2, we will need the generalisation of one of the implications of the Prohorov theorem.

Theorem 6.4.13 (Prohorov)**.** *Let $\mathcal{K}$ be a subset of the Borel measures on a completely regular Souslin space $\mathcal{S}$ that is uniformly bounded with respect to the total variation norm. If $\mathcal{K}$ is a tight family of measures, then $\mathcal{K}$ has a compact and sequentially compact closure with respect to the weak topology on $\mathcal{P}(\mathcal{S})$.*

The Prohorov theorem is given in [Bogachev, 2007, Theorem 8.6.7] and its specialisation to completely regular Souslin spaces follows from [Bogachev, 2007, Corollary 6.7.8 and Theorem 7.4.3]

Remark 6.4.14. The other implication of the ordinary Prohorov theorem does not necessarily hold in this generality [Bogachev, 2007, Proposition 8.10.19].

We will use the Prohorov theorem for measures on the product space $(\mathcal{P}(E) \times U \times [0,T])$, where the first two spaces are equipped with the weak* topology, and the last space with its standard topology.

Lemma 6.4.15. *The space* $(\mathcal{P}(E) \times U \times [0,T])$ *is completely regular and Souslin.*

Proof. We start with proving that $(\mathcal{P}(E) \times U \times [0,T])$ is completely regular. By Lemma [Köthe, 1969, 15.2.(3)] (D', wk^*) is completely regular, therefore, the subspace (U, wk^*) is completely regular. This yields the result as taking products preserves complete regularity.

By Condition 6.1.3 (a) and Lemma 6.7.6, we obtain that (U, wk^*) is Souslin. Clearly, $(\mathcal{P}(E), wk)$ and $[0,T]$ are Souslin, so that the product space $(\mathcal{P}(E) \times U \times [0,T])$ is Souslin by Lemma 6.6.5 in Bogachev Bogachev [2007]. □

Suppose that we have a weakly converging net of measures on $(\mathcal{P}(E) \times U \times [0,T])$. By definition, integrals of continuous and bounded functions with respect to this net of measures converges in $\mathbb{R}$. The next lemmas are aimed to extend this property to continuous functions, that are unbounded, but linear on U.

Definition 6.4.16. For the neighbourhood $\mathcal{N}$, we define the Minkowski functional on U

$$\|u\|_{\mathcal{N}} := \inf \{c \geq 0 \,|\, u \in c\mathcal{N}^\circ\}.$$

We have the following elementary results.

Lemma 6.4.17. $\|\cdot\|_{\mathcal{N}}$ *is a norm on* U, $\{u \,|\, \|u\|_{\mathcal{N}} \leq 1\} = \mathcal{N}^\circ$. *Furthermore, for* $u \in U$, *we have*

$$\sup_{f \in c\mathcal{N}} \frac{\langle f, u \rangle}{\|u\|_{\mathcal{N}}} = c.$$

We use this lemma to find functions ϕ of the type given in the following lemma, which is an analogue of the de la Vallée-Poussin lemma [Bogachev, 2007, Theorem 4.5.9].

Lemma 6.4.18. *For a net of measures π^α bounded in total variation norm, that weakly converges to a measure π, and a measurable function f, suppose that there exists a non-negative non-decreasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ which satisfies*

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

for which it holds that $\sup_\alpha \int \phi(|f|)\mathrm{d}\pi^\alpha \leq M < \infty$, then it holds that

$$\sup_\alpha \int |f| \mathrm{d}\pi^\alpha < \infty.$$

Also, we obtain that uniformly in α

$$\lim_{C\to\infty} \left| \int f \mathrm{d}\pi^\alpha - \int \Upsilon_C(f) \mathrm{d}\pi^\alpha \right| = 0, \tag{6.4.6}$$

where $\Upsilon_C(f) = (f \vee -C) \wedge C$.

Proof. Fix $\varepsilon > 0$ and pick $C(\varepsilon)$ big enough such that for $r \geq C(\varepsilon)$ we have $\frac{\phi(r)}{r} \geq \frac{M}{\varepsilon}$. Then, we obtain that

$$\sup_\alpha \int_{|f|\geq C} |f| \mathrm{d}\pi^\alpha \leq \sup_\alpha \frac{\varepsilon}{M} \int_{|f|\geq C} \phi(|f|)\mathrm{d}\pi^\alpha \leq \frac{\varepsilon}{M} M \leq \varepsilon.$$

As a consequence, we see

$$\sup_\alpha \int |f| \mathrm{d}\pi^\alpha \leq C(\varepsilon) \sup_\alpha \|\pi^\alpha\|_{TV} + \varepsilon < \infty.$$

The second statement follows by the observation that

$$\sup_\alpha \int |f - \Upsilon_C(f)| \, \mathrm{d}\pi^\alpha \leq \sup_\alpha \int_{|f|\geq C} |f| \mathrm{d}\pi^\alpha.$$

□

Lemma 6.4.19. *Under Condition 6.1.3 (f) that states that for every $c \geq 0$: $\Gamma(c) := \sup_{f\in c\mathcal{N}} \|Hf\| < \infty$, there exists a non-decreasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$, such that $\lim_{r\to\infty} \frac{\phi(r)}{r} = \infty$ and such that $\phi(|\langle f, u\rangle|) \leq \phi(\|u\|_\mathcal{N}) \leq \mathcal{L}(\mu, u)$ for every $f \in \mathcal{N}$, $u \in U$ and $\mu \in \mathcal{P}(E)$.*

The proof of this lemma is inspired by the proof of Lemma 10.21 in Feng and Kurtz [2006].

Proof. For $u \neq 0$ in U, Lemma 6.4.17 yields

$$\frac{\mathcal{L}(\mu,u)}{\|u\|_{\mathcal{N}}} \geq \sup_{f \in c\mathcal{N}} \left\{ \frac{\langle f, u\rangle}{\|u\|_{\mathcal{N}}} - \frac{\langle Hf, \mu\rangle}{\|u\|_{\mathcal{N}}} \right\} \geq c - \frac{\Gamma(c)}{\|u\|_{\mathcal{N}}}$$

for every $c > 0$. This directly yields for every $c > 0$

$$\lim_{r\to\infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u\,:\,\|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu,u)}{\|u\|_{\mathcal{N}}} \geq \lim_{r\to\infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u\,:\,\|u\|_{\mathcal{N}} \geq r} c - \frac{\Gamma(c)}{\|u\|_{\mathcal{N}}} = c,$$

which implies

$$\lim_{r\to\infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u\,:\,\|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu,u)}{\|u\|_{\mathcal{N}}} = \infty.$$

Consequently, the function

$$\phi(r) = r \inf_{\mu \in \mathcal{P}(E)} \inf_{u\,:\,\|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu,u)}{\|u\|_{\mathcal{N}}},$$

satisfies the claims in the lemma. □

6.4.5 *Proof of Proposition 6.4.2*

We now have the tools for the proof of Proposition 6.4.2. Essentially, the proof follows the approach as in [Feng and Kurtz, 2006, Proposition 8.13]. We give it for clarity and completeness as there are some notable differences. First of all, we work with absolutely continuous paths, instead of paths that satisfy a relaxed control equation. Second, the possible 'speeds' that we allow are elements of the completely regular Souslin subset U of a locally convex space instead of a metric space.

Proof of Proposition 6.4.2. Pick a sequence $\mu^n \in \mathcal{K}_M^T$. Because $\mathcal{P}(E)$ is compact, we assume that $\mu^n(0) \to \mu_0$. Define the occupation measures π^n on $\mathcal{P}(E) \times U \times [0,T] \subseteq \mathcal{P}(E) \times U \times [0,T]$ by

$$\pi^n(C \times [0,t]) = \int_0^t \mathbb{1}_C(\mu^n(s), \dot{\mu}^n(s)) \mathrm{d}s.$$

Proposition 6.4.1 tells us that π^n is tight in $\mathcal{P}\left(\mathcal{P}(E) \times U \times [0,T]\right)$ by considering the following calculation:

$$\begin{aligned} &C\pi^n\left\{(\mu,u,t) \in \mathcal{P}(E) \times U \times [0,T] \,|\, \mathcal{L}(\mu,u) \le C\right\}^c \\ &\le \int_0^T \mathcal{L}(\mu,u)\pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\ &\le M. \end{aligned}$$

In other words

$$\pi^n\left\{(\mu,u,t) \in \mathcal{P}(E) \times U \times [0,T] \,|\, \mathcal{L}(\mu,u) \le C\right\}^c \le \frac{M}{C}, \qquad (6.4.7)$$

and because C is arbitrary, we can choose it big enough such that this probability is smaller then any $\varepsilon > 0$ uniformly in n. This implies by Theorem 6.4.13 that π^n contains a weakly converging subsequence. Therefore, we assume without loss of generality that, there exists $\pi \in \mathcal{P}(\hat{K} \times U \times [0,T])$ such that $\pi^n \to \pi$ weakly.

We now show that π gives us a new path $s \mapsto \mu(s)$ in $\mathcal{K}_M^T$. Recall that for $c \ge 0$ $\Upsilon_c(g) = (g \wedge c) \vee -c$. So for a fixed $f \in D$, $u \mapsto \Upsilon_c(\langle f,u\rangle)$ is a bounded and continuous function. For an arbitrary $t \le T$, the set $\pi(\mathcal{P}(E) \times U \times \{t\})$ is a set of measure 0, so the function $(u,s) \mapsto \mathbb{1}_{\{s\le t\}}\Upsilon_c(\langle f,u\rangle)$ is a bounded Borel measurable functions that is continuous π almost everywhere.

Hence, by the weak convergence of π^n to π and Corollary 8.4.2 in Bogachev Bogachev [2007], we obtain for every $c \ge 0$ that

$$\int_{\{s\le t\}} \Upsilon_c(\langle f,u\rangle)\ \pi^n(\mathrm{d}\mu\times\mathrm{d}u\times\mathrm{d}s) \to \int_{\{s\le t\}} \Upsilon_c(\langle f,u\rangle)\ \pi(\mathrm{d}\mu\times\mathrm{d}u\times\mathrm{d}s).$$

By the Portmanteau theorem and the lower semi-continuity of $\mathcal{L}$, we obtain that

$$\int \mathcal{L}(\mu,u)\ \pi(\mathrm{d}\mu\times\mathrm{d}u\times\mathrm{d}s) \le \liminf_n \int \mathcal{L}(\mu,u)\ \pi^n(\mathrm{d}\mu\times\mathrm{d}u\times\mathrm{d}s) \le M.$$

Because $\phi(|\langle f,u\rangle|) \le \mathcal{L}(\mu,u)$ by Lemma 6.4.19, and the fact that ϕ satisfies the conditions of Lemma 6.4.18, we use the result in (6.4.6) to obtain that

$$\sup_n \left| \int_{\{s\le t\}} \langle f,u\rangle - \Upsilon_c(\langle f,u\rangle)\ \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \to 0,$$

as $c \to \infty$. This also follows for the limiting measure π:

$$\left| \int_{\{s \leq t\}} \langle f, u \rangle - \Upsilon_c(\langle f, u \rangle) \, \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \to 0.$$

Thus, by first sending c and then n to infinity, we get

$$\begin{aligned}
&\left| \int_{\{s \leq t\}} \langle f, u \rangle \, \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) - \int_{\{s \leq t\}} \langle f, u \rangle \, \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \\
&\leq \left| \int_{\{s \leq t\}} \langle f, u \rangle - \Upsilon_c(\langle f, u \rangle) \, \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \\
&\quad + \left| \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \; (\pi^n - \pi) \, (\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \\
&\quad + \left| \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) - \langle f, u \rangle \, \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \right| \\
&\quad \to 0.
\end{aligned} \tag{6.4.8}$$

Fix some $0 \leq t \leq T$ and pick a sequence $0 \leq t_n \leq T$ that converges to t. Because $\mu^n(t_n)$ is a sequence in the compact set $\mathcal{P}(E)$ it has a converging subsequence with limit ν. By Lemmas 6.4.18, 6.4.19, and the Dominated convergence theorem, we have

$$\lim_{n \to \infty} \int \mathbb{1}\{s \text{ between } t_n \text{ and } t\} |\langle f, u \rangle| \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \to 0,$$

which implies, using (6.4.8), that

$$\begin{aligned}
&\langle f, \nu \rangle - \langle f, \mu_0 \rangle \\
&= \lim_n \langle f, \mu^n(t_n) \rangle - \langle f, \mu^n(0) \rangle \\
&= \lim_n \int \mathbb{1}\{s \leq t\} \langle f, u \rangle \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&\quad - \int \mathbb{1}\{s \text{ between } t_n \text{ and } t\} \langle f, u \rangle \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&= \int \mathbb{1}\{s \leq t\} \langle f, u \rangle \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s).
\end{aligned}$$

Because D is dense in $C(E)$, this uniquely determines ν, and for every sequence $s_n \to t$, one gets $\mu^n(s_n) \to \nu$ weakly. Therefore, we will denote

$\mu(t) := \nu$. This way, we can construct $\mu(t)$ for a countable dense subset J of $[0, T]$ and $\mu(t)$ is continuous on J. As a consequence, $\mu(t)$ extends continuously to $[0, t]$ and satisfies

$$\langle f, \mu(t) \rangle - \langle f, \mu_0 \rangle = \int \mathbb{1}_{\{s \le t\}} \langle f, u \rangle \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s)$$

for every $f \in D$. This implies that for any sequence $s_n \to t$, we have $\mu(s_n) \to \mu(t)$, which yields that $\{\mu^n(t)\}_{0 \le t \le T}$ converges to $\{\mu(t)\}_{0 \le t \le T}$ in $C_{\mathcal{P}(E)}([0, T])$.

We proceed with extracting the speed of the trajectory $s \mapsto \mu(s)$ from the measure π. Let $\hat{\pi}$ be the measure π restricted to $U \times [0, T]$. By Corollary 10.4.6 in Bogachev Bogachev [2007], we can write $\hat{\pi}(\mathrm{d}u \times \mathrm{d}s)$ as $\lambda_s(\mathrm{d}u)\mathrm{d}s$. For Lebesgue almost every s, we know that $\int |\langle f, u \rangle| \lambda_s(\mathrm{d}u) < \infty$, so we can define the Gelfand integral $\bar{u}(s) = \int u \lambda_s(\mathrm{d}u)$, see Theorem 6.7.4. We show that $\bar{u}(s) = \dot{\mu}(s)$. First, by the measurability of $s \mapsto \lambda_s$, also $s \mapsto \bar{u}$ is measurable. Second, by Jensen's inequality in the first line, and the lower semi-continuity of $\mathcal{L}$ in the third,

$$\begin{aligned}
&\int_0^T |\langle f, \bar{u}(s) \rangle| \mathrm{d}s \\
&\qquad \le \int |\langle f, u \rangle| \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&\qquad \le T(\|Hf\| \vee \|H(-f)\|) + \int \mathcal{L}(\mu, u) \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&\qquad \le T(\|Hf\| \vee \|H(-f)\|) \\
&\qquad\qquad\qquad + \liminf_n \int \mathcal{L}(\mu, u) \pi^n(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&\qquad \le T(\|Hf\| \vee \|H(-f)\|) + M.
\end{aligned}$$

Last,

$$\begin{aligned}
\langle f, \mu(t) \rangle - \langle f, \mu(0) \rangle &= \int \mathbb{1}_{\{s \le t\}} \langle f, u \rangle \pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&= \int \mathbb{1}_{\{s \le t\}} \langle f, u \rangle \hat{\pi}(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) \\
&= \int_0^t \int \langle f, u \rangle \lambda_s(\mathrm{d}u) \mathrm{d}s \\
&= \int_0^t \langle f, \bar{u}(s) \rangle \mathrm{d}s.
\end{aligned}$$

This means that $\mu \in \mathcal{AC}^T$ and $\dot{\mu} = \bar{u}$.

We still need to show that $\mu \in \mathcal{K}_M^T$. By the construction of the path $s \mapsto \mu(s)$, it is clear that we have $\pi(\mathrm{d}\mu \times \mathrm{d}u \times \mathrm{d}s) = \mathbb{1}_{\{s \leq T\}}\delta_{\{\mu(s)\}}(\mathrm{d}\mu)\lambda_s(\mathrm{d}u)\mathrm{d}s$. This shows, using the convexity of $\mathcal{L}$ in the second line, and lower semi-continuity of $\mathcal{L}$ in the third line, that

$$\begin{aligned}\int_0^T \mathcal{L}(\mu(s), \dot{\mu}(s))\mathrm{d}s &= \int \mathcal{L}(\mu, u)\mathbb{1}\{s \leq T\}\delta_{\mu(s)}(\mathrm{d}\mu)\delta_{\bar{u}(s)}(\mathrm{d}u)\mathrm{d}s \\ &\leq \int \mathcal{L}(\mu, u)\mathbb{1}\{s \leq T\}\delta_{\mu(s)}(\mathrm{d}\mu)\lambda_s(\mathrm{d}u)\mathrm{d}s \\ &\leq \liminf_n \int_0^T \mathcal{L}(\mu^n(s), \dot{\mu}^n(s))\mathrm{d}s \\ &\leq M.\end{aligned}$$

So indeed $\mathcal{K}_M^T$ is compact in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. □

6.5 EXAMPLES

We give a number of examples on which Theorem 6.1.8 can be applied. First of all, we begin with a Markov jump process on a compact metric space. After that, interacting particle systems are considered, see Liggett [1985]. In that case, we also prove a representation theorem for D'. Finally, we consider diffusion processes.

6.5.1 *Markov pure jump process*

On a compact metric space (E, d), we have a Markov process $X(t)$ with associated semigroup $S(t) : C(E) \to C(E)$ generated by the bounded generator

$$Af(a) = \int r(a, \mathrm{d}b)\left[f(b) - f(a)\right],$$

where for every a $r(a, \cdot)$ is some non-negative measure, which is weakly continuous in a, satisfying $\|r\|_\infty = \sup_a r(a, E) < \infty$. We work with the space $(D, \tau_D) = (C(E), \|\cdot\|)$. In this case, the generators A^g and operator H are given by

$$A^g f(a) = \int r(a, \mathrm{d}b) e^{g(b)-g(a)} \left[f(b) - f(a)\right],$$
$$Hf(a) = \int r(a, \mathrm{d}b) \left[e^{f(b)-f(a)} - 1\right].$$

Lemma 6.5.1. *Conditions 6.1.2 and 6.1.3 are satisfied.*

Proof. Take $D = C(E)$, which clearly satisfies Conditions 6.1.2 (a) and (b').

Conditions 6.1.3 (a)-(c), (e) are clear. For (d), we only need to prove that $t \mapsto V(t)f$ is continuous for every $f \in C(E)$. So take a sequence $t_n \in \mathbb{R}^+$ converging to $t \in \mathbb{R}^+$. Then $S(t_n)e^f \to S(t)e^f$ by the strong continuity of $\{S(t)\}_{t \geq 0}$. Because f is bounded, the functions $S(t_n)e^f$ satisfy $e^{-\|f\|} \leq S(t_n)e^f(x) \leq e^{\|f\|}$ for all $x \in E$. On $[e^{-\|f\|}, e^{\|f\|}]$ the logarithm is uniformly continuous, which implies that $\|V(t_n)f - V(t)f\| \to 0$.

Finally, (f) is satisfied by taking $\mathcal{N} = \{g \in C(E) \mid \|g\| \leq \frac{1}{2}\log(\|r\|^{-1} + 1)\}$. □

6.5.2 *Interacting particle systems*

Let W be a compact metric space and let S be a countable set. Define $(E = W^S, d)$, the product space with d a metric that is compatible with the topology, on which we will define a Markov process $\{\eta(t)\}_{t \geq 0}$. Examples are the exclusion process, the contact process, etcetera. We follow the construction in Liggett Liggett [1985].

For Λ a finite subset of S and $\zeta \in W^\Lambda$ let $c_\Lambda(\eta, \mathrm{d}\zeta)$ be the rate at which the system makes a transformation from configuration η to η^ζ which is defined by

$$\eta^\zeta_x = \begin{cases} \eta_x & \text{if } x \notin \Lambda, \\ \zeta_x & \text{if } x \in \Lambda. \end{cases}$$

Put $c_\Lambda = \sup\{c_\Lambda(\eta, W^\Lambda) \mid \eta \in E\}$, the maximal total variation of $c_\Lambda(\eta, \cdot)$. We assume that $c_\Lambda(\eta, \mathrm{d}\zeta)$ is weakly continuous in the first variable. We define for finite $\Lambda \subseteq S$ and $u \in S$:

$$c_\Lambda(u) = \sup\left\{\|c_\Lambda(\eta, \mathrm{d}\zeta) - c_\Lambda(\hat{\eta}, \mathrm{d}\zeta)\|_{TV} \mid \eta_y = \hat{\eta}_y \text{ for } y \neq u\right\},$$

where $\|\cdot\|_{TV}$ refers to the total variation norm. This measures the amount that $\eta \mapsto c_\Lambda(\eta, \cdot)$ depends on the coordinate η_u. Furthermore, let $\gamma(x,u) = \sum_{\Lambda \ni x} c_\Lambda(u)$ for $u \neq x$ and $\gamma(x,x) = 0$ for all x.

For $f \in C(E)$, define

$$\Delta_f(x) = \sup\{|f(\eta) - f(\zeta)| \mid \text{ for } y \neq x : \; \eta_y = \zeta_y\}$$

the variation of f at $x \in S$. For a function in $C(E)$ let $\mathcal{D}(f) := \{x \in S \mid \Delta_f(x) > 0\}$ be the dependence set of f and define the space of local functions by

$$\{f \in D \mid |\mathcal{D}(f)| < \infty\}.$$

and the space of test functions by

$$D = \left\{ f \in C_b(E) \,\middle|\, \|\!|f|\!\| := \sum_{x \in S} \Delta_f(x) < \infty \right\}, \tag{6.5.1}$$

which is the closure of the space of local functions with respect to the $\|\!|\cdot|\!\|$ semi-norm.

For functions $f \in D$, define the formal generator A to be

$$Af(\eta) = \sum_\Lambda \int c_\Lambda(\eta, \mathrm{d}\zeta) \left[f(\eta^\zeta) - f(\eta) \right]. \tag{6.5.2}$$

Theorem I.3.9 in Liggett [1985] shows that the closure of A generates a Feller semigroup $\{S(t)\}_{t \geq 0}$. Using this semigroup a Markov process $(\eta(t))_{t \geq 0}$ is constructed such that $S(t)f(\eta) = \mathbb{E}[f(\eta(t)) \mid \eta(0) = \eta]$.

Theorem 6.5.2 (Liggett I.3.9). *Assume that*

$$\sup_x \sum_{\Lambda \ni x} c_\Lambda < \infty, \tag{6.5.3}$$

and

$$M := \sup_{x \in S} \sum_{\Lambda \ni x} \sum_{u \neq x} c_\Lambda(u) = \sup_{x \in S} \sum_u \gamma(x,u) < \infty. \tag{6.5.4}$$

Finally, define the quantity

$$\varepsilon = \inf_{u \in S} \inf_{\substack{\eta_1 = \eta_2 \\ \text{off } u \\ \eta_1(u) \neq \eta_2(u)}} \sum_{\Lambda \ni u} \left[c_\Lambda\left(\eta_1, \{\zeta \mid \zeta(u) = \eta_2(u)\}\right) \right.$$
$$\left. + \, c_\Lambda\left(\eta_2, \{\zeta \mid \zeta(u) = \eta_1(u)\}\right) \right].$$

Then, we have the following:

(a) *The closure of $\overline{A}$ of A generates a strongly continuous positive contraction semigroup $S(t)$.*
(b) *D is a core for $\overline{A}$.*
(c) *If $f \in D$, then $S(t)f \in D$ for all $t \geq 0$ and*

$$|||S(t)f||| \leq e^{t(M-\varepsilon)} |||f||| .$$

To make the notation a bit easier, we do not distinguish between $\overline{A}$ and A. A calculation gives the expressions for $A^g f$ and Hf for $f, g \in D$.

$$\begin{aligned} A^g f(\eta) &= \sum_{\Lambda} \int c_\Lambda(\eta, \mathrm{d}\zeta) e^{g(\eta^\zeta)-g(\eta)} \left[f(\eta^\zeta) - f(\eta) \right], \\ Hf(\eta) &= \sum_{\Lambda} \int c_\Lambda(\eta, \mathrm{d}\zeta) \left[e^{g(\eta^\zeta)-g(\eta)} - 1 \right]. \end{aligned}$$

Remark 6.5.3. It is also possible to consider interacting particle systems where a bounded operator is added to A, without changing the core D. For example, one can consider

$$A_\theta f(\eta) = Af(\eta) + \sum_i c_i(\eta) \left[f(\theta_i \eta) - f(\eta) \right]$$

where θ_i is a shift: $(\theta_i \eta)_j = \eta_{i+j}$, and $\sum_i \|c_i\| < \infty$.
This includes processes like the environment process seen from a random walker in a dynamic random environment and the tagged particle process.

Our first goal is to equip D with a topology τ_D. The semi-norm $|||\cdot|||$ defined on D will be our starting point for τ_D as (6.5.3) implies $\|Af\| \leq \sup_x \sum_{\Lambda \ni x} c_\lambda |||f|||$. Note that $|||\mathbb{1}||| = 0$, so $|||\cdot|||$ alone can not define a topology. We do have the following result.

Lemma 6.5.4. *Let $\mathcal{C}$ be the space of constant functions and let $\|\cdot\|_Q$ be the norm on the quotient space $C(E)/\mathcal{C}$. For $f \in D$, we have that $2\|f\|_Q \leq |||f|||$.*

Proof. It is sufficient to prove the statement for local functions, because every $f \in D$ can be approximated by local f_n for which it holds that $|||f_n||| \to |||f|||$ and $\|f_n\| \to \|f\|$.
Suppose that f is a local function and let $\mathcal{D}(f) = \{x_1, \ldots, x_n\}$. Now pick the function $f' \in D$ such that $f = f' + c$ for some $c \in \mathbb{R}$, such that the range of f' is contained in $[0, 2\|f\|_Q]$. Pick η and ζ such that $f'(\eta) = 2\|f\|_Q$

and $f'(\zeta) = 0$. For $0 \leq k \leq n$ define $\Lambda_k = \{x_1, \dots, x_k\}$ and let ξ_k be equal to ζ on Λ_k and equal to η off Λ_k. Then it holds that

$$2 \|f\|_Q = f'(\eta) = f'(\xi_0) = \sum_{k=0}^{n-1} f(\xi_k) - f(\xi_{k+1}) \leq \sum_{k=1}^{n} \Delta_f(x_k) = \|\!|f|\!\|\,.$$

□

The Lemma shows that one additional semi-norm is sufficient to topologise D. Let τ_D be the topology induced by $\|\cdot\|_D := \|\!|\cdot|\!\| + \|\cdot\|$.

Lemma 6.5.5. (D, τ_D) *is a separable Banach space.*

Proof. We start by proving that $(D, \|\cdot\|_D)$ is a Banach space, by using the following characterisation of completeness [Conway, 2007, Exercise III.4.2]. D is complete if and only if, for every sequence $f_n \in D$ such that

$$\sum_n \|f_n\|_D < \infty$$

the sum $\sum_{n=1}^N f_n$ converges in D.

So suppose that $\sum_n \|f_n\|_D < \infty$, then $\sum_n \|f\| < \infty$. Therefore, $\sum_n f_n \in C(E)$ as $(C(E), \|\cdot\|)$ is a Banach space. We need to show that $\sum_n f_n \in D$. By the definition of D, we need to check whether $\|\!|\sum_n f_n|\!\| < \infty$. But this follows from

$$\left|\!\left|\!\left| \sum_n f_n \right|\!\right|\!\right| \leq \sum_n \|\!|f_n|\!\| < \sum_n \|f_n\|_D < \infty.$$

So $(D, \|\cdot\|_D)$ is a Banach space and thus barrelled [Treves, 1967, Corollary 2 of Proposition 33.2].

We now prove separability of $(D, \|\cdot\|_D)$. For a finite box $\Lambda \subseteq S$, $w \in W^\Lambda$, and $\eta \in W^S$, define

$$\eta_\Lambda w_{\Lambda^c}(x) = \begin{cases} \eta_x & \text{for } x \in \Lambda \\ w_x & \text{for } x \in \Lambda^c \end{cases}.$$

Then define the local function $f_\Lambda \in D$ by $f_\Lambda(\eta) = f(\eta_\Lambda w_{\Lambda^c})$. Because f is uniformly continuous, these local functions approximate f with respect to $\|\cdot\|_D$ as can be seen from the following computation.

$$\|f - f_\Lambda\|_D = \sum_{x \in \Lambda^c} \Delta_f(x) + \|f - f_\Lambda\| \to 0.$$

For a fixed and finite region $\Lambda \subseteq S$, the norm $\|\cdot\|_D$ restricted to the local functions depending on coordinates in Λ is equivalent to the sup norm. Therefore, this set of local functions is separable. By taking a sequence of finite regions $\Lambda_n \to S$. We obtain that the set of local functions is separable. By the argument above, every function in D can be approximated by local functions in the $|||\cdot|||$ semi-norm, so indeed $(D, \|\cdot\|_D)$ is separable. □

Proposition 6.5.6. $(D, \|\cdot\|_D)$ *satisfies Conditions 6.1.2 and 6.1.3.*

Proof. Conditions 6.1.3 (a) and (b) follow from Lemma 6.5.5. Conditions 6.1.2 and 6.1.3 (c) follows from a number of straightforward calculations using the semi-norm $|||\cdot|||$.

By Theorem 6.5.2 (a) and (c), we obtain that $S(t) \in \mathcal{L}(D, \tau_D)$. An elementary calculation shows that for every $f \in D$ and $x \in S$, we have that $t \mapsto \Delta_{S(t)f}(x)$ is continuous. This implies, by using the Dominated convergence theorem and Theorem 6.5.2(c) that $t \mapsto S(t)f$ is continuous for $|||\cdot|||$. We conclude that $\{S(t)\}_{t \geq 0}$ is a strongly continuous semigroup for (D, τ_D).

Because $S(t)D \subseteq D$, Condition 6.1.2 (b') implies that also $V(t)D \subseteq D$. For a sequence of functions g_n that are uniformly bounded away from 0, we have that if $\|g_n - g\|_D \to 0$, then also $\|\log g_n - \log g\|_D \to 0$. Together with the continuity of $f \mapsto e^f$ by Condition 6.1.3 (c), we obtain the desired continuity properties of $V(t)$ from the properties of $S(t)$.

Condition 6.1.3 (e) is a direct consequence of Assumption (6.5.3) in Theorem 6.5.2. For (f), fix $f \in D$, then the function $\alpha \mapsto e^\alpha$ defined on $[-\|f\|, \|f\|]$ is Lipschitz continuous, with Lipschitz constant $e^{\|f\|}$. This means that $|e^\alpha - 1| \leq |\alpha| e^{\|f\|}$. Applying this to $\|Hf\|$, we obtain

$$
\begin{aligned}
\|Hf\| &\leq e^{2\|f\|}\, |||f||| \sum_{\Lambda} \left| \int c_\Lambda(\eta, \mathrm{d}\zeta) \left[f(\eta^\zeta) - f(\eta) \right] \right| \\
&\leq e^{|||f|||}\, |||f||| \sup_x \sum_{\Lambda \ni x} c_\Lambda
\end{aligned}
$$

Using that for $x \geq 0$ $xe^x \leq e^{2x}$, (f) is satisfied by taking

$$
\mathcal{N} := \left\{ f \in D \,\middle|\, |||f||| \leq -\frac{1}{2} \log \left(\sup_x \sum_{\Lambda \ni x} c_\Lambda \right) \right\}.
$$

□

Proposition 6.5.6 implies that Theorem 6.1.8 holds for interacting particle systems where the derivative of the trajectory $t \mapsto \mu(t)$ lies in D'.

Because we can always choose $\mathcal{N}$ in Condition 6.1.3 such that it contains all constant functions, we can restrict our attention to $(D/\mathcal{C})'$, where $\mathcal{C}$ is the space of constant functions. This is reasonable, because the only derivatives of a path of probability measures that we will find satisfy $\langle \mathbb{1}, u \rangle = 0$. In the next section, we give a representation theorem for $(D/\mathcal{C})'$.

6.5.2.1 *A representation theorem for* $((D/\mathcal{C})', \|\!|\cdot|\!\|)$

We identify the dual of $D/\mathcal{C}$ the space of equivalence classes $D/\mathcal{C} \subseteq C(E)/\mathcal{C}$, where $\mathcal{C} := \{c\mathbb{1} \mid c \in \mathbb{R}\}$. Additionally, we equip $D/\mathcal{C}$ with the norm $\|\!|\cdot|\!\|$, which is equivalent to the quotient norm $\|\cdot\|_{D,Q} = \|\cdot\|_Q + \|\!|\cdot|\!\|$ as $\|\!|\cdot|\!\| \leq \|\cdot\|_{D,Q} \leq \frac{3}{2} \|\!|\cdot|\!\|$ by Lemma 6.5.4.

We consider the dual of $D/\mathcal{C}$, which is equipped with the operator norm

$$\|\!|\alpha|\!\| = \sup_{f \in D/\mathcal{C}} \frac{|\langle f, \alpha \rangle|}{\|\!|f|\!\|}.$$

The goal of the following discussion is to identify both this dual space and its norm. First of all, the dual $(D/\mathcal{C})'$ can be seen as a subspace of functionals on D that are constant on the equivalence classes $f + \mathcal{C}$. Therefore,

$$\|\!|\alpha|\!\| = \sup_{f \in D} \frac{\langle f, \alpha \rangle}{\|\!|f|\!\|},$$

for α such that $\langle \mathbb{1}, \alpha \rangle = 0$.

We introduce some notation. For $\Lambda \subseteq S$, let $\mathcal{E}_\Lambda := \sigma(\eta_x \mid x \in \Lambda)$. Furthermore, $\tilde{\Pi}$ is the space of additive set functions α on the algebra $\mathcal{E}_a := \bigcup_{\Lambda : |\Lambda| < \infty} \mathcal{E}_\Lambda$, for which it holds that $\alpha(E) = 0$. Note that the σ-algebra $\mathcal{E}$ is given by $\sigma(\mathcal{E}_a)$.

For $\alpha \in \tilde{\Pi}$ and a finite subset $\Lambda \subseteq S$, we denote the restriction of α to $\mathcal{E}_\Lambda$ by $P_\Lambda \alpha$ and we set $P_x := P_{\{x\}}$. Also, we define the function $\|\alpha\|_\Pi = \sup_x \|P_x \alpha\|_{TV}$ taking values in $[0, \infty]$.

Definition 6.5.7. Let Π be the set

$$\Pi := \left\{ \alpha \in \tilde{\Pi} \,\middle|\, \|\alpha\|_\Pi < \infty \right\}.$$

It follows that Π is a vector space and that $\|\cdot\|_\Pi$ is a norm on Π. The following technical lemma enables us to show that $(\Pi, \|\cdot\|_\Pi)$ is a Banach space.

Lemma 6.5.8. *For a finite set $T \subseteq S$: $\|P_\Lambda \alpha\|_{TV} \leq |\Lambda| \, \|\alpha\|_\Pi$.*

Proof. Pick a local function f with dependence set $\mathcal{D}(f) = \{x_1, \ldots, x_n\}$, $\sup_\eta f(\eta) = 2\,\|f\|_Q$ and $\inf_\eta f(\eta) = 0$. Pick ζ such that $f(\zeta) = 0$, and define for $k \leq n$ the sets $\Lambda_k = \{x_1, \ldots, x_k\}$. For $\eta \in E$, let $\eta(k)$ be equal to ζ on Λ_k, and equal to η outside Λ_k. Furthermore, let $f_k(\eta) = f(\eta_k)$. Then it follows that

$$\begin{aligned}
\int f \mathrm{d}\alpha &= \int f_0(\eta) - f_n(\eta) \mathrm{d}\alpha(\eta) \\
&= \sum_{k=0}^{n-1} \int f_k(\eta) - f_{k+1}(\eta) \mathrm{d}\alpha(\eta) \\
&\leq \frac{1}{2} \sum_{k=0}^{n-1} \Delta_f(x_{k+1}) \left\|P_{x_{k+1}} \alpha\right\|_{TV} \qquad (6.5.5) \\
&\leq \sum_{k=0}^{n-1} \|f\| \left\|P_{x_{k+1}} \alpha\right\|_{TV} \\
&\leq n \, \|f\| \, \|\alpha\|_\Pi .
\end{aligned}$$

The bound obtained in line three of (6.5.5) is stronger then necessary, for this lemma, but we will use it again for the proof of Theorem 6.5.10. □

Lemma 6.5.9. $(\Pi, \|\cdot\|_\Pi)$ *is a Banach space.*

Proof. We apply exercise III.4.2 in Conway Conway [2007] that states that that $(\Pi, \|\cdot\|_\Pi)$ is complete if we can show for an arbitrary sequence $(\alpha_n)_{n\in\mathbb{N}}$ in Π such that $\sum_n \|\alpha_n\|_\Pi < \infty$, that the partial sums $\sum_n \alpha_n$ converge in Π.

So pick a sequence α_n in Π such that $\sum_n \|\alpha_n\|_\Pi < \infty$. Furthermore, take a sequence of finite sets Λ_k that is increasing to S. By Lemma 6.5.8, we see that

$$\sum_n \|P_{\Lambda_k} \alpha_n\|_{TV} \leq |\Lambda_k| \sum_n \|\alpha_n\|_\Pi < \infty.$$

The space of measures on $\mathcal{E}_{\Lambda_k}$ of bounded variation is a Banach space. Hence, $\alpha_{\Lambda_k} := \sum_n P_{\Lambda_k} \alpha_n$ exists and is a measure of bounded variation on $\mathcal{E}_{\Lambda_k}$. Furthermore, it is easy to see that this leads to a consistent sequence in k, so there exists a additive set function α on $\bigcup_{\Lambda : |\Lambda| < \infty} \mathcal{E}_\Lambda$, which, if restricted to finite regions, is a measure of bounded variation.

It follows that $\left\|\alpha - \sum_{k=1}^{N} \alpha_k\right\|_\Pi \to 0$, because

$$\left\| P_x \left(\alpha - \sum_{n=1}^{N} \alpha_n \right) \right\|_{TV} = \left\| P_x \left(\sum_{n=N+1}^{\infty} \alpha_n \right) \right\|_{TV} \leq \sum_{n=N+1}^{\infty} \|\alpha_n\|_\Pi \to 0.$$

□

We are now able to prove a representation theorem for $((D/\mathcal{C})', \|\!|\cdot|\!\|)$.

Theorem 6.5.10. $((D/\mathcal{C})', \|\!|\cdot|\!\|) = (\Pi, \frac{1}{2}\|\cdot\|_\Pi)$, *hence,* $\|\!|\alpha|\!\| = \frac{1}{2}\sup_x \|P_x\alpha\|_{TV}$.

Proof. First, we show that $(D/\mathcal{C})'$ can be seen as a space of set functions. Take a finite set $\Lambda_0 \subseteq S$, then restricted the space $D_{\Lambda_0} := \{f \in D \,|\, \mathcal{D}(f) \subseteq \Lambda_0\}$ α is a continuous and linear function.

The space D_{Λ_0} with the topology induced by $\|\!|\cdot|\!\|$ is isomorphic to $C(W^{\Lambda_0})$ with the topology induced by $\|\cdot\|_Q$, as

$$2\|\cdot\|_Q \leq \|\!|\cdot|\!\| \leq 2|\Lambda_0|\,\|\cdot\|_Q\,.$$

Therefore, by the Riesz representation theorem, Theorem 7.10.4 in Bogachev [2007], it follows that for $f \in D_{\Lambda_0}$, $\alpha(f) = \langle f, \hat{\alpha}_{\Lambda_0}\rangle$ where $\hat{\alpha}_{\Lambda_0}$ is a measure of bounded variation on $\mathcal{E}_{\Lambda_0}$ such that $\hat{\alpha}_{\Lambda_0}(E) = 0$. This can be done consistently for every finite set $\Lambda \subseteq S$, which implies that $\hat{\alpha}$ can be seen as a set function on $\bigcup_{\Lambda:|\Lambda|<\infty} \mathcal{E}_\Lambda$ for which the restriction to finite regions is a measure of bounded variation.

We proceed by showing that $\|\!|\alpha|\!\| \geq \frac{1}{2}\sup_x \|P_x\alpha\|_{TV}$. For $x \in S$, let $C(W^{\{x\}})$ be the set of continuous functions on W, but seen as local functions in D which depend only on the coordinate η_x.

$$\begin{aligned}\|\!|\alpha|\!\| &= \sup_{f\in D} \frac{|\langle f,\alpha\rangle|}{\|\!|f|\!\|} \geq \sup_{f\in C(W^{\{x\}})} \frac{|\langle f,\alpha\rangle|}{\|\!|f|\!\|} \\ &= \sup_{f\in C(W^{\{x\}})} \frac{|\langle f,\alpha\rangle|}{2\|f\|_Q} = \sup_{f\in C(W^{\{x\}})} \frac{|\langle f,\hat{\alpha}\rangle|}{2\|f\|_Q} = \frac{1}{2}\|P_x\hat{\alpha}\|_{TV}\end{aligned}$$

This means that the function $\Phi : (B/\mathcal{C})' \to \Pi$, mapping α to $\hat{\alpha}$, is well defined, injective and continuous. So, we identify α and $\hat{\alpha}$.

For the other inequality note that by continuity we can restrict the supremum to local functions:

$$\|\alpha\| = \sup_{f \text{ local}} \frac{|\langle f, \alpha\rangle|}{\|f\|}.$$

For local functions f, the result in (6.5.5) yields:

$$\frac{|\langle f, \alpha\rangle|}{\|f\|} \leq \frac{\|f\| \frac{1}{2} \sup_x \|P_x\alpha\|_{TV}}{\|f\|} = \frac{1}{2} \sup_x \|P_x\alpha\|_{TV}.$$

This means that Φ is an isometry with respect to $\|\cdot\|$ and $\frac{1}{2}\|\cdot\|_{\Pi}$. We show that it is also surjective. Pick a local function f, then clearly $\alpha(f)$ is well defined, because α restricted to $\mathcal{E}_{\mathcal{D}(f)}$ is a measure of bounded variation. By the calculation above we see that $|\langle f, \alpha\rangle| \leq \frac{1}{2}\|f\| \sup_x \|P_x\alpha\|_{TV}$. Hence, α defines a bounded linear functional on the local functions. Thus, it extends by continuity to a continuous linear functional on $D/\mathcal{C}$. □

6.5.3 *Diffusion processes on* $\mathbb{R}^d$

We now show that our result partly reproduces the Dawson and Gärtner theorem Dawson and Gärtner [1987]. First of all, we prove the result for a time-homogeneous case, but more importantly, we need to assume more regularity on the diffusion and drift terms.

Let $C_0^m(\mathbb{R}^d)$ be the space of m times continuously differentiable functions, for which all derivatives up to order m are in $C_0(\mathbb{R}^d)$.

For every $x \in \mathbb{R}^d$, let $\{\sigma_{i,j}(x)\}_{i,j}$ be non-negative definite matrices, $\sigma_{i,j}(x)$ continuous in x. Denote with $a_{i,j}(x) = \sigma_{i,j}(x)\sigma_{i,j}(x)^T$. For each i, let $b_i \in C(\mathbb{R}^d)$. Define for every $f \in C_c^\infty(\mathbb{R}^d)$ the infinitesimal operator

$$Af(x) = \frac{1}{2}\sum_{i,j} a_{i,j}(x)\partial_i\partial_j f(x) + \sum_i b_i(x)\partial_i f(x).$$

Denote with

$$C_b^m(\mathbb{R}^d) := \left\{ f \in C_b(\mathbb{R}^d) \,\middle|\, D^\alpha f \in C_b(\mathbb{R}^d) \text{ if } |\alpha| \leq m \right\},$$
$$C_0^m(\mathbb{R}^d) := \left\{ f \in C_0(\mathbb{R}^d) \,\middle|\, D^\alpha f \in C_0(\mathbb{R}^d) \text{ if } |\alpha| \leq m \right\},$$

and equip $C_0^m(\mathbb{R}^d)$ with the norm

$$\|f\|_m := \sum_{0\leq|\alpha|\leq m} \|D^\alpha f\|.$$

Using the methods obtained to prove Theorem 8.2.5 in Ethier and Kurtz [1986], we obtain the following theorem.

Theorem 6.5.11. *If $\sigma_{i,j}, b_i \in C_b^3(\mathbb{R}^d)$, then the closure of A generates a strongly continuous contraction semigroup on $C_0(\mathbb{R}^d)$. Additionally, $S(t)C_0^{(}\mathbb{R}^d) \subseteq C_0^2(\mathbb{R}^d)$ and the restriction of $S(t)$ to $C_0^2(\mathbb{R}^d)$ is strongly continuous for $\|\cdot\|_2$.*

We calculate $A^g f$ and Hf for $f, g \in \mathcal{D} = C_0^2(\mathbb{R}^d)$. Again, the calculation of A^g gives us a new generator with a changed drift.

$$
\begin{aligned}
A^g f(x) &= Af(x) + \sum_{i,j} \frac{a_{i,j}(x) + a_{j,i}(x)}{2} \partial_j g(x) \partial_i f(x) \\
&= \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_i \partial_j f(x) \\
&\quad + \sum_i \left(b_i(x) + \sum_j \frac{a_{i,j}(x) + a_{j,i}(x)}{2} \partial_j g(x) \right) \partial_i f(x).
\end{aligned}
$$

Hf introduces a quadratic term:

$$
Hf(x) = Af(x) + \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_i f(x) \partial_j f(x). \tag{6.5.6}
$$

As a corollary to Theorem 6.5.11, we obtain the next result.

Corollary 6.5.12. *$(C_0^2(\mathbb{R}^d), |||\cdot|||_2$ satisfies Condition 6.1.3.*

Proof. Conditions (a) to (e) are straightforward to check. For (f), we put

$$
\begin{aligned}
\mathcal{N} := \Bigg\{ f \in C_0^2(\mathbb{R}^d) \,\Bigg|\, & \text{ for all } x \in \mathbb{R}^d, \text{ we have } d \sup_i |b_i(x)||\partial_i f(x)| \\
& + \frac{d^2}{2} \sup_{i,j} |a_{i,j}(x)| \left(|\partial_i \partial_j f(x)| + |\partial_i f(x)||\partial_j f(x)| \right) \leq 1 \Bigg\}.
\end{aligned}
$$

Clearly, $\mathcal{N}$ is closed, convex and balanced. We prove that $\mathcal{N}$ is absorbing, which follows by showing that $\mathcal{N}$ contains a ball $\{f \in C_0^2(\mathbb{R}^d) \mid \|f\|_2 \leq c\}$ for some c.

Let $\bar{a} = \sup_{i,j} \sup_{x \in \mathbb{R}^d} |a_{i,j}(x)|$ and $\bar{b} = \sup_i \sup_{x \in \mathbb{R}^d} |b_i(x)|$. Pick $c > 0$ such that

$$\frac{d^2}{2}\bar{a}(c^2 + c) + d\bar{b}c \leq 1.$$

This choice implies that

$$\left\{ f \in C_0^2(\mathbb{R}^d) \,\middle|\, \|f\|_2 \leq c \right\} \subseteq \mathcal{N} \cap C^2(K_n).$$

We obtain that $\mathcal{N}$ is a barrel and by construction of $\mathcal{N}$ and the form of H, see (6.5.6), that $\sup_{f \in \mathcal{N}} \|Hf\| \leq 1$. Also, for $c \geq 1$, we obtain $\sup_{f \in c\mathcal{N}} \|Hf\| \leq c^2$. □

A similar approach would give the result for $D = \mathcal{S}$ the space of rapidly decreasing smooth functions with its Fréchet space topology. This would need the extension for separable barrelled spaces in Condition 6.1.3. Note that $\mathcal{S}$ is separable by the discussion following Proposition A.9 in Treves [1967] and barrelled by Corollary 1 of Proposition 33.2 in Treves [1967]

6.5.4 *The Dawson and Gärtner theorem*

As a consequence of the discussion above, we re-obtain a time-homogeneous and smooth version of Theorem 4.5 by Dawson and Gärtner Dawson and Gärtner [1987].

Let $(x_1, \ldots, x_d)$ be the Euclidean coordinates. For $f \in C_0^2(\mathbb{R}^d)$, define $(\nabla f)^i = \sum_{j=1^d} a_{i,j}(\cdot) \frac{\partial f}{\partial x_j}$. Then it follows from (6.5.6) that $\langle Hf, \mu \rangle = \langle Af, \mu \rangle + \frac{1}{2} \langle |\nabla f|^2, \mu \rangle$.

We introduce two new spaces,

$$D_\mu := \left\{ f \in D = C_0^2(\mathbb{R}^d) \,\middle|\, \langle |\nabla f|^2, \mu \rangle \neq 0 \right\}$$
$$T_\mu := \left\{ \alpha \in C_0^2(\mathbb{R}^d)' \,\middle|\, \|\alpha\|_\mu < \infty \right\},$$

where $\|\cdot\|_\mu$ is defined on $C_0^2(\mathbb{R}^d)'$ by

$$\|\alpha\|_\mu := \sup_{f \in D_\mu} \frac{|\langle f, \alpha \rangle|^2}{\langle |\nabla f|^2, \mu \rangle}.$$

The next proposition shows the connection between Theorem 6.1.8 and Theorem 4.5 by Dawson and Gärtner Dawson and Gärtner [1987].

Proposition 6.5.13. *If $\mathcal{L}(\mu,\alpha) < \infty$, then $\alpha \in T_\mu$ and $\mathcal{L}(\mu,\alpha) = \frac{1}{2}\left\|\alpha - A'\mu\right\|_\mu$. As a consequence for a trajectory $\nu \in \mathcal{AC}$*

$$\int_0^\infty \mathcal{L}(\nu(s),\dot\nu(s))\mathrm{d}s = \frac{1}{2}\int_0^\infty \left\|\dot\nu(s) - A'\nu(s)\right\|_{\nu(s)}.$$

Proof. Pick $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\alpha \in C_0^2(\mathbb{R}^d)'$ such that $\mathcal{L}(\mu,\alpha) < \infty$. Define $\hat\alpha = \alpha - A'\mu$. Consider

$$\begin{aligned}
\mathcal{L}(\mu,\alpha) &= \sup_{f \in C_0^2(\mathbb{R}^d)} \left\{\langle f,\alpha\rangle - \langle Hf,\mu\rangle\right\} \\
&= \sup_{f \in C_0^2(\mathbb{R}^d)} \left\{\langle f,\alpha\rangle - \langle Af,\mu\rangle - \frac{1}{2}\langle|\nabla f|^2,\mu\rangle\right\} \\
&= \sup_{f \in C_0^2(\mathbb{R}^d)} \left\{\langle f,\hat\alpha\rangle - \frac{1}{2}\langle|\nabla f|^2,\mu\rangle\right\} \\
&= \sup_{f \in C_0^2(\mathbb{R}^d)} \sup_{c\in\mathbb{R}} \left\{c\langle f,\hat\alpha\rangle - c^2\frac{1}{2}\langle|\nabla f|^2,\mu\rangle\right\} \qquad (6.5.7)
\end{aligned}$$

By assumption, the supremum in the equation above is finite. Then if $f \in D_\mu^c$, it must be that $\langle f,\hat\alpha\rangle = 0$. Therefore, these f yield 0 as an argument in the supremum.

For a given $f \in D_\mu$, optimising over c yields $c = \frac{\langle f,\hat\alpha\rangle}{\langle|\nabla f|^2,\mu\rangle}$. Therefore, we can rewrite (6.5.7) as

$$\mathcal{L}(\mu,\alpha) = 0 \vee \frac{1}{2}\sup_{f\in D_\mu}\frac{|\langle f,\hat\alpha\rangle|^2}{\langle|\nabla f|^2,\mu\rangle} = \frac{1}{2}\left\|\alpha - A'\mu\right\|_\mu.$$

□

6.6 APPENDIX: DECOMPOSITION OF THE RATE FUNCTION ON PRODUCT SPACES

In this appendix, (E,d) is a complete separable metric space.

Suppose $\mathbb{P}$ is the law of a Markov process on $D_E(\mathbb{R}^+)$. Suppose that the sequence $(L_n^{X(0)},\ldots L_n^{X(t_k)})$ satisfies the large deviation principle on $\mathcal{P}(E)^{k+1}$. The following lemma is a multidimensional version of exercise 6.2.26 of Dembo and Zeitouni [1998].

Lemma 6.6.1. *The large deviation rate function* $I[0, t_1, \dots, t_k]$ *of the LDP of the sequences* $(L_n^{X(0)}, \dots L_n^{X(t_k)})$ *on* $\mathcal{P}(E)^{k+1}$ *is given by*

$$I[0, t_1, \dots, t_k](\nu_0, \dots, \nu_k) = \sup_{f_0, \dots, f_k \in C_b(E)} \sum_{i=0}^{k} \langle f_i, \nu_i \rangle - \log \mathbb{E}\left[e^{f_0(X(0)) + \dots + f_k(X(t_k))} \right]. \tag{6.6.1}$$

Also, we can restrict to a smaller class of functions, see [Dembo and Zeitouni, 1998, Definition 4.4.7 and exercise 4.4.14].

Corollary 6.6.2. *The supremum over* $C_b(E)$ *in* (6.6.1) *can be restricted to any class of functions* M *that separates points and is closed under taking point-wise minima. In particular, this holds for* $C_0(E)$ *if* E *is locally compact.*

Denote with $V(s,t)f(x) = \log \mathbb{E}_{X(s)=x}\left[e^{f(X(t))}\right]$ and put

$$I_{t_1,t_2}(\nu_1 \,|\, \nu_0) = \sup_{f \in M} \langle f, \nu_1 \rangle - \langle V(t_1, t_2) f, \nu_0 \rangle.$$

Clearly, if X is a time-homogeneous process, we can simplify to $V(t-s) := V(s,t)$ and $I_{t_2 - t_1} := I_{t_1,t_2}$. The following proposition can be verified in a straightforward way, see for example Lemma 4.7 in Dawson and Gärtner [1987].

Proposition 6.6.3. *Let* $M \subseteq C_b(E)$ *be a set of functions that separates points, and which is closed under taking point-wise minima. Denote with* $V(s,t)f(x) = \log \mathbb{E}_{X(s)=x}\left[e^{f(X(t))}\right]$ *and let* M *be such that for every* $t \geq 0$: $V(t)M \subseteq M$. *Define*

$$I_{t_1,t_2}(\nu_1 \,|\, \nu_0) = \sup_{f \in M} \langle f, \nu_1 \rangle - \langle V(t_1, t_2) f, \nu_0 \rangle.$$

Then, it holds that

$$I\left[0, t_1, \dots, t_k\right](\nu_0, \dots, \nu_k) = I_0(\nu_0) + \sum_{i=1}^{k} I_{t_{i-1},t_i}(\nu_i \,|\, \nu_{i-1}).$$

6.7 APPENDIX: SOUSLIN SPACES, BARRELLED SPACES, AND GELFAND INTEGRATION

6.7.1 *Barrelled spaces and Gelfand integration*

Definition 6.7.1. A locally convex space $\mathcal{X}$ is called barrelled if every barrel is a neighbourhood of 0. A set S is a barrel if it is convex, balanced, absorbing and closed. S is balanced if we have the following: if $x \in S$ and $\alpha \in \mathbb{R}$, $|\alpha| \leq 1$ then $\alpha x \in S$. S is absorbing if for every $x \in \mathcal{X}$ there exists a $r \geq 0$ such that if $|\alpha| \geq r$ then $x \in \alpha S$.

For example, Banach, Fréchet and LF(limit Fréchet) spaces are barrelled [Treves, 1967, Chapter 33]. The space of Schwartz functions is Fréchet and the space $C_c^\infty(\mathbb{R}^d)$ with is usual topology is LF.

The importance of barrelled spaces follows from the fact that the closed graph theorem holds for them [Carreras and Bonet, 1987, Proposition 7.1.11], [Robertson and Robertson, 1973, Theorem VI.7].

Theorem 6.7.2 (Closed graph theorem)**.** *Let $\mathcal{X}$ be a barrelled locally convex space, and let F be a Fréchet space. Suppose that $T : F \to \mathcal{X}$ is a linear operator with closed graph in $F \times \mathcal{X}$, then T is continuous.*

The closed graph theorem is of importance for integration of functions with values the dual of a barrelled space. Let $(\Omega, \mathcal{F}, \mu)$ be a complete and finite measure space, and let $\mathcal{X}$ be a barrelled space with continuous dual $\mathcal{X}'$. We equip X' with $\sigma(\mathcal{X}', \mathcal{X})$, the weak* topology.

Definition 6.7.3. A function $f : \Omega \to \mathcal{X}'$ is called weak* measurable if the scalar function

$$\omega \mapsto \langle x, f(\omega) \rangle$$

is $\mathcal{F}$ measurable for every $x \in \mathcal{X}$. Such a function f is called *Gelfand* or weak* integrable if $\langle x, f \rangle \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ for every $x \in \mathcal{X}$.

For Gelfand integrable functions, we obtain, using the Closed graph theorem, the following result [Diestel and Uhl, 1977, pages 52-53].

Theorem 6.7.4. *Let $\mathcal{X}$ be a barrelled space and $(\Omega, \mathcal{F}, \mu)$ a complete and finite measure space. For every measurable set $A \in \mathcal{F}$ and Gelfand integrable function $f : \Omega \to \mathcal{X}'$, there exists a unique $x'_A \in \mathcal{X}'$ such that*

$$\langle x, x'_A \rangle = \int_A \langle x, f(\omega) \rangle \mu(\mathrm{d}\omega)$$

for all $x \in \mathcal{X}$. This element x'_A will be denoted by $\int_A f \mathrm{d}\mu$.

6.7.2 *Souslin spaces*

Definition 6.7.5. A space (Y, τ_Y) is called Souslin, if $Y = f(X)$ for some complete separable metric space (X, τ_X) and some continuous function $f : (X, \tau_X) \to (Y, \tau_y)$.

For more background on Souslin spaces, see Chapters 6 and 7 in Bogachev [2007].

Lemma 6.7.6. *Let (X, τ) be a separable barrelled locally convex Hausdorff space and T a barrel in (X, τ). Then $(\bigcup_n nT^\circ, wk^*) \subseteq (X', wk^*)$ is a Souslin space.*

In particular, as the unit ball in a Banach space B is a barrel, the dual (B', wk^*) of separable Banach space is Souslin.

Proof. Because (X, τ) is barrelled, T is a neighbourhood of 0. Consequentially, T° is an equi-continuous set in (X', wk^*) by 21.3.(1) in Köthe [1969]. By the Bourbaki-Alaoglu theorem, 20.9.(4) Köthe [1969], this set is weak* compact.

Furthermore, by 39.4.(7) in Köthe [1979], T° is metrisable. (T°, wk^*) is compact and metric, which implies that it is complete separable metric and as a consequence Souslin. We can do the same for $n\mathcal{N}^\circ$ for every $n \in \mathbb{N}$, so we obtain that $(\bigcup_n n\mathcal{N}^\circ, wk^*)$ is Souslin [Bogachev, 2007, Theorem 6.6.6]. □

7 LARGE DEVIATIONS ON THE PROCESS LEVEL

This Chapter is based on work jointly with Frank Redig.

In Chapter 3, we considered large deviations for the empirical magnetization of a mean-field spin-flip model. The main example in this chapter is a nearest-neighbour spin-flip model on the lattice $\mathbb{Z}^d$. To obtain a path-space large deviation principle to study the evolution of the magnetization, it is not sufficient to only consider the evolution of the magnetization as this evolution is not autonomous.

Therefore, to study the behaviour for large n, one replaces the empirical magnetization by the empirical measures. In the limit, these empirical measures satisfy a autonomous equation, which makes this evolution a suitable object to study. The main result in this chapter is a large deviation principle for the trajectories of the empirical measure around the autonomous equation.

We first consider the much studied fixed time issue, see for example Georgii [2011] or Pfister [2002].

7.1 FIXED TIME PROCESS LEVEL LARGE DEVIATIONS

Let W be a compact metric space and denote with E the product space $E = W^{\mathbb{Z}^d}$. In examples, W will be a finite set (interacting particle systems), or $W = I^n$, for some bounded interval I, or more generally a compact finite dimensional manifold (interacting diffusions). Elements of E are denoted by Greek letters η, σ, ξ. For a configuration $\eta \in E$ and $i \in \mathbb{Z}^d$, η_i denotes evaluation of η in i. On E we have translations defined by $(\theta_i \eta)_j = \eta_{i+j}$. We set $\mathcal{P}_\theta(E)$ to be the set of *translation invariant measures*, i.e. measures μ such that $\mu \circ \theta_i = \mu$ for all $i \in \mathbb{Z}^d$. We say that $\mu \in \mathcal{P}_\theta(E)$ is *ergodic* if it is an extreme element of the convex set $\mathcal{P}_\theta(E)$. In other words, μ is ergodic if and only if we have that $\mu = c\nu_1 + (1-c)\nu_2$ for $\nu_1, \nu_2 \in \mathcal{P}_\theta(E)$ and $c \in (0,1)$, then $\mu = \nu_1 = \nu_2$. We denote the set of all ergodic measures by $\mathcal{P}_e(E)$.

Let $\Lambda_n \subset \mathbb{Z}^d$ be the box defined by $\Lambda_n := \mathbb{Z}^d \cap [-n, n]^d$ and for $A \subseteq \mathbb{Z}^d$, let $\mathcal{B}_A$ be the σ-algebra generated by the variables η_i, for $i \in A$. Denote with $\mathcal{B}_n := \mathcal{B}_{\Lambda_n}$.

For $\sigma \in E$, consider the empirical measures $L_n(\sigma)$, defined by

$$L_n(\sigma) := \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} \delta_{\theta_i \sigma}$$

We are interesting in the large deviation behaviour of the sequence $L_n(\sigma)$ in the setting that σ has an ergodic distribution $\mu \in \mathcal{P}_\theta(E)$. If μ is an ergodic measure, it follows by the Ergodic theorem that the sequence $L_n(\sigma)$ converges weakly μ almost surely to μ. Thus, compared to earlier chapters of this thesis, the large deviations of L_n are around the ergodic limit instead of the usual law of large numbers.

Instead of studying L_n, one can also study the large deviation behaviour of averages of shifts of a periodization of σ. As noted in Chapter 6 of Rassoul-Agha and Seppäläinen [2015], this sequence of objects and $L_n(\sigma)$ and are equivalent on an exponential scale, so their large deviation behaviour is the same.

To describe the rate function and conditions to obtain the large deviation principle, we introduce some more notation. As in Chapter 6, we denote the relative entropy $S(\nu \mid \mu)$, for two measures $\mu, \nu \in \mathcal{P}(E)$, by

$$S(\nu \mid \mu) = \begin{cases} \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathrm{d}\nu & \text{if } \nu << \mu \\ \infty & \text{otherwise.} \end{cases}$$

We will denote by $S_n(\nu \mid \mu) = S_n(\nu_n \mid \mu_n)$, where μ_n, ν_n are the measures μ and ν restricted to $\mathcal{B}_n$.

Because L_n is obtained by 'dividing out' translations, and if the measure μ satisfies the natural condition of asymptotic decoupledness, to be introduced below, it is to be expected that the corresponding rate function should be the entropy density.

Definition 7.1.1. Let $\mathcal{X} = X^{\mathbb{Z}^d}$ for some Polish space X. Fix $\nu, \mu \in \mathcal{P}_\theta(\mathcal{X})$. If the limit

$$\lim_{n \to \infty} \frac{1}{|\Lambda_n|} S_n(\nu \mid \mu)$$

exists, we call it the *relative entropy density* and denote this limit by $s(\nu \mid \mu)$.

Definition 7.1.2. Let $\mathcal{X} = X^{\mathbb{Z}^d}$ for some Polish space X. A probability measure $\mu \in \mathcal{P}(\mathcal{X})$ is called *asymptotically decoupled* (AD) if there exists sequences $d(n)$, $c(n)$ such that

$$\lim_{n\to\infty} \frac{c(n)}{|\Lambda_n|} = 0, \quad \lim_{n\to\infty} \frac{d(n)}{n} = 0 \tag{7.1.1}$$

and for all $i \in \mathbb{Z}^d$, $A \in \mathcal{F}_{i+\Lambda_n}$ and $B \in \mathcal{F}_{(i+\Lambda_{n+d(n)})^c}$, such that $\mu(A)\mu(B) \neq 0$:

$$\begin{aligned} e^{-c(n)} \leq & \frac{\mu(A\cap B)}{\mu(A)\mu(B)} \\ & \frac{\mu(A\cap B)}{\mu(A)\mu(B)} \leq e^{c(n)}. \end{aligned}$$

The first line respectively second line refer to AD from below respectively AD from above.

Note that the class of asymptotically decoupled measures includes the class of Gibbs measures. In [Pfister, 2002, Proposition 3.2] it is shown that the large deviation principle can be proven for measures satisfying the AD property.

Proposition 7.1.3. *Let $\mathcal{X} = X^{\mathbb{Z}^d}$ for some Polish space X. Let $\mu \in \mathcal{P}_\theta(\mathcal{X})$ be asymptotically decoupled. Then the limit*

$$s(\nu \,|\, \mu) = \lim_{n\to\infty} \frac{1}{|\Lambda_n|} S_n(\nu \,|\, \mu)$$

exists for all $\nu \in \mathcal{P}_\theta(E)$. Additionally, the function $\nu \mapsto s(\nu \,|\, \mu)$ is lower semi-continuous with respect to the weak topology and has weakly compact level sets.

For an asymptotically decoupled measure μ, we extend $s(\cdot \,|\, \mu)$ to $\mathcal{P}(E)$ setting $s(\nu \,|\, \mu) = \infty$ for $\nu \in \mathcal{P}(E) \setminus \mathcal{P}_\theta(E)$. With this extension, Pfister obtained the large deviation principle for L_n.

Theorem 7.1.4 (Theorem 3.3 in Pfister [2002])**.** *Let μ be asymptotically decoupled. Then the sequence $\{L_n\}_{n\geq 1}$ satisfies the large deviation principle on $\mathcal{P}(E)$ with normalization $|\Lambda_n|$ and rate function $\nu \mapsto s(\nu \,|\, \mu)$.*

7.2 TRANSLATION INVARIANT DYNAMICS

We now introduce dynamics. In particular, we consider a translation invariant Feller process $\{\sigma(t)\}_{t\geq 0}$ on E, having a transition semigroup $\{S(t)\}_{t\geq 0}$ on $C(E)$. As before, we write $(A, \mathcal{D}(A))$ for the generator of $S(t)$. To aid the exposition and to stress the translation invariance of A, we will assume that there exists some 'source' generator $(A_0, \mathcal{D}(A_0))$ with $\mathcal{D}(A) \subseteq \mathcal{D}(A_0)$ such that

$$A = \sum_{i\in\mathbb{Z}^d} \theta_{-i} A_0 \theta_i.$$

The main non-trivial setting where our results hold is in the context of spin-flip dynamics. On the other hand, for product dynamics, the results also hold for diffusion processes.

Example 7.2.1 (Spin-flip system)**.** We consider a spin-flip system on $\{-1,1\}^{\mathbb{Z}^d}$. As in Chapter 6, let D be the set of functions with bounded triple norm $|\!|\!|\cdot|\!|\!|$. For $f \in D$, we consider a generator of the type

$$Af(\sigma) = \sum_{i\in\mathbb{Z}^d} c_i(\sigma)\left[f(\sigma^i) - f(\sigma)\right].$$

We will assume that c_i is continuous in σ and that c_i only depends on a finite set of coordinates close to i. The configuration σ^i is defined by

$$\sigma^i_j = \begin{cases} \sigma_j & \text{if } j \neq i, \\ -\sigma_j & \text{if } j = i. \end{cases}$$

The spin-flip process generated by this generator is translation invariant if $c_i(\sigma) = c_0(\theta_{-i}\sigma)$ and in this case, we have $A = \sum_i \theta_{-i} A_0 \theta_i$ with A_0 given by

$$A_0 f(\sigma) = c_0(\sigma)\left[f(\sigma^0) - f(\sigma)\right].$$

Example 7.2.2 (Diffusion processes on a torus)**.** Set $E := (\mathbb{R}/\mathbb{Z})^{\mathbb{Z}^d}$. In this context, let D be the space of functions f that only depend on a finite set of coordinates $\mathcal{D}(f)$, and that are twice continuously differentiable for coordinates in $\mathcal{D}(f)$. We consider the product dynamics defined by the generator A with source generator A_0 given by

$$A_0 f(x) = \frac{\mathrm{d}^2}{\mathrm{d}x_0^2} f(x), \qquad f \in D,$$

where x_0 is the 0 coordinate of $x \in E$.

Below, we will prove, under some conditions, the large deviation principle for $\{L_n\}_{n\geq 0}$, where $L_n := \{L_n(t)\}_{t\geq 0} \in D_{\mathcal{P}(E)}(\mathbb{R}^+)$ and where $L_n(t) := \frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n} \delta_{\theta_i\sigma(t)}$.
The representation of the rate function will be of the form as in Theorem 2.4.10. To do this, we will construct a limiting semigroup $V(t)$. Because this semigroup should correspond to dynamics on $\mathcal{P}_\theta(E)$, it should act on the dual of $\mathcal{M}_\theta(E)$. A variational representation of $V(t)$, as in other Chapters of this thesis has not been obtained yet. Some ideas on this problem are mentioned in Section 7.5.

Denote by $C_\theta(E)$ the space $C(E)/\mathcal{I}$, where

$$\mathcal{I} := \{f \in C(E) \,|\, |\langle f, \mu\rangle| = 0 \text{ for all } \mu \in \mathcal{P}_\theta(E)\}.$$

In Section 7.6.2, we explore some properties of this space. Importantly, the quotient norm $\|\cdot\|_\theta$ turns $(C_\theta(E), \|\cdot\|_\theta)$ into a Banach space and Lemma 7.6.2 characterises the norm on $C_\theta(E)$ by

$$\|f\|_\theta := \inf_{g\in\mathcal{I}} \|f-g\| = \sup_{\mu\in\mathcal{P}_\theta(E)} |\langle f,\mu\rangle|.$$

By construction, the continuous dual space of $(C_\theta(E), \|\cdot\|_\theta)$ equals $\mathcal{P}_\theta(E)$.

7.3 MAIN RESULTS

7.3.1 *The large deviation principle and some consequences*

Denote by $r_t(\sigma, \mathrm{d}\zeta)$ the kernel of the Markov process generated by $S(t)$, i.e. $S(t)f(\sigma) = \int f(\eta) r(\sigma, \mathrm{d}\eta)$. We introduce a notion of asymptotically decoupledness for Markov processes. This notion will allow us to obtain the large deviation principle at later times.

Assumption 7.3.1. The Markov process $\{\sigma(t)\}_{t\geq 0}$ is called uniformly asymptotically decoupled(UAD) if for every time t there exist two sequences $\{c_t(n)\}_{n\geq 0}$ and $\{d_t(n)\}_{n\geq 0}$ satisfying (7.1.1) such that for every measure $\mu \in \mathcal{P}(E^k)$ that is asymptotically decoupled with sequences $\{c_\mu(n)\}_{n\geq 0}$ and $\{d_\mu(n)\}_{n\geq 0}$, the measure

$$\mu \otimes S(t)(\mathrm{d}\sigma_1, \ldots, \mathrm{d}\sigma_k, \mathrm{d}\zeta) := \int \mu(\mathrm{d}\sigma_1, \ldots, \mathrm{d}\sigma_k) r_t(\sigma_k, \mathrm{d}\zeta)$$

is asymptotically decoupled with sequences $\{c_\mu(n) + c_t(n)\}_{n\geq 0}$ and $\{d_\mu(n) + d_t(n)\}_{n\geq 0}$.

Clearly, the assumption above is satisfied for product dynamics as in Example 7.2.2. By cluster expansion methods this assumption can be verified for interacting systems as in Le Ny and Redig [2004]. In that paper, translation invariant nearest-neighbour spin-flip dynamics as in Example 7.2.1 are considered. Starting with a measure $\mu \in \mathcal{P}_\theta(E)$ that is asymptotically decoupled, it is shown that the law of the process at a later time is also AD. The same proof, without any changes, also works for any starting law that is AD. Additionally the same proof, with minor changes, gives the AD property law of the process at two or, finitely many, times.

The proof of the result in Le Ny and Redig [2004] relies on the property that the rate of any spin flip is bounded away from 0. Thus, it is unclear whether such an result can be proven for a translation invariant exclusion process.

Define

$$V_n(t)f = \frac{1}{|\Lambda_n|} \log S(t) e^{\sum_{i\in\Lambda_n} f\circ\theta_i}. \tag{7.3.1}$$

These $V_n(t)f$ have the interpretation of conditional log-Laplace transforms. This semigroup can be rewritten in terms of the semigroup in Chapter 6. Denote by

$$\begin{aligned} \mathbb{V}(t)f &= \log S(t)e^f & & f \in C(E) \\ \mathbb{H}f &= e^{-f}Ae^f & & f : e^f \in \mathcal{D}(A), \end{aligned}$$

the semigroup and Hamiltonian that were used in Chapter 6.

Adapting the approach of Pfister [2002], we use our conditional AD property to show that as $n \to \infty$ the conditional Log-Laplace transforms converge. By a projective limit theorem argument, we obtain the large deviation result on $\prod_{t\in\mathbb{R}} \mathcal{P}(E)$. A stochastic Lyapunov technique based on the generators $H_n f = \frac{1}{|\Lambda_n|}\mathbb{H}(\sum_{i\in\Lambda_n} f\circ\theta_i)$ of the semigroups $V_n(t)$ is used to prove exponential tightness of L_n in $D_{\mathcal{P}_\theta(E)}(\mathbb{R}^+)$. We start with this result first.

Lemma 7.3.2. *Let D be a core for $(A, \mathcal{D}(A))$. Suppose that for every $\lambda \in \mathbb{R}$ and $f \in D$, we have $e^{\lambda|\Lambda_n|f_n} \in \mathcal{D}(A)$, where $f_n = \frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n} f\circ\theta_i$. Furthermore, suppose that*

$$C(\lambda, f) := \sup_n \|H_n(\lambda f)\| < \infty$$

Then L_n is exponentially tight in $D_{\mathcal{P}(E)}(\mathbb{R}^+)$.

This result is based on Theorem 4.4 in Feng and Kurtz [2006] of which the conditions are verified below. Exponential tightness, together with convergence of the conditional log-Laplace transforms yields the large deviation principle with a representation of the rate function as in Theorem 2.4.10.

Theorem 7.3.3. *Suppose that we have a translation invariant Markov process with the UAD property.*

Then we have that for every $f \in C(E)$, the sequence $V_n(t)f$ defined in (7.3.1) *has a limit $V(t)f$ in C_θ, uniformly for t in compact intervals.*

Additionally, suppose that the initial law of the Markov process is $\mathbb{P}_0$ is asymptotically decoupled and that the conditions for Lemma 7.3.2 are satisfied. Then the large deviation principle holds for $\{L_n\}_{n\geq 0}$ with normalization $|\Lambda_n|$ on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ and the rate function is given by

$$I(\gamma) = \sup_{\{t_i\}} s(\gamma(0) \,|\, \mathbb{P}_0) + \sum_{i=1}^{k} I_{t_i - t_{i-1}}(\gamma(t_i) \,|\, \gamma(t_{i-1}),$$

where $\{t_i\}$ runs over collections of ordered times $0 = t_0 < t_1 < \cdots < t_k$ and where

$$I_t(\nu \,|\, \mu) = \sup_f \left\{ \langle f, \nu \rangle - \langle V(t) f, \mu \rangle \right\}.$$

As a corollary, we obtain the semigroup property for $\{V(t)\}_{t\geq 0}$.

Corollary 7.3.4. *The collection of operators $\{V(t)\}_{t\geq 0}$ forms a semigroup on C_θ, i.e. $V(s)V(t) = V(s+t)$ and $V(0) = \mathbb{1}$.*

Note that both structural conditions of Theorem 7.3.3 can be checked in the example of invariant spin-flip dynamics, see Example 7.2.1, where $E = \{-1, 1\}^{\mathbb{Z}^d}$ and where for $f \in D$

$$Af(\sigma) = \sum_{i \in \mathbb{Z}^d} c_i(\sigma) \left[f(\sigma^i) - f(\sigma) \right]$$

and

(a) Translation invariance: for all $i \in \mathbb{Z}^d$, it holds that $c_i(\sigma) = c_0(\theta_{-i}\sigma)$.

In this setting, the verification of the conditions for Lemma 7.3.2 are straightforward. The UAD property can be verified via cluster expansion methods, following Le Ny and Redig [2004], under the additional conditions that

(b) Nearest neighbour interaction: c_0 only depends on $D_{c_0} = \{j \in \mathbb{Z}^d \mid |j| \leq 1\}$.

(c) Strict positivity:

$$0 < \min_\sigma c_0(\sigma) < \max_\sigma c_0(\sigma) < \infty.$$

7.4 PROOFS OF THE RESULTS IN SECTION 7.3.1

We start with the proof of Lemma 7.3.2.

Proof of Lemma 7.3.2. To simplify notation, set $f_n = \frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n} f \circ \theta_i$. Theorem 4.4 in Feng and Kurtz [2006], shows, using the compactness of $\mathcal{P}(E)$, that the sequence L_n is exponentially tight in $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ if for every $f \in D$, D closed under addition and separating points in $\mathcal{P}(E)$, the sequence of trajectories $t \mapsto \langle f, L_n(t)\rangle = f_n(\sigma(t))$ is exponentially tight in $D_{\mathbb{R}}(\mathbb{R}^+)$.

To do this, we use Theorem 4.1, (b) to (a), combined with Remark 4.2 in Feng and Kurtz [2006]. We use the notation of Feng and Kurtz [2006]. On $\mathbb{R}$ the metric r is simply the Euclidean distance, and as we are considering trajectories in $[-\|f\|, \|f\|] \subseteq \mathbb{R}$, it is not necessary to replace the metric r by $q = r \wedge 1$.

For every $T > 0$, pick $\beta = 1$ and let $\gamma_n(\delta, \lambda, T) = |\Lambda_n|\delta C(\lambda, f)$. $\gamma_n(\delta, \lambda, T)$ satisfies the condition in equation (4.2) of Feng and Kurtz [2006]. Because the condition in equation (4.3) follows from equation (4.6), we prove the latter.

$e^{\lambda|\Lambda_n|f_n} \in \mathcal{D}(A)$ which implies by Lemma 4.3.2 in Ethier and Kurtz [1986] that

$$\exp\left\{|\Lambda_n|\lambda f_n(\sigma(t)) - \int_0^t |\Lambda_n| H_n(\lambda f)(\sigma(s))\mathrm{d}s\right\}$$

is a martingale. This implies that

$$\begin{aligned}
&\mathbb{E}\left[e^{|\Lambda_n|\lambda(f_n(\sigma(t+u))-f_n(\sigma(t)))} \,\middle|\, \mathcal{F}_t\right] \\
&\quad\leq e^{u|\Lambda_n|\|H_n(\lambda f)\|} \\
&\qquad\times \mathbb{E}\left[e^{|\Lambda_n|\lambda\left(f_n(\sigma(t+u))-f_n(\sigma(t))-\int_t^{t+u}|\Lambda_n|H_n(\lambda f)(\sigma(s))\mathrm{d}s\right)} \,\middle|\, \mathcal{F}_t\right] \\
&\quad\leq e^{\gamma_n(\delta,\lambda,T)},
\end{aligned}$$

which proves equation (4.6) in Feng and Kurtz [2006]. □

7.4.1 *The large deviation principle and decomposition of the relative entropy density*

Provided that the starting distribution is AD, we have the AD property for the finite dimensional distributions of the Markov process $\{\sigma(t)\}_{t\geq 0}$ in $\mathcal{P}(E^k)$ for any k by Assumption 7.3.1. This yields by Theorem 7.1.4 in Pfister [2002] that we have the large deviation principle for the sequence

$$n \mapsto L_n(0, t_1, \ldots, t_k) := \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} \delta_{\theta_i \sigma(0)} \times \delta_{\theta_i \sigma(t_1)} \times \cdots \times \delta_{\theta_i \sigma(t_k)}$$

in $\mathcal{P}(E^{k+1})$ for some sequence of times $t_0 = 0 < t_1 < \cdots < t_k$. Let $\mu_{0,t_1,\ldots,t_k} \in \mathcal{P}(E^{k+1})$ be the distribution of the process $t \mapsto \sigma(t)$ restricted to the times $t_0, \ldots, t_k$. The rate function of this large deviation principle is given by the relative entropy density

$$s(\nu_{0,t_1,\ldots,t_k} \,|\, \mu_{0,t_1,\ldots,t_k}) = \lim_n \frac{1}{\Lambda_n} S_n(\nu_{0,t_1,\ldots,t_k} \,|\, \mu_{0,t_1,\ldots,t_k}).$$

We will decompose this relative entropy density in $k+1$ terms, one for every time component. We start with $k = 1$. Fix some time t_1 and some measure $\nu_{0,t_1} \in \mathcal{P}(E^2)$. Denote by ν_0 and μ_0 the time 0 marginals of ν_{0,t_1} and μ_{0,t_1}. Define

$$\tilde{\nu}_{0,t_1}(\mathrm{d}\hat{\sigma}, \mathrm{d}\zeta) = \int \delta_\sigma \otimes S'(t_1)\delta_\sigma(\mathrm{d}\hat{\sigma}, \mathrm{d}\zeta)\, \nu_0(\mathrm{d}\sigma) \quad \in \mathcal{P}(E^2)$$

the measure ν_0 composed with the Markovian evolution $S'(t_1)$. In other words, $\tilde{\nu}_{0,t_1}$ is the measure of which the first marginal coincides with ν_0, and which has regular conditional probabilities given the first coordinate that coincide with those of the Markov process. The application of Lemma 7.6.1, applied for $\mathcal{F}$ the σ-algebra generated by all variables for the first time coordinate, yields for an arbitrary $n \in \mathbb{N}$ that

$$\begin{aligned} S_n(\nu_{0,t_1} \,|\, \mu_{0,t_1}) &= S_n(\nu_0 \,|\, \mu_0) + \int_E S_n(\nu^{\mathcal{F}}(\sigma, \cdot) \,|\, r_\sigma(\cdot))\nu_0(\mathrm{d}\sigma) \\ S_n(\nu_{0,t_1} \,|\, \tilde{\nu}_{0,t_1}) &= S_n(\nu_0 \,|\, \nu_0) + \int_E S_n(\nu^{\mathcal{F}}(\sigma, \cdot) \,|\, r_\sigma(\cdot))\nu_0(\mathrm{d}\sigma) \\ &= \int_E S_n(\nu^{\mathcal{F}}(\sigma, \cdot) \,|\, r_\sigma(\cdot))\nu_0(\mathrm{d}\sigma), \end{aligned}$$

where $\nu^{\mathcal{F}}(\sigma, \cdot)$ is the regular conditional probability of ν given $\mathcal{F}$. Note that because of the Markov property, this regular conditional probability equals $r(\sigma, \cdot)$ and only depends on the first coordinate.

Combining the two statements yields

$$S_n(\nu_{0,t_1} \,|\, \mu_{0,t_1}) = S_n(\nu_0 \,|\, \mu_0) + S_n(\nu_{0,t_1} \,|\, \tilde{\nu}_{0,t_1}).$$

When divided by $|\Lambda_n|$, the first two terms converge as $n \to \infty$ by the AD property of μ_0 and μ_{0,t_1}, which implies that

$$s(\nu_{0,t_1} \,|\, \mu_{0,t_1}) = s(\nu_0 \,|\, \mu_0) + s(\nu_{0,t_1} \,|\, \tilde{\nu}_{0,t_1}). \tag{7.4.1}$$

We will iterate this procedure in the next lemma. Fix some k and times $t_0 = 0 < t_1, \cdots < t_k$ and $\nu_{0,t_1,\ldots,t_k} \in \mathcal{P}(E^{k+1})$. Denote by $\nu_{0,t_1,\ldots,t_{p-1}}$ the restriction of $\nu_{0,t_1,\ldots,t_k}$ to the first p coordinates. As above, we compose the measures $\nu_{0,t_1,\ldots,t_{p-1}}$ with Markovian evolutions of time $t_p - t_{p-1}$:

$$\begin{aligned}
&\tilde{\nu}_{0,t_1,\ldots,t_p}(\mathrm{d}\hat{\sigma}_0, \ldots, \mathrm{d}\hat{\sigma}_{p-1}, \mathrm{d}\zeta) \\
&= \int \delta_{\sigma_0,\ldots,\sigma_{p-1}} \times S'(t_p - t_{p-1})\delta_{\sigma_{p-1}}(\mathrm{d}\hat{\sigma}_0, \ldots, \mathrm{d}\hat{\sigma}_{p-1}, \mathrm{d}\zeta) \\
&\qquad\qquad\qquad \times \nu_{0,t_1,\ldots,t_{p-1}}(\mathrm{d}\sigma_0, \ldots, \mathrm{d}\sigma_{p-1}).
\end{aligned}$$

Note that $\tilde{\nu}_{0,t_1,\ldots,t_p} \in \mathcal{P}(E^{p+1})$.

Lemma 7.4.1. *We have*

$$s(\nu_{0,t_1,\ldots,t_k} \,|\, \mu_{0,t_1,\ldots,t_k}) = s(\nu_0 \,|\, \mu_0) + \sum_{p=1}^{k} s(\nu_{0,t_1,\ldots,t_p} \,|\, \tilde{\nu}_{0,t_1,\ldots,t_p}).$$

Proof. We reconsider the argument that led to (7.4.1). We have $k+1$ marginals now, and we start by decomposing the times into $\{0\}$ and $\{t_1, \ldots, t_k\}$, this yields

$$s(\nu_{0,t_1,\ldots,t_k} \,|\, \mu_{0,t_1,\ldots,t_k}) = s(\nu_0 \,|\, \mu_0) + s(\nu_{0,t_1,\ldots,t_k} \,|\, \overline{\nu}_0),$$

where $\overline{\nu}_0$ is the measure ν_0 composed with the Markovian evolution for the remaining k coordinates. The same step can be repeated for relative entropy density on the right, now decomposing the time marginals into $\{0, t_1\}$ and $\{t_2, \ldots, t_k\}$ and so on.

Let $\overline{\nu}_{0,t_1,\dots,t_p} \in \mathcal{P}(E^{p+1})$ be the measure $\nu_{0,t_1,\dots,t_p}$ composed with the Markovian evolution for the remaining coordinates. This yields inductively that

$$
\begin{aligned}
&s(\nu_{0,t_1,\dots,t_k} \mid \mu_{0,t_1,\dots,t_k}) \\
&\quad = s(\nu_0 \mid \mu_0) + s(\nu_{0,t_1,\dots,t_k} \mid \overline{\nu}_0) \\
&\quad = s(\nu_0 \mid \mu_0) + h(\nu_{0,t_1} \mid \tilde{\nu}_{0,t_1}) + s(\nu_{0,t_1,\dots,t_k} \mid \overline{\nu}_{0,t_1}) \\
&\qquad\quad \vdots \\
&\quad = s(\nu_0 \mid \mu_0) + \sum_{p=1}^{k} s(\nu_{0,t_1,\dots,t_p} \mid \tilde{\nu}_{0,t_1,\dots,t_p}).
\end{aligned}
$$

Note that we have used $\tilde{\nu}_{0,t_1,\dots,t_k} = \overline{\nu}_{0,t_1,\dots,t_k-1}$ for the last equality. □

By the contraction principle, Theorem 2.4.6, we obtain the large deviations behaviour of sequences $(L_n(0), L_n(t_1), \dots, L_n(t_k)) \in \mathcal{P}(E)^{k+1}$, where $L_n(t) = \frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n} \delta_{\theta_i\sigma(t)}$, which has the rate function $I_{0,t_1,\dots,t_k}$ defined for $(\gamma_0, \dots, \gamma_{t_k}) \in \mathcal{P}(E)^{k+1}$ by

$$
I_{0,t_1,\dots,t_k}(\gamma_0, \dots, \gamma_{t_k}) = \inf_{\substack{\nu\in\mathcal{P}(E^{k+1}) \\ \forall\, p\in\{0,\dots,k\} : \nu_p = \gamma_p}} h(\nu_{0,t_1,\dots,t_k} \mid \mu_{0,t_1,\dots,t_k})
$$

where ν_p is the restriction of ν to the p-th coordinate.

To study the contracted rate function, we turn our attention to Section 7.4.2 and the approximating conditional pressures $V_n(t)f$ defined by

$$
V_n(t)f(\sigma) = \frac{1}{|\Lambda_n|} \log \mathbb{E}_\sigma \left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i(\sigma(t))} \right].
$$

The convergence of this sequence in C_θ follows by abstract arguments developed in Section 7.4.2 below.

If we set $\Upsilon(\sigma)(\mathrm{d}\sigma_1, \mathrm{d}\sigma_2) = \delta_\sigma \otimes S'(t)\delta_\sigma(\mathrm{d}\sigma_1, \mathrm{d}\sigma_2)$, then by Assumption 7.3.1 and the translation invariance of the process, Condition 7.4.4 is satisfied and the convergence of $V_n(t)f$ to $V(t)f$ in C_θ follows by Theorem 7.4.10.

The next lemma will show that the sum in Lemma 7.4.1 simplifies.

Lemma 7.4.2. *The set of empirical measures* $(L_n(0), L_n(t_1), \dots, L_n(t_k))$ *satisfies the large deviation principle on* $\mathcal{P}(E)^{k+1}$ *with rate function*

$$I_{0,t_1,\dots,t_k}(\gamma_0,\dots,\gamma_{t_k}) = \begin{cases} s(\gamma_0 \mid \mu_0) + \sum_{p=1}^k \mathcal{I}_{t_p - t_{p-1}}(\gamma_{t_p} \mid \gamma_{t_{p-1}}) & \\ \qquad\qquad \text{if } \gamma_0,\dots,\gamma_{t_k} \in \mathcal{P}_\theta(E) \\ \infty \qquad\qquad \text{otherwise} \end{cases}$$

where

$$\mathcal{I}_t(\gamma_1 \mid \gamma_0) = \sup_{f \in C_\theta} \langle f, \gamma_1 \rangle - \langle V(t) f, \gamma_0 \rangle.$$

Proof. Lemma 7.4.1 gives

$$I_{0,t_1,\dots,t_k}(\gamma_0,\dots,\gamma_k) = \inf_{\substack{\nu \in \mathcal{P}(E^{k+1}) \\ \forall\, p \in \{0,\dots,k\} \,:\, \nu_{t_p} = \gamma_p}} s(\nu_0 \mid \mu_0) + \sum_{p=1}^k s(\nu_{0,t_1,\dots,t_p} \mid \tilde{\nu}_{0,t_1,\dots,t_p}).$$

Taking apart the last term gives

$$\begin{aligned} I_{0,t_1,\dots,t_k}(\gamma_0,\dots,\gamma_k) = \inf_{\substack{\nu' \in \mathcal{P}(E^k) \\ \forall\, p \in \{0,\dots,k-1\} \,:\, \nu_{t_p} = \gamma_p}} \Bigg\{ & s(\nu'_0 \mid \mu_0) + \sum_{p=1}^{k-1} s(\nu'_{0,t_1,\dots,t_p} \mid \tilde{\nu}'_{0,t_1,\dots,t_p}) \\ & + \inf_{\substack{\nu \in \mathcal{P}(E^{k+1}) \\ \nu_{0,\dots,t_{k-1}} = \nu',\, \nu_{t_k} = \gamma_k}} s(\nu \mid \tilde{\nu}_{0,\dots,t_k}) \Bigg\} \end{aligned}$$

If we apply Lemma 7.4.13, with $\mathcal{X} = E^k, \mathcal{Y} = E$, $\mu = \tilde{\nu}_{0,\dots,t_k}$ and $\Upsilon(\sigma_0,\dots,\sigma_{k-1}) = r_{\sigma_{k-1}}$, we see that this last term equals

$$\sup_{f \in C(E)} \langle f, \gamma_k \rangle - \langle p(f), v' \rangle. \tag{7.4.2}$$

Now note that as $f \in C(E)$, we have $p(f) = V(t_k - t_{k-1}) f$, and that $V(t_k - t_{k-1}) f$ is an equivalence class of functions on the t_{k-1} marginal. Hence, (7.4.2) equals

$$\sup_{f \in C(E)} \langle f, \gamma_k \rangle - \langle V(t_k - t_{k-1}) f, \gamma_{k-1} \rangle.$$

Repeating this step inductively yields the final result. ☐

Lemma 7.4.2 is the final ingredient for the proof of Theorem 7.3.3, which is now straightforward.

Proof of Theorem 7.3.3. The large deviation statement follows from Lemma's 7.3.2 and 7.4.2 and Theorem 4.28 in Feng and Kurtz [2006]. □

For the proof of Corollary 7.3.4, we first show that $\mathcal{P}_\theta(E)$ has a well behaved dense subset.

Lemma 7.4.3. *The translation invariant ergodic AD measures are weakly dense in $\mathcal{P}_\theta(E)$.*

This lemma can be proven as in Lemma 6.9 in Rassoul-Agha and Seppäläinen [2015], where it is shown that the ergodic translation invariant measures are dense in $\mathcal{P}_\theta(E)$. Lemma 7.4.3 follows by observing that the approximating measures in the proof of Lemma 6.9 in Rassoul-Agha and Seppäläinen [2015] are AD.

Proof of Corollary 7.3.4. As a direct consequence of Theorem 7.3.3, the contraction principle yields for any $s, t \geq 0$ and measures $\mu, \nu \in \mathcal{P}(E)$ that

$$I_{t+s}(\nu \mid \mu) = \inf_{\lambda \in \mathcal{P}_\theta(E)} \{I_t(\nu \mid \lambda) + I_s(\lambda \mid \mu)\}. \tag{7.4.3}$$

Let $f \in C_\theta(E)$. We first prove that $V(t)V(s)f(\mu) = V(s+t)f(\mu)$ for an asymptotically decoupled measure μ. By Hölders inequality $V_n(t)f$ is convex in f, which implies that also $V(t)f$ is convex in f. We also know that $f \mapsto V(t)f$ is continuous in $C_\theta(E)$ by Theorem 7.4.10. This implies that the double Legendre-Fenchel transform of $V(t)f$ is $V(t)f$, in other words:

$$\langle V(t)f, \mu\rangle = \sup_{\nu \in \mathcal{P}_\theta(E)} \{\langle f, \nu\rangle - I_t(\nu|\mu)\}. \tag{7.4.4}$$

Therefore, we have, using equation (7.4.3), that

$$\begin{aligned} V(t+s)f(\mu) &= \sup_{\nu \in \mathcal{P}_\theta(E)} \sup_{\lambda \in \mathcal{P}_\theta(E)} \{\langle f, \nu\rangle - I_s(\nu \mid \lambda) - I_t(\lambda \mid \mu)\} \\ &= \sup_{\lambda \in \mathcal{P}_\theta(E)} \{\langle V(s)f, \lambda\rangle - I_t(\lambda \mid \mu)\} \\ &= \langle V(t)V(s)f, \mu\rangle \end{aligned}$$

In other words, we have

$$V(t+s)f(\mu) = V(t)(V(s)f)(\mu) \tag{7.4.5}$$

for all asymptotically decoupled μ. Because the AD measures are dense in $\mathcal{P}_\theta(E)$ by Lemma 7.4.3, we obtain that $V(t)V(s)f = V(t+s)f$. □

7.4.2 *Existence of the conditional pressure density and the relation to the relative entropy density*

In Section 7.4.1, we used that the limit of

$$V_n(t)f(\sigma) = \frac{1}{|\Lambda_n|} \log \mathbb{E}_\sigma \left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i(\sigma(t))} \right]$$

exists as $n \to \infty$. A priori it is not clear that for a given σ this limit exists, even if the measure $S(t)'\delta_\sigma$ is AD. This is the case because $S(t)'\delta_\sigma$ is in general not translation invariant, which implies that the standard argument based on the work by Pfister [2002] does not apply.

However, if we consider the sequence $V_n(t)f$ in the quotient space C_θ, we are able to the standard argument in an adapted way.

This section can be read independently of the other sections as we will consider the conditional pressure density and conditional relative entropy density on arbitrary spaces with a product structure.

Let X, Y be two Polish spaces and define the product spaces $\mathcal{X} = X^{\mathbb{Z}^d}, \mathcal{Y} = Y^{\mathbb{Z}^d}$. Suppose we are given a measurable map $\Upsilon : \mathcal{X} \to \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ that satisfies the following condition.

Condition 7.4.4. The map Υ has the following two properties.

(a) $\eta \mapsto \Upsilon(\eta)$ is continuous for the topology on $\mathcal{X}$ to the weak topology on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

(b) For all $\eta \in \mathcal{X}$ and $i \in \mathbb{Z}^d$, we have $\Upsilon(\theta_i \eta) = \Upsilon(\eta) \circ \theta_i$. Additionally, we have that the projection of $\Upsilon(\eta)$ on $\mathcal{X}$ equals δ_η.

(c) $\Upsilon(\eta)$ is asymptotically decoupled with sequences $\{c(n)\}_{n\geq 0}$ and $\{d(n)\}_{n\geq 0}$ that do not depend on η.

For $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and $f \in C(\mathcal{X} \times \mathcal{Y})$, define

$$p_n(f \mid \mu) = \frac{1}{|\Lambda_n|} \log \mathbb{E}_\mu \left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i} \right].$$

Clearly, if Υ does not depend on $\eta \in \mathcal{X}$, then it is well known that $p_n(f) = p_n(f \mid \Upsilon(\eta))$ converges in some appropriate sense, see for example Pfister [2002]. Note that by (a) of the condition above, we have for $f \in C(\mathcal{X} \times \mathcal{Y})$ that the function $\eta \mapsto p_n(f \mid \Upsilon(\eta))$ is an element of $C(\mathcal{X})$.

We will show that under Condition 7.4.4 (b) and (c) that the sequence $p_n(f \mid \Upsilon(\cdot))$ converges in $C_\theta(\mathcal{X})$.

Remark 7.4.5. Note that if we replace Condition 7.4.4 (a) by merely assuming measurability of the map, Proposition 7.4.7 holds also if the space of continuous functions is replaced by the space of bounded and measurable functions.

We start with an auxiliary lemma that has a straightforward proof.

Lemma 7.4.6. *Let $(B, \|\cdot\|)$ be some Banach space. Let $\{x_n\}_{n\geq 1}$ be a sequence in B that satisfies $\|x_n - x_m\| \leq c_{n,m}$, where $\{c_{n,m}\}_{m,n\geq 1}$ satisfies $\lim_{m\to\infty} \lim_{n\to\infty} c_{n,m} = 0$. Then, the sequence x_n converges.*

Proposition 7.4.7. *Let Υ satisfy Condition 7.4.4 and let $f \in C(\mathcal{X} \times \mathcal{Y})$. Then the sequence $n \mapsto p_n(f \mid \Upsilon(\cdot))$ converges in $C_\theta(\mathcal{X})$.*

Proof. Let f be $\mathcal{F}_r$ measurable. Given some fixed m and large n, we introduce a decomposition of Λ_n into smaller boxes of size Λ_m and corridors between the translates of Λ_m to exploit the asymptotically decoupledness of $\Upsilon(\eta)$. We adapt the approach as in Pfister [2002], which is the canonical way to prove show the existence of the pressure.

Define $r' = r'(m) := \lceil \frac{d(m+r)}{2} \rceil$. For $n > m + r + r'$, there exists a unique maximal $k = k(n, m)$ such that

$$2n + 1 = k \left[2(m + r + r') + 1\right] + j, \tag{7.4.6}$$

where $0 \leq j < 2(m+r+r')+1$. We split up the box Λ_n into k^d translates of the box $\Lambda_{m+r+r'}$, and the complement of these boxes in Λ_n. Now, we split up the boxes of size $\Lambda_{m+r+r'}$ into a smaller box located exactly in the center of the box of size $\Lambda_{m+r+r'}$, which is a shift of Λ_m, and a boundary of width $r + r'$. These small boxes are denoted with $\Lambda^1, \ldots, \Lambda^{k^d}$, their centers by $x_1, \ldots, x_{k^d}$ and the complement of these boxes in Λ_n is denoted by Λ^{k^d+1}.

This decomposition yields

$$
\begin{aligned}
&p_n(f \mid \Upsilon(\eta)) \\
&= \frac{1}{|\Lambda_n|} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i} \right] \\
&\leq \frac{1}{|\Lambda_n|} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{q=1}^{k^d} \sum_{i\in\Lambda^q} f\circ\theta_i + |\Lambda^{k^d+1}| \|f\|} \right] \\
&= \frac{1}{|\Lambda_n|} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{q=1}^{k^d} \sum_{i\in\Lambda^q} f\circ\theta_i} \right] + \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \\
&\leq \frac{1}{|\Lambda_n|} \log \left\{ e^{k^d c(m+r)} \prod_{q=1}^{k^d} \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{i\in\Lambda^q} f\circ\theta_i} \right] \right\} + \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \\
&= \frac{k^d c(m+r)}{|\Lambda_n|} + \frac{|\Lambda_m|}{|\Lambda_n|} \sum_{q=1}^{k^d} \frac{1}{|\Lambda_m|} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{i\in\Lambda^q} f\circ\theta_i} \right] \\
&\quad + \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \\
&= \frac{k^d c(m+r)}{|\Lambda_n|} + \frac{|\Lambda_m|}{|\Lambda_n|} \sum_{q=1}^{k^d} p_m(f \mid \Upsilon(\theta^q(\eta))) + \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \qquad (7.4.7)
\end{aligned}
$$

θ^q is the shift such that the box Λ^q is centred at the origin. Note that we have only used AD from above in the fourth line.

Using AD from below, we obtain that

$$
\begin{aligned}
&p_n(f \mid \Upsilon(\eta)) \geq \\
&\qquad - \frac{k^d c(m+r)}{|\Lambda_n|} + \frac{|\Lambda_m|}{|\Lambda_n|} \sum_{q=1}^{k^d} p_m(f \mid \Upsilon(\theta^q(\eta))) - \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| .
\end{aligned}
\qquad (7.4.8)
$$

Because $\frac{1}{k^d}\sum_{q=1}^{k^d} p_m(f \mid \Upsilon(\theta^q(\eta)))$ is bounded by $\|f\|$, we get as upper and lower bounds

$$\begin{aligned} p_n(f \mid \Upsilon(\eta)) &\le \frac{|\Lambda_m| k^d}{|\Lambda_n|} \frac{1}{k^d} \sum_{q=1}^{k^d} p_m(f \mid \Upsilon(\theta^q(\eta))) \\ &\quad + \frac{k^d c(m+r)}{|\Lambda_n|} + \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \\ p_n(f \mid \Upsilon(\eta)) &\ge \frac{|\Lambda_m| k^d}{|\Lambda_n|} \frac{1}{k^d} \sum_{q=1}^{k^d} p_m(f \mid \Upsilon(\theta^q(\eta))) \\ &\quad - \frac{k^d c(m+r)}{|\Lambda_n|} - \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| . \end{aligned}$$

Because Υ is translation invariant, integration with respect to a translation invariant measure simplifies the sum. Consequentially, we obtain

$$\begin{aligned} &\|p_n(f \mid \Upsilon(\cdot)) - p_m(f \mid \Upsilon(\cdot))\|_\theta \\ &\quad \le \left\| p_n(f \mid \Upsilon(\cdot)) - \frac{|\Lambda_m| k^d}{|\Lambda_n|} p_m(f \mid \Upsilon(\cdot)) \right\|_\theta \\ &\quad\quad + \left| \frac{|\Lambda_m| k^d}{|\Lambda_n|} - 1 \right| \|p_m(f \mid \Upsilon(\cdot))\|_\theta \\ &\quad \le \frac{k^d c(m+r)}{|\Lambda_n|} + 2\frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \\ &\quad \le \frac{c(m+r)}{|\Lambda_m|} + 2\frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| . \end{aligned} \tag{7.4.9}$$

By Lemma 7.4.6, we are done if we can show that the constants

$$c_{n,m}(f) = \frac{c(m+r)}{|\Lambda_m|} + 2\frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} \|f\| \tag{7.4.10}$$

satisfy $\lim_{m\to\infty} \lim_{n\to\infty} c_{n,m}(f) = 0$. By definition of asymptotically decoupledness, the first component of $c_{n,m}(f)$ converges to 0 as m goes to infinity. This means that we are left to prove that

$$\lim_m \lim_n \frac{|\Lambda^{k^d+1}|}{|\Lambda_n|} = 0. \tag{7.4.11}$$

This statement is equivalent to $\lim_m \lim_n \frac{k^d|\Lambda_m|}{|\Lambda_n|} = 1$, which in turn is equivalent to $\lim_m \lim_n \frac{k(2m+1)}{2n+1} = 1$. By equation (7.4.6), this yields the equivalence to showing

$$\lim_m \lim_n \frac{2k(r+r')+j}{2n+1} = 0.$$

Because j only depends on m, the term involving j converges taking n to infinity to 0. Finally, the definition of r' and k imply

$$\begin{aligned} &= \lim_{m\to\infty} \lim_{n\to\infty} \frac{2k(r+r')}{2n+1} = \lim_{m\to\infty} \lim_{n\to\infty} \frac{k(2r+d(m+r))}{2n+1} \\ &= \lim_{m\to\infty} \frac{2r+d(m+r)}{2m+1} = 0. \end{aligned}$$

□

Lemma 7.4.8. *Let Υ satisfy Condition 7.4.4. For every $t \geq 0$ and $f, g \in C(\mathcal{X} \times \mathcal{Y})$*

$$\|p_n(f \,|\, \Upsilon(\cdot)) - p_n(g \,|\, \Upsilon(\cdot))\| \leq \|f-g\|.$$

If f, g are local, we have

$$\|p(f) - p(g)\|_\theta \leq \|f-g\|_\theta .$$

Proof. Let $f, g \in C(\mathcal{X} \times \mathcal{Y})$. We examine the approximating sequences $p_n(f \,|\, \Upsilon(\eta))$ and $p_n(g \,|\, \Upsilon(\eta))$.

$$\begin{aligned} &\frac{1}{|\Lambda_n|} \log\left\{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i}\right]\right\} - \frac{1}{|\Lambda_n|} \log\left\{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i}\right]\right\} \\ &= \frac{1}{|\Lambda_n|} \log\left\{\frac{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i}\right]}{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i}\right]}\right\} \\ &= \frac{1}{|\Lambda_n|} \log\left\{\frac{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i} e^{\sum_{i\in\Lambda_n} (f-g)\circ\theta_i}\right]}{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i}\right]}\right\} \\ &= \frac{1}{|\Lambda_n|} \log\left\{\frac{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i} e^{|\Lambda_n|\langle f-g, L_n(\sigma)\rangle}\right]}{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i}\right]}\right\} \end{aligned}$$

where σ is a random variable distributed as $\Upsilon(\eta)$. In the last line one recognises a tilted measure. This means that

$$\left|\frac{1}{|\Lambda_n|}\log\left\{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} f\circ\theta_i}\right]\right\} - \frac{1}{|\Lambda_n|}\log\left\{\mathbb{E}_{\Upsilon(\eta)}\left[e^{\sum_{i\in\Lambda_n} g\circ\theta_i}\right]\right\}\right| \leq \sup_\sigma |\langle f-g, L_n(\sigma)\rangle|.$$

This proves the first statement. If f, g are local, the approach that was given in Lemma 7.6.2, we get that

$$\limsup_{n\to\infty}\sup_\eta |p_n(f\,|\,\Upsilon(\eta)) - p_n(g\,|\,\Upsilon(\eta))| \leq \limsup_{n\to\infty}\sup_\sigma |\langle f-g, L_n(\sigma)\rangle| = |\langle f-g,\nu\rangle| \leq \|f-g\|_\theta\,,$$

where ν is a specific weak limit point of a converging subsequence of $L_n(\sigma_n)$ and where σ_n is the configuration that maximises $\sup_\sigma |\langle f - g, L_n(\sigma)\rangle|$. In other words,

$$\lim_{n\to\infty}\|p_n(f\,|\,\Upsilon(\cdot)) - p_n(g\,|\,\Upsilon(\cdot))\|_\theta \leq \limsup_{n\to\infty}\|p_n(f\,|\,\Upsilon(\cdot)) - p_n(g\,|\,\Upsilon(\cdot))\| \leq \|f-g\|_\theta\,.$$

Taking limits, we see $\|p(f) - p(g)\|_\theta \leq \|f-g\|_\theta$. □

The results of the lemma show that $f \mapsto p(f)$ can be considered as a continuous and contractive map from the image of the local functions in $C_\theta(\mathcal{X}\times\mathcal{Y})$ to $C_\theta(\mathcal{X})$. This implies that $f \mapsto p(f)$ can be extended as a continuous and contractive map from $C_\theta(\mathcal{X}\times\mathcal{Y})$ to $C_\theta(\mathcal{X})$.

Definition 7.4.9. Define $f \mapsto p(f)$ for $f \in C_\theta(\mathcal{X}\times\mathcal{Y})$ by the continuous extension of $p(f)$ for local functions f.

Theorem 7.4.10. *Let Υ satisfy Condition 7.4.4. The map $f \mapsto p(f)$ is contractive if considered as a map from $C_\theta(\mathcal{X}\times\mathcal{Y})$ to $C_\theta(\mathcal{X})$, i.e. for $f, g \in C_\theta(\mathcal{X}\times\mathcal{Y})$, we have*

$$\|p(f) - p(g)\|_\theta \leq \|f-g\|_\theta\,.$$

Also, for every $f \in C(\mathcal{X}\times\mathcal{Y})$, we have

$$\|p_n(f\,|\,\Upsilon(\cdot)) - p(f)\|_\theta \to 0.$$

Proof. The contractivity property follows directly from the definition and Lemma 7.4.8. We prove the second statement. Let $f \in C_\theta(\mathcal{X} \times \mathcal{Y})$, represented by some function in $C(\mathcal{X} \times \mathcal{Y})$ that we will also denote by f and let f_r be local functions that approximate the representant of f in norm. We obtain

$$\begin{aligned}
&\|p(f) - p_n(f \,|\, \Upsilon(\cdot))\|_\theta \\
&\leq \|p(f) - p(f_r)\|_\theta + \|p(f_r) - p_n(f_r \,|\, \Upsilon(\cdot))\|_\theta \\
&\qquad\qquad\qquad\qquad + \|p_n(f_r \,|\, \Upsilon(\cdot)) - p_n(f \,|\, \Upsilon(\cdot))\|_\theta \\
&\leq \|f - f_r\|_\theta + \|p(f_r) - p_n(f_r \,|\, \Upsilon(\cdot))\|_\theta \\
&\qquad\qquad\qquad\qquad + \|p_n(f_r \,|\, \Upsilon(\cdot)) - p_n(f \,|\, \Upsilon(\cdot))\| \\
&\leq \|f - f_r\| + \|p(f_r) - p_n(f_r \,|\, \Upsilon(\cdot))\|_\theta + \|f_r - f\| \,.
\end{aligned}$$

The second statement follows by sending first r and then n to infinity. □

Suppose that we have a translation invariant measure $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ that is asymptotically decoupled. Furthermore, suppose that the regular conditional probability of μ given the first coordinate is given by $\Upsilon(\cdot)$, where Υ satisfies Condition 7.4.4.

Pick some translation invariant measure $\nu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and denote by ν_0 and μ_0 the restrictions of ν and μ to the $\mathcal{X}$ component. Denote $\nu_0 \otimes \Upsilon(\mathrm{d}\eta_1, \mathrm{d}\eta_2) := \int \Upsilon(\mathrm{d}\zeta)(\mathrm{d}\eta_1, \mathrm{d}\eta_2)\nu_0(\mathrm{d}\zeta)$. The argument that led to (7.4.1) gives

$$s(\nu \,|\, \mu) = s(\nu_0 \,|\, \mu_0) + s(\nu \,|\, \nu_0 \otimes \Upsilon). \tag{7.4.12}$$

It is well known that the relative entropy is given by the Legendre transform of the pressure. Our next step is to show that the conditional relative entropy density is given by the Legendre transform of the conditional pressure density.

Proposition 7.4.11. *Let Υ satisfy Condition 7.4.4 and define $\tilde{\nu} = \nu_0 \otimes \Upsilon$. Then we have that*

$$s(\nu \,|\, \tilde{\nu}) = \sup_{f \in C_b(\mathcal{X} \times \mathcal{Y})} \left\{ \langle f, \nu \rangle - \langle p(f), \nu_0 \rangle \right\}.$$

The proof of this lemma is based on the following error bound.

Lemma 7.4.12. *Let Υ satisfy Condition 7.4.4. There exists sequences $\{b_1(m)\}_{m\geq 1}$ and $\{b_2(m)\}_{m\geq 1}$, $b_1(m)\uparrow 1$, $b_2(m)\downarrow 0$ such that for every $f\in C(\mathcal{X}\times\mathcal{Y})$ that satisfies $\mathcal{D}(f)\subseteq\Lambda_m$ for some m, we have*

$$\langle p(f),\mu\rangle \leq b_1(m)\frac{1}{|\Lambda_m|}\langle \log\langle e^{|\Lambda_m|f},\Upsilon(\cdot)\rangle,\mu\rangle + b_2(m)$$

for every $\mu\in\mathcal{P}_\theta(\mathcal{X})$.

Proof of Proposition 7.4.11. Let f be $\mathcal{F}_{\Lambda_m}$ measurable. Define the $\mathcal{F}_{\Lambda_{n+m}}$ measurable function $\bar f = \sum_{i\in\Lambda_n} f\circ\theta_i$. This yields by Lemma 2.19 Seppäläinen [1993] that

$$S_{n+m}(\nu\,|\,\tilde\nu) \geq \langle \bar f,\nu\rangle - \int \log\langle e^{\bar f},\Upsilon(\eta)\rangle \nu_0(\mathrm{d}\eta),$$

which implies

$$\frac{1}{|\Lambda_n|}S_{n+m}(\nu\,|\,\tilde\nu) \geq \langle f,\nu\rangle - \int p_n(f\,|\,\Upsilon(\eta))\nu_0(\mathrm{d}\eta).$$

Taking the limit n to infinity yields

$$s(\nu\,|\,\tilde\nu) \geq \sup_{f\text{ local}} \langle f,\nu\rangle - \langle p(f),\tilde\nu\rangle.$$

Because the local functions are dense in $C(\mathcal{X}\times\mathcal{Y})$, and p is continuous, we obtain that the supremum can actually be taken over all $f\in C(\mathcal{X}\times\mathcal{Y})$,

On the other hand, by Lemma 7.4.12, we see

$$\begin{aligned}\sup_g\langle g,\nu\rangle - \langle p(g),\tilde\nu\rangle &\geq \langle \frac{f}{|\Lambda_m|},\nu\rangle - \langle p\left(\frac{f}{|\Lambda_m|}\right),\tilde\nu\rangle \\ &\geq \frac{1}{|\Lambda_m|}\left\{\langle f,\nu\rangle - b_1(m)\int\log\langle e^f,\Upsilon(\eta)\rangle\nu_0(\mathrm{d}\eta)\right\} - b_2(m).\end{aligned}$$

By taking the supremum over all such $f\in\mathcal{F}_{\Lambda_m}$, Lemma 2.19 in Seppäläinen [1993] gives us

$$\begin{aligned}\sup_g\langle g,\nu\rangle - \langle p(g),\tilde\nu\rangle &\geq \frac{b_1(m)}{|\Lambda_m|}S_m\left(\frac{\nu}{b_1(m)}\,\middle|\,\tilde\nu\right) - b_2(m) \\ &= \frac{1}{|\Lambda_m|}S_m(\nu\,|\,\tilde\nu) - \frac{\log b_1(m)}{|\Lambda_m|} - b_2(m).\end{aligned}$$

Taking the limit m to infinity, we obtain

$$\sup_g\{\langle g,\nu\rangle - \langle p(g),\tilde\nu\rangle\} \geq s(\nu\,|\,\tilde\nu).$$

□

We proceed with the proof of our auxiliary lemma.

Proof of Lemma 7.4.12. Consider the sequence $n_k = k(m + \lceil \frac{d(m)}{2} \rceil)$. The box Λ_{n_k} can be split up into k^d boxes of size Λ_m and equally sized corridors of size $2\lceil \frac{d(m)}{2} \rceil$ in between them. Let $x_1, \dots, x_{k^d}$ be the centers of these boxes.

By Hölders inequality and the fact that $\Upsilon(\eta)$ is asymptotically decoupled from above to obtain, we obtain

$$\begin{aligned}
&p_{n_k}(f \mid \Upsilon(\eta)) \\
&= \frac{1}{|\Lambda_{n_k}|} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{\sum_{i \in \Lambda_m} \sum_{q=1}^{k^d} f \circ \theta_{x_q + i}} \right] \\
&\leq \frac{1}{|\Lambda_{n_k}|} \frac{1}{|\Lambda_m|} \sum_{i \in \Lambda_m} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{|\Lambda_m| \sum_{q=1}^{k^d} f \circ \theta_{x_q + i}} \right] \\
&\leq \frac{1}{|\Lambda_{n_k}|} \frac{1}{|\Lambda_m|} \sum_{i \in \Lambda_m} \sum_{q=1}^{k^d} \log \mathbb{E}_{\Upsilon(\eta)} \left[e^{|\Lambda_m| f \circ \theta_{x_q + i}} \right] \\
&\qquad + \frac{1}{|\Lambda_{n_k}|} \frac{1}{|\Lambda_m|} \sum_{i \in \Lambda_m} k^d c(m).
\end{aligned}$$

Using that $\Upsilon(\theta_i \eta) = \Upsilon(\eta) \circ \theta_i$, integration with respect to $\mu \in \mathcal{P}_\theta(\mathcal{X})$ yields

$$\begin{aligned}
\langle p_{n_k}(f \mid \Upsilon(\cdot)), \mu \rangle \leq \frac{|\Lambda_{n_k}| - |\Lambda^{k^d+1}|}{|\Lambda_{n_k}|} \frac{1}{|\Lambda_m|} \langle \log \mathbb{E}_{\Upsilon(\cdot)} \left[e^{|\Lambda_m| f} \right], \mu \rangle \\
+ \frac{c(m)}{(m + \lceil \frac{d(m)}{2} \rceil)^d}
\end{aligned}$$

which implies the result by taking n to infinity. □

Important for the large deviations question introduced in the introduction, is how the relative entropy density on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ behaves under the contraction to the product space $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. Consider $\gamma_0 \in \mathcal{P}_\theta(\mathcal{X})$ and $\gamma_1 \in \mathcal{P}_\theta(\mathcal{Y})$. Recall that ν_0 is the restriction of ν to the first coordinate. Let ν_1 be the restriction of ν to the second coordinate. Define the quantity

$$I(\gamma_0, \gamma_1) = \inf_{\substack{\nu \in \mathcal{P}_\theta(\mathcal{X} \times \mathcal{Y}) \\ \nu_0 = \gamma_0, \nu_1 = \gamma_1}} s(\nu \mid \mu).$$

Lemma 7.4.13. *Let $\mu \in \mathcal{P}_\theta(\mathcal{X} \times \mathcal{Y})$ be AD. Let Υ be the regular conditional probability of μ given the $\mathcal{X}$ coordinate. Suppose that Υ satisfies Condition 7.4.4.*

For $\gamma_0 \in \mathcal{P}_\theta(\mathcal{X})$ and $\gamma_1 \in \mathcal{P}_\theta(\mathcal{Y})$. Define $\tilde{\gamma}_{0,1}(\mathrm{d}\xi, \mathrm{d}\zeta) = \Upsilon(\eta)(\mathrm{d}\xi)\gamma_0(\mathrm{d}\eta)$. Then we have

$$I(\gamma_0, \gamma_1) = s(\gamma_0 \mid \mu_0) + \inf_{\substack{\nu \in \mathcal{P}_\theta(\mathcal{X}\times\mathcal{Y}) \\ \nu_0 = \gamma_0, \nu_1 = \gamma_1}} s(\nu \mid \tilde{\gamma}_{0,1}).$$

and for the right hand part, it holds that

$$I_{\mathcal{Y}\mid\mathcal{X}}(\gamma_1 \mid \gamma_0) := \inf_{\substack{\nu \in \mathcal{P}_\theta(\mathcal{X}\times\mathcal{Y}) \\ \nu_0 = \gamma_0, \nu_1 = \gamma_1}} s(\nu \mid \tilde{\gamma}_{0,1}) = \sup_{f \in C(\mathcal{Y})} \left\{ \langle f, \gamma_1 \rangle - \langle p(f), \gamma_0 \rangle \right\}.$$

Proof. The first statement is clear from equation (7.4.12). Note that $f \mapsto p(f)$ is continuous and convex on $C_\theta(\mathcal{X} \times \mathcal{Y})$, so the second statement follows as in the proof of the second statement of Lemma 2.19 in Seppäläinen [1993]. □

7.5 CONJECTURE: A VARIATIONAL EXPRESSION FOR THE RATE FUNCTION

We have seen that the non-linear semigroup often has a second representation as a variational semigroup. This representation is an important step to obtain a variational representation of the large deviation rate function. Even though there is no proof of this representation at the moment, in the setting of a one-dimensional nearest-neighbour spin flip model a conjecture on the form of this semigroup can be made, following the general structure seen in this thesis.

7.5.1 *The conjecture*

As in Chapter 6, we will need a number of conditions.

Condition 7.5.1. D is a core for $(A, \mathcal{D}(A))$ that satisfies

(a) D is an algebra, i.e. if $f, g \in D$ then $fg \in D$.

(b) If $f \in D$ and $\phi : \mathbb{R} \to \mathbb{R}$ a smooth function on the range of f, then $\phi \circ f \in D$.

By property (b), D can be used as a domain of definition of $\mathbb{H}f = e^{-f}Ae^{f}$, the generator of the semigroup $\{\mathbb{V}(t)\}_{t\geq 0}$. Similar to the connection between $\mathbb{V}(t)$ and $V(t)$, there is a clear connection between $\mathbb{H}$ and H. Recall that $H_n f = \frac{1}{|\Lambda_n|}\mathbb{H}\left(\sum_{i\in\Lambda_n} f\circ\theta_i\right)$. By definition, the finite volume approximations of the semigroup $\{V(t)\}_{t\geq 0}$ on $C(E)$ satisfy

$$V_n(t)f = \frac{1}{|\Lambda_n|}\mathbb{V}(t)\left(\sum_{i\in\Lambda_n} f\circ\theta_i\right),$$

which implies for $f \in D$ that

$$\lim_{t\to 0}\left\|\frac{V_n(t)f - \frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n} f\circ\theta_i}{t} - H_n f\right\| = 0.$$

In the space C_θ, this yields

$$\lim_{t\to 0}\left\|\frac{V_n(t)f - f}{t} - H_n f\right\|_\theta = 0.$$

Therefore, we expect that if $\{V(t)\}_{t\geq 0}$ is strongly continuous and $Hf := \lim_n H_n f$ exists in C_θ, this limit is the generator of the semigroup $\{V(t)\}_{t\geq 0}$. We start with a condition, which allows the identification of the limit $\lim_n H_n f$.

Condition 7.5.2. Let Condition 7.5.1 be satisfied. For every $f \in D$, the net, based on the finite subsets $\Lambda \subseteq \mathbb{Z}^d$ ordered by set-inclusion, defined by

$$\Lambda \mapsto H_\Lambda f := e^{-\sum_{i\in\Lambda} f\circ\theta_i} A_0 e^{\sum_{x\in\Lambda} f\circ\theta_i}$$

is a bounded Cauchy net in $C(E)$. We denote the limit by Hf and write formally

$$Hf = e^{-\sum_{i\in\mathbb{Z}^d} f\circ\theta_i} A_0 e^{\sum_{i\in\mathbb{Z}^d} f\circ\theta_i}.$$

Note that this condition implies the conditions for Lemma 7.3.2 which yields exponential tightness of the sequence L_n. Using the condition, we can prove the following proposition.

Proposition 7.5.3. *Let Condition 7.5.2 be satisfied. For $f \in D$, we have $\sup_n \|H_n f\| < \infty$ and $H_n f \to Hf$ in C_θ. Additionally, if $f, g \in D$ and $f = g$ in C_θ, then $Hf = Hg$.*

As in Proposition 6.4.10 in Chapter 6, the aim is to use the Hamiltonian of the semigroup to write down a variational semigroup that hopefully equals the semigroup $V(t)$. To write down this variation semigroup in terms of a Lagrangian, we need additional properties of D. However, as there is no rigorous proof yet, we restrict ourselves to what representation is to be expected.

As in Chapter 6, we assume that the topology τ_D is finer then the $\|\cdot\|$ topology restricted to D and that $H : (D, \tau_D) \to (C(E), \|\cdot\|)$ is continuous. Finally, we set $D_\theta := (D/\mathcal{I})$.

Definition 7.5.4. Define the *Lagrangian* $\mathcal{L} : \mathcal{P}_\theta(E) \times D_\theta^* \to \mathbb{R}^+$ by

$$\mathcal{L}(\mu, u) = \sup_{f \in D_\theta} \left\{ \langle f, u \rangle - \langle Hf, \mu \rangle \right\}.$$

Clearly, $\mathcal{L}$ is lower semi-continuous and convex. Additionally, Furthermore, it is straightforward to establish properties like Proposition 6.3.13 in Chapter 6, i.e. to find a set U such that $\mathcal{L} = \infty$ for directions outside U. Because $f \mapsto \langle Hf, \mu \rangle$ is convex and continuous, the double Legendre-Fenchel transform of $f \mapsto \langle Hf, \mu \rangle$ is coincides with $f \mapsto \langle Hf, \mu \rangle$. We state the result as a lemma.

Lemma 7.5.5. *For $\mu \in \mathcal{P}_\theta(E)$ and $f \in D$, we have*

$$\langle Hf, \mu \rangle = \sup_{u \in U} \left\{ \langle f, u \rangle - \mathcal{L}(\mu, u) \right\}.$$

We now introduce a set of paths in $C_{\mathcal{P}_\theta(E)}(\mathbb{R}^+)$ that are absolutely continuous in a suitable way.

Definition 7.5.6. Define $D_\theta - \mathcal{AC}^\theta$, or if there is no chance of confusion, $\mathcal{AC}^\theta$, the space of absolutely continuous paths in $C_{\mathcal{P}_\theta(E)}(\mathbb{R}^+)$. A path $\nu \in C_{\mathcal{P}_\theta(E)}(\mathbb{R}^+)$ is absolutely continuous if there exists a (D_θ^*, wk^*) measurable curve $s \mapsto u(s)$ in D_θ^* with the following properties:

(i) for every $f \in D_\theta$ and $t \geq 0$ $\int_0^t |\langle f, u(s) \rangle| \mathrm{d}s < \infty$,

(ii) for every $t \geq 0$, $\nu(t) - \nu(0) = \int_0^t u(s) \mathrm{d}s$ as a D_θ^* Gelfand integral.

We denote $\dot{\nu}(s) := u(s)$. Furthermore, we will denote $\mathcal{AC}^\theta_{\mu_0}$ for the space of absolutely continuous trajectories starting at μ_0, and $\mathcal{AC}^{\theta,T}$ for trajectories that are only considered up to time T. Similarly, we define $\mathcal{AC}^{\theta,T}_{\mu_0}$.

We have introduced sufficient notation to define the Nisio variational semigroup.

Definition 7.5.7. The Nisio semigroup $\mathbf{V}$ mapping upper semi-continuous functions on $\mathcal{P}_\theta(E)$ to upper semi-continuous functions on $\mathcal{P}_\theta(E)$ is defined by

$$\mathbf{V}(t)G(\mu) = \sup_{\nu \in \mathcal{AC}_\mu^\theta} \left\{ G(\nu(t)) - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s))\mathrm{d}s \right\}.$$

For a function $f \in C_\theta(E)$, we denote $[f]$ for the function in $C(\mathcal{P}_\theta(E))$, defined by $[f](\mu) = \langle f, \mu \rangle$. As in Chapter 6, we would like to show that

$$\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle.$$

This, however, is not proven yet. Given a proof of this statement, we would obtain the following result as in Chapter 6.

Conjecture 7.5.8. The rate function I from Theorem 7.3.3 can be rewritten as

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}^\theta, \\ \infty & \text{otherwise.} \end{cases}$$

7.5.2 *Motivation for the conjecture*

We give an argument similar to that in Chapter 6, based on two inequalities:

$$\mathbf{V}(t)[f](\mu) \leq \langle V(t)f, \mu \rangle, \tag{7.5.1}$$

$$\mathbf{V}(t)[f](\mu) \geq \langle V(t)f, \mu \rangle, \tag{7.5.2}$$

for all $f \in C(E)$ and $t \geq 0$. Denote

$$R_n(\lambda)f = \frac{1}{|\Lambda_n|} \log \mathbb{E}\left[e^{\sum_{i \in \Lambda_n} f \circ \theta_i(\sigma_\tau)} \right]$$

where τ is an exponential stopping time with expectation λ. The proof that $V_n(t)f$ converges to $V(t)f$ in C_θ also works to prove that $R_n(\lambda)f$ converges as $n \to \infty$. Denote this limit by $R(\lambda)f$. In particular, the argument in Le Ny and Redig [2004] to obtain the AD property for the law of a nearest-neighbour spin-flip Markov process at some later time can be adapted to obtain the AD property at some exponential random time.

For the one-dimensional spin flip context the results of Redig and Wang [2010] state that under the evolution of time, exponentially decaying translation invariant potentials get mapped into exponentially decaying translation invariant potentials. As translation invariant exponentially decaying potentials can be naturally mapped into D_θ, rewriting the definitions into what they mean in our context, this gives a class of functions D_{exp} in D_θ such that $V(t)D_{exp} \subseteq D_{exp}$.

The subspace D_{exp} comes naturally equipped with a collection of semi-norms that are stronger than the norm on D_θ. If additionally, the map $V(t)$ is continuous from D_{exp} to D_{exp} with respect to this norm, this yields regularity that we can use for approximation arguments below. In fact, we need a similar statement for the resolvent.

Conjecture 7.5.9. For a one-dimensional nearest-neighbour spin flip model, we have that

$$R_n(\lambda)f \to R(\lambda)f, \qquad V_n(t)f \to V(t)f$$

in D_{exp}.

Arguments in favour of (7.5.1). Because Lemma 6.4.9 also holds in this context: for $t \geq 0$, $f \in D$ and $\mu \in \mathcal{P}_\theta(E)$, we have

$$\lim_{n\to\infty} \mathbf{R}(n)^{\lfloor nt \rfloor}[f](\mu) = \mathbf{V}(t)[f](\mu),$$

the inequality in (7.5.1) follows if we can prove the following two claims:

(a) $(\mathbb{1} - \lambda H)R(\lambda)f \geq f$

(b) $R\left(\frac{t}{m}\right)^m f \to V(t)f$.

For (a), note that $(\mathbb{1} - \lambda H_n)R_n(\lambda)f \geq f$ can be proven for all f and all $n \geq 1$ as in Lemma 6.3.3. The Hamiltonians $H_n : D \to C(E)$ can be shown to be uniformly continuous in n. Thus, the inequality carries over to the limit by Conjecture 7.5.9 for all $f \in D_{exp}$.

For (b), we rewrite $R\left(\frac{t}{k}\right)^k f - V(t)f$ as

$$\begin{aligned} R\left(\frac{t}{k}\right)^k f - V(t)f = R\left(\frac{t}{k}\right)^k f - R_n\left(\frac{t}{k}\right)^k f \\ + \left(R_n\left(\frac{t}{k}\right)^k f - V_n(t)f\right) + \left(V_n(t)f - V(t)f\right). \end{aligned}$$

The first and third term of the right-hand side converge in C_θ to 0 as $n \to \infty$. Thus, we need to prove that the middle term converges uniformly in n as k goes to infinity.

Motivated by the arguments based on the bound obtained in (7.4.9), we rewrite the middle term as

$$R_n\left(\frac{t}{k}\right)^k f - V_n(t)f = R_n\left(\frac{t}{k}\right)^k f - R_m\left(\frac{t}{k}\right)^k f + \left(R_m\left(\frac{t}{k}\right)^k f - V_m(t)f\right) + \left(V_m(t)f - V_n(t)f\right).$$

By choosing a sufficiently large m, the C_θ norms of the first and third term on the right-hand side are small. Thus $R_n\left(\frac{t}{k}\right)^k f - V_n(t)f$ converges to 0 as $k \to \infty$ uniformly in n. □

Arguments in favour of (7.5.2). In Theorem 7.3.3, we saw that the semi-groups

$$V_n(t)f = \frac{1}{|\Lambda_n|} \log S(t) e^{\sum_{i \in \Lambda_n} f \circ \theta_i}.$$

converge in C_θ to the limiting semigroup $V(t)f$. In Chapter 6, we obtained an explicit expression for $V_n(t)f$ in terms of a Doob-h transform.

Denote by $f_n = \sum_{i \in \Lambda_n} f \circ \theta_i$ and $f_n(s) = \mathbb{V}(t-s)f_n$. Denote by $\mathbb{Q}^{f,n}$ the measure on $D_E(\mathbb{R}^+)$ defined by

$$\frac{\mathrm{d}\mathbb{Q}^{f,n}}{\mathrm{d}\mathbb{P}}(X) = e^{f_n(X(t)) - f_n(s)(X(0))}.$$

Denote by $\mu_n(s)$ the law of $\mathbb{Q}^{f,n}$ at time s. For a translation invariant measure μ, we obtain from Corollary 6.3.16 and the proof of Proposition 6.4.10 that

$$\begin{aligned} &\langle V_n(t)f, \mu\rangle \\ &= \frac{1}{|\Lambda_n|}\langle f_n, \mu_n(s)\rangle - \int_0^t \frac{1}{|\Lambda_n|}\langle A^{f_n} f_n, \mu_n(s)\rangle - \langle H_n f, \mu_n(s)\rangle \mathrm{d}s \end{aligned}$$

The measures $\mu_n(s)$ are not translation invariant as the processes generated by $A^{f_n(s)}$ are not translation invariant. However, the laws should be close to translation invariant measures because for every g the function $A^{f_n(s)}g$ converges to some object $A^F g$ where A^F is a translation invariant generator.

In the setting that $\mu_n(s)$ would be translation invariant, operator duality techniques as used in Chapter 6, would give that the integrand equals

$\mathcal{L}(\mu_n(s), \dot{\mu}_n(s))$, where $\mathcal{L}$ is the Lagrangian introduced in Definition 7.5.4. Also, for a translation invariant measure $\mu_n(s)$, we would have

$$\frac{1}{|\Lambda_n|}\langle f_n, \mu_n(s)\rangle = \langle f, \mu_n(s)\rangle.$$

Thus, we can conclude that

$$\mathbf{V}(t)[f](\mu) \geq \langle V(t)f, \mu\rangle.$$

□

7.6 APPENDIX: ENTROPY DECOMPOSITION AND QUOTIENT SPACES

7.6.1 *Entropy decomposition*

Consider a Polish space X equipped with its Borel σ-algebra $\mathcal{B}_X$. Given a countably generated sub-σ-algebra $\mathcal{F}$ of $\mathcal{B}_X$ and a measure $\mu \in \mathcal{P}(X)$, there is a map $x \in X \mapsto \mu^{\mathcal{F}}(x, \cdot) \in \mathcal{P}(E)$ with the properties that

(a) $x \mapsto \mu^{\mathcal{F}}(x, B)$ is $\mathcal{F}$ measurable for every $B \in \mathcal{B}_X$.

(b) If $B \in \mathcal{F}$ then $\mu^{\mathcal{F}}(x, B) = \mathbb{1}_B(x)$.

(c) $\mu(A \cap B) = \int_A \mu^{\mathcal{F}}(x, B)\mu(x)$ for $A \in \mathcal{F}$ and $B \in \mathcal{B}_X$.

This map is called a *regular conditional probability* of μ given $\mathcal{F}$.

We state Lemma 4.4.7 from Deuschel and Stroock [1989].

Lemma 7.6.1. *Let X be a Polish space and let $\mathcal{F}$ be a countably generated sub-σ-algebra of $\mathcal{B}_X$. Given $\mu, \nu \in \mathcal{P}(X)$, let $x \mapsto \mu^{\mathcal{F}}(x, \cdot)$ and $x \mapsto \nu^{\mathcal{F}}(x, \cdot)$ be regular conditional probabilities of μ and ν given $\mathcal{F}$. Then $x \mapsto S(\nu^{\mathcal{F}}(x, \cdot) \,|\, \mu^{\mathcal{F}}(x, \cdot))$ is $\mathcal{F}$ measurable and*

$$S(\nu \,|\, \mu) = S(\nu_{\mathcal{F}} \,|\, \mu_{\mathcal{F}}) + \int_X S(\nu^{\mathcal{F}}(x, \cdot) \,|\, \mu^{\mathcal{F}}(x, \cdot))\nu_{\mathcal{F}}(\mathrm{d}x), \qquad (7.6.1)$$

where $\mu_{\mathcal{F}}, \nu_{\mathcal{F}}$ are the restrictions of μ and ν to $\mathcal{F}$.

7.6.2 *The quotient space of functions*

A Markov semigroup $\{S(t)\}_{t\geq 0}$ naturally acts on the space $C(E)$. We are interested, however, in how the system behaves after dividing out all translations. Therefore, we must also consider the quotient space of $C(E)$. Recall that $(C_\theta(E), \|\cdot\|_\theta) := (C(E)/\mathcal{I}, \|\cdot\|_\theta)$ is the quotient space where

$$\mathcal{I} := \{f \in C(E) \,|\, |\langle f, \mu\rangle| = 0 \text{ for all } \mu \in \mathcal{P}_\theta(E)\}.$$

The quotient norm is defined by

$$\|f\|_\theta = \inf_{g \in \mathcal{I}} \|f - g\|$$

The next lemma gives a second representation of the quotient norm.

Lemma 7.6.2. *It holds that*

$$\|f\|_\theta = \sup_{\mu \in \mathcal{P}_\theta(E)} |\langle f, \mu \rangle|.$$

This implies that elements of $C(E)/\mathcal{I}$ can on one hand be viewed as classes with a representant in $C(E)$, e.g. the class of $f_0(\sigma) := \sigma_0$ contains $f_i(\sigma) := \sigma_i$ and $f_A(\sigma) := \frac{1}{|A|} \sum_{i \in A} \sigma_i$, for all $A \subseteq \mathbb{Z}^d$ finite. On the other hand, elements of $C(E)/\mathcal{I}$ can be viewed as a linear subspace of $C(\mathcal{P}(E))$, and in that view we have that if $\|F_n - F\|_\theta \to 0$ means $F_n(\mu) - F(\mu) \to 0$, uniformly in the choice of $\mu \in \mathcal{P}_\theta(E)$.

Proof of Lemma 7.6.2. Pick $f \in C(E)$, then

$$\begin{aligned} \|f\|_\theta &= \inf_{g \in \mathcal{I}} \|f - g\| \\ &\geq \inf_{g \in \mathcal{I}} \sup_{\mu \in \mathcal{P}_\theta(E)} |\langle f, \mu \rangle - \langle g, \mu \rangle| = \sup_{\mu \in \mathcal{P}_\theta(E)} |\langle f, \mu \rangle|. \end{aligned}$$

For the other inequality, define for every $n > 0$ and $i \in \Lambda_n$ the function $g_{i,n} := |\Lambda_n|^{-1}(f - f \circ \theta_i)$, and the sum $g_n = \sum_{i \in \lambda_n, i \neq 0} g_i$. Clearly, $g_{i,n} \in \mathcal{I}$ and $g_n \in \mathcal{I}$.

$$\inf_{g \in \mathcal{I}} \|f - g\| \leq \liminf_n \|f - g_n\| = \liminf_n \left\| \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} f \circ \theta_i \right\|.$$

Let l be the value of this $\liminf$. Pick a subsequence n_k and configurations $\sigma_k \in E$ such that $l = \lim_k |\langle f, L_{n_k}(\sigma_k) \rangle|$. Because E is a compact space, the sequence $\{L_{n_k}(\sigma_k)\}$ is relatively compact. Pick a converging subsequence, and let ν be its limit. It is clear that ν is translation invariant. As a consequence, we obtain that

$$\inf_{g \in \mathcal{I}} \|f - g\| \leq l = |\langle f, \nu \rangle| \leq \sup_{\mu \in \mathcal{P}_\theta(E)} |\langle f, \mu \rangle|.$$

□

The next lemma can be found without proof as Proposition 2.34 in Enter et al. [1993]. We repeat the result here, and prove it for completeness.

Lemma 7.6.3. *$\mathcal{I}$ is equal to the closure of the linear span of functions of the type $f - \theta_i f$, where $i \in \mathbb{Z}^d$, $f \in C(E)$.*

For a closed subspace $Y \subseteq X$, denote with

$$Y^{\perp} := \{y^* \in X^* \ : \ \langle y, y^* \rangle = 0 \text{ for all } y \in Y\},$$

the annihilator of Y.

Proof. Denote with

$$S := \overline{span \{f - \theta_i f \ i \in \mathbb{Z}^d, f \in C_b(E)\}}.$$

Suppose that $S \neq \mathcal{I}$. By the Hahn-Banach theorem, it follows that there exists $\mu \in \mathcal{M}_\theta(E)$, such that $\mu \in S^{\perp}$, $\mu \notin \mathcal{I}^{\perp}$.

Because $\mu \in S^{\perp}$, we see that $\langle f \circ \theta_i, \mu \rangle = \langle f, \mu \rangle$ for every $i \in \mathbb{Z}^d$ and $f \in C_b(E)$. However, this implies that $\mu = \theta_i \mu$ for every $i \in \mathbb{Z}^d$, so μ is translation invariant. It is easy to check that the Hahn-Jordan decomposition into μ^+ and μ^- is such that μ^+ and μ^- are translation invariant, see the construction in [Bogachev, 2007, Theorem 3.1.1.].

Using that $\langle f, \nu \rangle = 0$ for all translation invariant probability measures, we obtain the same result for μ^+ and μ^-, which implies that $\langle f, \mu \rangle = 0$ for all $f \in \mathcal{I}$. It follows that $\mu \in \mathcal{I}^{\perp}$, which is a contradiction. Therefore $S = \mathcal{I}$. □

Part III

FUNCTIONAL ANALYTIC METHODS FOR PROBABILITY ON POLISH SPACES

8 STRONGLY CONTINUOUS AND LOCALLY EQUI-CONTINUOUS SEMIGROUPS ON LOCALLY CONVEX SPACES

The study of Markov processes on complete separable metric spaces (E, d) naturally leads to transition semigroups on $C_b(E)$ that are not strongly continuous with respect to the norm. Often, these semigroups turn out to be strongly continuous with respect to the weaker locally convex strict topology and in Chapter 9 we will prove that the transition semigroup of the solution to a well-posed martingale problem is continuous for the strict topology.

This naturally leads to the study of strongly continuous semigroups on more general locally convex spaces. This chapter, with the exception of the Trotter-Kato approximation results and its corollaries, is based on

Richard Kraaij. Strongly continuous and locally equi-continuous semigroups on locally convex spaces. *Semigroup Forum*, 92(1):158–185, 2016a. ISSN 1432-2137. doi: 10.1007/s00233-015-9689-1,

in which such a class of semigroups is considered. We immediately note that the results in this chapter and Chapter 9 serve as a starting point for the extension of the results in Chapter 6 to the setting of Polish spaces.

We start out with some historical context. The theory of equi-continuous semigroups is developed analogously to the Banach space situation for example in Yosida [1978]. When characterising the operators that generate a semigroup, the more general context of locally equi-continuous semigroups introduces new technical challenges. Notably, the integral representation of the resolvent is not necessarily available. To solve this problem Kōmura [1968], Ōuchi [1973], Dembart [1974] have studied various generalised resolvents. More recently, Albanese and Kühnemund [2002] also study asymptotic pseudo-resolvents and give a Trotter-Kato approximation result and the Lie-Trotter product formula.

A different approach is used in recent papers where a subclass of locally convex spaces (X, τ) is considered for which the ordinary representation of the resolvent can be obtained. Essentially, these spaces are also equipped with a norm $\|\cdot\|$ such $(X, \|\cdot\|)$ is Banach and such that the dual $(X, \tau)'$

is norming for $(X, \|\cdot\|)$. Bi-continuous semigroups have been studied in Kühnemund [2003], Albanese and Mangino [2004], Farkas [2004], in which the Hille-Yosida, Trotter Approximation theorem and perturbation results have been shown. Bi-continuity has the drawback, however, that it is a non-topological notion. Kunze [2009, 2011] studies semigroups of which he assumes that the resolvent can be given in integral form. His notions are topological, and he gives a Hille-Yosida theorem for equi-continuous semigroups.

In Section 8.2, we start with some minor results for locally convex spaces (X, τ) that are strong Mackey. These spaces are of interest, because a strongly continuous semigroup on a strong Mackey space is automatically locally equi-continuous, which extends a result by Kōmura [1968] for barrelled spaces.

From that point onward, we will consider sequentially complete locally convex spaces (X, τ) that are additionally equipped with an 'auxiliary' norm. We assume that the norm topology is finer than τ, but that the norm and τ bounded sets coincide. In Section 8.3, we define $\mathcal{N}$ as the set of τ continuous semi-norms that are bounded by the norm. We say that the space satisfies Convexity Condition C if $\mathcal{N}$ is closed under taking countable convex combinations. This property allows the generalisation of a number of results in the Banach space theory. First of all, strong continuity of a semigroup on a space satisfying Condition C implies the exponential boundedness of the semigroup. Second, in Section 8.4, we show that the resolvent can be expressed in integral form. Third, in Section 8.5, we give a straightforward proof of the Hille-Yosida theorem for strongly continuous and locally equi-continuous semigroups. Finally, in Sections 8.7 and 8.6, we prove the Trotter-Kato theorem and the Chernoff and Trotter product formulas.

The strength of spaces that satisfy Condition C and the set $\mathcal{N}$ is that results from the Banach space theory generalise by replacing the norm by semi-norms from $\mathcal{N}$. Technical difficulties arising from working with the set $\mathcal{N}$ instead of the norm are overcome by the probabilistic techniques of stochastic domination and Chernoff's bound, see Appendix 8.10.

In Section 8.8, we consider τ bi-continuous semigroups. We show that if the so called mixed topology $\gamma = \gamma(\|\cdot\|, \tau)$, introduced by Wiweger [1961], has good sequential properties, bi-continuity of a semigroup for τ is equivalent to strong continuity and local equi-continuity for γ.

In Section 8.9, we show that the spaces $(C_b(E), \beta)$ and $(\mathcal{B}(\mathfrak{H}), \beta)$, where E is a Polish space, $\mathfrak{H}$ a Hilbert space and where β is their respective strict topology, are strong Mackey and satisfy Condition C. This implies that our results can be applied to Markov transition semigroups on $C_b(E)$ and quantum dynamical semigroups on $\mathcal{B}(\mathfrak{H})$.

8.1 PRELIMINARIES

We recall some notation. Let (X, τ) be a locally convex space. We call the family of operators $\{T(t)\}_{t\geq 0}$ a *semigroup* if $T(0) = \mathbb{1}$ and $T(t)T(s) = T(t+s)$ for $s, t \geq 0$. A family of (X, τ) continuous operators $\{T(t)\}_{t\geq 0}$ is called a *strongly continuous semigroup* if $t \mapsto T(t)x$ is continuous and *weakly continuous* if $t \mapsto \langle T(t)x, x' \rangle$ is continuous for every $x \in X$ and $x \in X'$.

We call $\{T(t)\}_{t\geq 0}$ a *locally equi-continuous* family if for every $t \geq 0$ and continuous semi-norm p, there exists a continuous semi-norm q such that $\sup_{s\leq t} p(T(s)x) \leq q(x)$ for every $x \in X$.

Furthermore, we call $\{T(t)\}_{t\geq 0}$ a *quasi equi-continuous* family if there exists $\omega \in \mathbb{R}$ such that for every continuous semi-norm p, there exists a continuous semi-norm q such that $\sup_{s\geq 0} e^{-\omega t} p(T(s)x) \leq q(x)$ for every $x \in X$. Finally, we abbreviate strongly continuous and locally equi-continuous semigroup to *SCLE* semigroup.

We use the following notation for duals and topologies. X^* is the algebraic dual of X and X' is the continuous dual of (X, τ). Finally, X^+ is the sequential dual of X:

$$X^+ := \{f \in X^* \mid f(x_n) \to 0, \text{ for every sequence } x_n \in X \text{ converging to } 0\}.$$

We write $(X, \sigma(X, X'))$, $(X, \mu(X, X'))$, $(X, \beta(X, X'))$, for X equipped with the weak, Mackey or strong topology. Similarly, we define the weak, Mackey and strong topologies on X'. For any topology τ, we use τ^+ to denote the strongest locally convex topology having the same convergent sequences as τ, Webb [1968].

8.2 STRONG MACKEY SPACES: CONNECTING STRONG CONTINUITY AND LOCAL EQUI-CONTINUITY

We start with a small exposition on a subclass of locally convex spaces that imply nice 'local' properties of semigroups.

[Kōmura, 1968, Proposition 1.1] showed that on a barrelled space a strongly continuous semigroup is automatically locally equi-continuous. This fact is proven for the smaller class of Banach spaces in [Engel and Nagel, 2000, Proposition I.5.3], where they use the strong continuity of $\{T(t)\}_{t\geq 0}$ at $t=0$ and the Banach Steinhaus theorem.

This approach disregards the fact that $\{T(t)\}_{t\geq 0}$ is strongly continuous for all $t \geq 0$ and [Kunze, 2009, Lemma 3.8] used this property to show that, in the case that every weakly compact subset of the dual is equi-continuous, strong continuity implies local equi-continuity.

Definition 8.2.1. We say that a locally convex space (X, τ) is *strong Mackey* if all $\sigma(X', X)$ compact sets in X' are equi-continuous.

Following the proof of Lemma 3.8 in Kunze [2009], we obtain the following result.

Lemma 8.2.2. *If a semigroup $\{T(t)\}_{t\geq 0}$ of continuous operators on a strong Mackey space is strongly continuous, then the semigroup is locally equi-continuous.*

Proof. Fix $T \geq 0$. It follows from 39.3.(4) in Köthe [1979] that $\{T(t)\}_{t\leq T}$ is equi-continuous if the set

$$\mathcal{T}'(U) := \left\{T'(t)x' \mid t \leq T, x' \in U\right\}$$

is equi-continuous in X' for every equi-continuous set $U \subseteq X'$. So pick an equi-continuous set U in X'. First of all, note that we can replace U by its $\sigma(X', X)$ closure, because the $\sigma(X', X)$ closure of an equi-continuous set is equi-continuous. We show that $\mathcal{T}'(U)$ is relatively compact, so that the fact that (X, τ) is of type A implies that $\mathcal{T}'(U)$ is equi-continuous.

Pick a net $\alpha \mapsto T'(t_\alpha)\mu_\alpha$, where $t_\alpha \leq T$ and $\mu_\alpha \in U$. The interval $[0, T]$ is compact, and because U is closed and equi-continuous it is $\sigma(X', X)$ compact by the Bourbaki-Alaoglu theorem [Köthe, 1969, 20.9.(4)], which implies that we can restrict ourselves to a net α such that $t_\alpha \to t_0$ for some $t_0 \leq T$ and $\mu_\alpha \to \mu_0$ weakly, where $\mu_0 \in U$.

We show that $T'(t_\alpha)\mu_\alpha \to T'(t_0)\mu_0$ weakly. For every $x \in X$, we have

$$\begin{aligned} &|\langle T(t_\alpha)x, \mu_\alpha\rangle - \langle T(t_0)x, \mu_0\rangle| \\ &\leq |\langle T(t_\alpha)x, \mu_\alpha\rangle - \langle T(t_0)x, \mu_\alpha\rangle| + |\langle T(t_0)x, \mu_\alpha\rangle - \langle T(t_0)x, \mu_0\rangle| \end{aligned} \tag{8.2.1}$$

The second term converges to 0, because $\mu_\alpha \to \mu_0$ in $(X', \sigma(X', X))$ and the first term goes to zero because the set U is equi-continuous and $\{T(t)\}_{t\geq 0}$ is strongly continuous. □

We start with a proposition that gives sufficient conditions for a space to be strong Mackey.

Proposition 8.2.3. *Any of the following properties implies that (X, τ) is strong Mackey.*

(a) (X, τ) is barrelled.

(b) (X, τ) is sequentially complete and bornological.

(c) The space (X, τ) is sequentially complete, Mackey and the continuous dual X' of X is equal to the sequential dual X^+ of X.

A space for which $X^+ = X'$ is called a *Mazur* spaceWilansky [1981], or *weakly semi bornological*Beatty and Schaefer [1996]. Note that a Mackey Mazur space satisfies $\tau = \tau^+$ by Corollary 7.6 in Wilansky [1981]. On the other hand, a space such that $\tau^+ = \tau$ is Mazur.

Proof. By [Köthe, 1969, 21.2.(2)], the topology of a barrelled space coincides with the strong topology $\beta(X, X')$, in other words, all weakly bounded, and thus all weakly compact, sets are equi-continuous.

Statement (b) follows from (a) as a sequentially complete bornological space is barrelled, see 28.1.(2) in Köthe [1969].

We now prove (c). The sequential completeness of (X, τ) and $X' = X^+$ imply that $(X', \mu(X', X))$ is complete by Corollary 3.6 in Webb [1968].

Let $K \subseteq X'$ be $\sigma(X', X)$ compact. By Krein's theorem [Köthe, 1969, 24.4.(4)], the completeness of $(X', \mu(X', X))$ implies that the absolutely convex cover of K is also $\sigma(X', X)$ compact. By the fact that τ is the Mackey topology, every absolutely convex compact set in $(X', \sigma(X', X))$ is equi-continuous [Köthe, 1969, 21.4.(1)]. This implies that K is also equi-continuous. □

As an application of Lemma 8.2.2, we have the following proposition, which states that strong continuity is determined by local properties of the semi-group.

Proposition 8.2.4. *A semigroup $\{T(t)\}_{t\geq 0}$ of continuous operators on a strong Mackey space is strongly continuous if and only if the following two statements hold* *(i) There is a dense subset* $D \subseteq X$ *such that* $\lim_{t\to 0} T(t)x = x$ *for every* $x \in D$.

(ii) $\{T(t)\}_{t\geq 0}$ *is locally equi-continuous.*

In the Banach space setting, strong continuity of the semigroup is equivalent to strong continuity at $t = 0$, see Proposition I.5.3 in Engel and Nagel [2000]. In the more general situation, this equivalence does not hold, see Example 5.2 in Kunze [2009].

Proof. Suppose that $\{T(t)\}_{t\geq 0}$ is strongly continuous. (i) follows immediately and (ii) follows from Lemma 8.2.2.

For the converse, suppose that we have (i) and (ii) for the semigroup $\{T(t)\}_{t\geq 0}$. First, we show that $\lim_{t\downarrow 0} T(t)x = x$ for every $x \in X$. Pick some $x \in X$ and let x_α be an approximating net in D and let p be a continuous semi-norm and fix $\varepsilon > 0$. We have

$$p(T(t)x - x) \leq p(T(t)x - T(t)x_\alpha) + p(T(t)x_\alpha - x_\alpha) + p(x_\alpha - x).$$

Choose α large enough such that the first and third term are smaller than $\varepsilon/3$. This can be done independently of t, for t in compact intervals, by the local equi-continuity of $\{T(t)\}_{t\geq 0}$. Now let t be small enough such that the middle term is smaller than $\varepsilon/3$.

We proceed with the proving the strong continuity of $\{T(t)\}_{t\geq 0}$. The previous result clearly gives us $\lim_{s\downarrow t} T(s)x = T(t)x$ for every $x \in X$, so we are left to show that $\lim_{s\uparrow t} T(s)x = T(t)x$.

For $h > 0$ and $x \in X$, we have $T(t-h)x - T(t)x = T(t-h)\,(x - T(h)x)$, so the result follows by the right strong continuity and the local equi-continuity of the semigroup $\{T(t)\}_{t\geq 0}$. □

A second consequence of Lemma 8.2.2, for quasi complete spaces, follows from Proposition 1 in Albanese et al. [2012].

Proposition 8.2.5. *Suppose that we have a semigroup of continuous operators* $\{T(t)\}_{t\geq 0}$ *on a quasi complete strong Mackey space. Then the semigroup is strongly continuous if and only if it is weakly continuous and locally equi-continuous.*

Proposition 8.2.6. *Suppose that* $(X, \mu(X, X'))$ *and* $(X', \mu(X', X))$ *are quasi-complete and let* $\{T(t)\}_{t\geq 0}$ *be a SCLE semigroup. Then, the dual semigroup* $\{T'(t)\}_{t\geq 0}$ *is SCLE for the Mackey topology* $\mu(X', X)$ *on* X'.

Proof. Clearly, the dual semigroup is strongly continuous for $\sigma(X', X)$. Therefore, $\{T'(t)\}_{t\geq 0}$ is strongly continuous for $\mu(X', X)$ if we can show that it is locally equi-continuous.

Pick a $\sigma(X, X')$ compact set $K \subseteq X$ and fix $t_0 > 0$. We need to prove that $U = \{T(t)f \mid f \in K, t \leq t_0\}$ is weakly compact. Pick a net $g_\alpha = T(t_\alpha)f_\alpha \in U$, for $t_\alpha \leq t_0$ and $f_\alpha \in K$. Pick a subnet $\beta \subseteq \alpha$ such that $f_\beta \to f$ and $t_\beta \to t$. We show that $T(t_\beta)f_\beta$ converges to $T(t)f$. Fix $\mu \in X'$.

$$
\begin{aligned}
&|\langle T(t_\beta)f_\beta, \mu\rangle - \langle T(t)f, \mu\rangle| \\
&\quad \leq |\langle T(t_\beta)f_\beta, \mu\rangle - \langle T(t_\beta)f, \mu\rangle| - |\langle T(t_\beta)f, \mu\rangle - \langle T(t)f, \mu\rangle| \\
&\quad \leq \left|\langle f_\beta, T'(t_\beta)\mu\rangle - \langle f, T'(t_\beta)\mu\rangle\right| - |\langle T(t_\beta)f, \mu\rangle - \langle T(t)f, \mu\rangle| \,.
\end{aligned}
$$

The first term on the right hand side converges to 0 by the local equi-continuity of $\{T(t)\}_{t\geq 0}$. The second term clearly also converges to 0. The result follows by Proposition 8.2.5. □

As in the Banach space situation, it would be nice to have some condition that implies that the semigroup, suitably rescaled is globally bounded. We directly run into major restrictions.

Example 8.2.7. Consider $C_c^\infty(\mathbb{R})$ the space of test functions, equipped with its topology as a countable strict inductive limit of Fréchet spaces. This space is complete [Treves, 1967, Theorem 13.1], Mackey [Treves, 1967, Propositions 34.4 and 36.6] and $C_c^\infty(\mathbb{R})^+ = C_c^\infty(\mathbb{R})'$ as a consequence of [Treves, 1967, Corollary 13.1.1].

Define the semigroup $\{T(t)\}_{t\geq 0}$ by setting $(T(t)f)(s) = f(t+s)$. This semigroup is strongly continuous, however, even if exponentially rescaled, it can never be globally bounded by 19.4.(4) Köthe [1969].

So even if (X, τ) is strong Mackey, we can have semigroups that have undesirable properties. This issue is serious. For example, in the above example, formally writing the resolvent corresponding to the semigroup in its integral form, yields a function which is not in $C_c^\infty(\mathbb{R})$. One can work around this problem, see for example Dembart [1974], Kōmura [1968], Ōuchi [1973] which were already mentioned in the introduction.

However, motivated by the study of Markov processes, where the resolvent informally corresponds to evaluating the semigroup at an exponential random time, we would like to work in a framework in which the ordinary integral representation for the resolvent holds.

8.3 A SUITABLE STRUCTURE OF BOUNDED SETS

In this section, we shift our attention to another type of locally convex spaces. As a first major consequence, we are able to show in Corollary 8.3.7 an analogue of the exponential boundedness of a strongly continuous semigroup on a Banach space. This indicates that we may be able to mimic major parts of the Banach space theory.

Suppose that (X, τ) is a locally convex space, and suppose that X can be equipped with a norm $\|\cdot\|$, such that τ is weaker than the norm topology. It follows that bounded sets for the norm are bounded sets for τ. This means that if we have a τ-continuous semi-norm p, then there exists some $M > 0$ such that $\sup_{x:\|x\|\leq 1} p(x) \leq M$. Therefore, $p(x) \leq M \|x\|$ for every x, i.e. every τ-continuous semi-norm is dominated by a constant times the norm.

Definition 8.3.1. Let (X, τ) be equipped with a norm $\|\cdot\|$ such that τ is weaker than the norm topology. Denote by $\mathcal{N}$ the τ-continuous semi-norms that satisfy $p(\cdot) \leq \|\cdot\|$. We say that $\mathcal{N}$ is *countably convex* if for any sequence p_n of semi-norms in $\mathcal{N}$ and $\alpha_n \geq 0$ such that $\sum_n \alpha_n = 1$, we have that $p(\cdot) := \sum_n \alpha_n p_n(\cdot) \in \mathcal{N}$.

We start with exploring the situation where τ and $\|\cdot\|$ have the same bounded sets.

Condition (Boundedness condition B)**.** A locally convex space (X, τ) also equipped with a norm $\|\cdot\|$, denoted by $(X, \tau, \|\cdot\|)$, satisfies *Condition B* if

(a) τ is weaker than the norm topology.

(b) Both topologies have the same bounded sets.

Remark 8.3.2. Suppose that (X, τ) is a locally convex space, and suppose that $\|\cdot\|$ is a norm on X such that the norm topology is stronger than τ, but such that the norm topology has less bounded sets than τ.

In this case, it is useful to consider the *mixed* topology $\gamma = \gamma(\|\cdot\|, \tau)$, introduced in Wiweger [1961]. In Section 8.8, we study the relation of bi-continuous semigroups for τ with SCLE semigroups for γ.

We introduce some notation. We write $X'_n := (X, \|\cdot\|)'$ and $X'_\tau := (X, \tau)'$. Also, we denote $B_n := \{x' \in X'_n \mid \|x'\|' \leq 1\}$, where $\|\cdot\|'$ is the operator norm on X'_n. Finally, we set $B_\tau = B_n \cap X'_\tau$. We start with a well known theorem that will aid our exposition.

Theorem 8.3.3 (Bipolar Theorem)**.** *Let (X, τ) be a locally convex space and let $\|\cdot\|$ be a norm on X. Let p be a τ lower semi-continuous semi-norm* *such that $p \leq \|\cdot\|$. Then there exists a absolutely convex weakly bounded set $\mathfrak{S} := \{p \leq 1\}^\circ \subseteq B_\tau$ such that*

$$p(x) = \sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle|.$$

Furthermore, p is continuous if and only if $\mathfrak{S}$ is an equi-continuous set.

Proof. The result follows from 20.8.(5) and 21.3.(1) in Köthe [1969]. The fact that $\mathfrak{S} \subseteq B_\tau$ is a consequence of $p \in \mathcal{N}$. □

Lemma 8.3.4. *Let (X, τ) be sequentially complete locally convex space, and $\|\cdot\|$ a norm on X such that the norm topology is stronger than τ. Then the following are equivalent.*

(a) The norm bounded sets equal the τ bounded sets.

(b) $\|\cdot\|$ is τ lower semi-continuous.

(c) The norm can be expressed as $\|x\| = \sup_{x' \in B_\tau} |\langle x, x' \rangle|$.

In all cases, the topology generated by $\|\cdot\|$ is the $\beta(X, X'_\tau)$ topology and is Banach. The norm can equivalently be written as

$$\|x\| = \sup_{p \in \mathcal{N}} p(x). \tag{8.3.1}$$

Proof. We start with the proof of (a) to (b). Define the $\beta(X, X'_\tau)$ continuous norm $|||x||| := \sup_{x' \in B_\tau} |\langle x, x' \rangle|$. Note that $|||\cdot||| \leq \|\cdot\|$ by construction. It follows that the $\|\cdot\|$ topology is stronger than the $|||\cdot|||$ topology, which is in turn stronger than τ. The bounded sets of the two extremal topologies are the same, so the $\|\cdot\|$ and the $|||\cdot|||$ bounded sets coincide. Thus, there is some $c \geq 1$ such that $|||\cdot||| \leq \|\cdot\| \leq c\,|||\cdot|||$. But this means that $\|\cdot\|$ is $\beta(X, X'_\tau)$ continuous, and thus τ lower semi-continuous.

Now assume (b), we prove (a). Because τ is weaker than the norm topology, it follows that the norm topology has less bounded sets. On the other hand, as the norm is τ lower semi-continuous, it is continuous for the strong topology $\beta(X, X'_\tau)$. Therefore, the strong topology has less bounded sets than the norm topology. Because (X, τ) is sequentially complete, the Banach-Mackey theorem, 20.11.(3) in Köthe [1969] shows that the strongly bounded sets and the τ bounded sets coincide, which implies (a).

(c) clearly implies (b) and (b) implies (c) by the Bipolar theorem.

(X, τ) is Banach by 18.4.(4) in Köthe [1969]. □

The usefulness of $\mathcal{N}$ becomes clear from the next three results. Intuitively, the next two lemmas tell us that in the study of semigroups on these locally convex spaces the collection $\mathcal{N}$ replaces the role that the norm plays for semigroups on Banach spaces.

Lemma 8.3.5. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition B. Let I be some index set and let $(T_\alpha)_{\alpha\in I}$ be (X, τ) to (X, τ) continuous operators. Then the following are equivalent*

(a) The family $\{T_\alpha\}_{\alpha\in I}$ is τ-equi-continuous and $\sup_{\alpha\in I} \|T_\alpha\| \leq M$.

(b) For every $p \in \mathcal{N}$, there is $q \in \mathcal{N}$ such that $\sup_{\alpha\in I} p(T_\alpha x) \leq Mq(x)$ for all $x \in X$.

Furthermore, if the family $\{T_\alpha\}_{\alpha\in I}$ is τ-equi-continuous, then there exists $M \geq 0$ such that these properties hold.

Proof. The implication (b) to (a) follows from (8.3.1). For the proof of (a) to (b), fix some semi-norm $p \in \mathcal{N}$. Because the family $\{T_\alpha\}_{\alpha\in I}$ is τ-equi-continuous, there is some continuous semi-norm $\hat{q}$ such that $\sup_{\alpha\in I} p(T_\alpha x) \leq \hat{q}(x)$. This implies that $q(x) := M^{-1} \sup_{\alpha\in I} p(T_\alpha x)$ is τ-continuous. We conclude that $q \in \mathcal{N}$ by noting that

$$q(x) = \frac{1}{M} \sup_{\alpha\in I} p(T_\alpha x) \leq \frac{1}{M} \sup_{\alpha\in I} \|T_\alpha x\| \leq \|x\| .$$

If the family $\{T_\alpha\}_{\alpha\in I}$ is τ-equi-continuous, it is τ-equi-bounded which implies that there is some $M \geq 0$ such that $\sup_{\alpha\in I} \|T_\alpha\| \leq M$ by Condition B. □

In particular, we have the following result.

Lemma 8.3.6. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition B and $\{T(t)\}_{t\geq 0}$ be a semigroup of continuous operators. Then the following are equivalent.*

(a) $\{T(t)\}_{t\geq 0}$ is locally equi-continuous.

(b) For every $t \geq 0$ there exists $M \geq 1$, such that for every $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that for all $x \in X$

$$\sup_{s\leq t} p(T(s)x) \leq Mq(x).$$

As a corollary, we obtain an exponential growth bound.

Corollary 8.3.7. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition B. For a locally equicontinuous semigroup $\{T(t)\}_{t\geq 0}$, there is $M \geq 1$ and $\omega \in \mathbb{R}$ such that for every $T \geq 0$ and every $p \in \mathcal{N}$ there is a $q \in \mathcal{N}$ such that for all $x \in X$*

$$\sup_{t\leq T} e^{-\omega t} p(T(t)x) \leq Mq(x).$$

Proof. Pick $M \geq 1$ such that for every $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that

$$\sup_{t\leq 1} p(T(t)x) \leq Mq(x) \tag{8.3.2}$$

for every $x \in X$. Without loss of generality, we can always choose $q \in \mathcal{N}$ to dominate p. We use this property to construct an increasing sequence of semi-norms in $\mathcal{N}$.

Fix some $p \in \mathcal{N}$ and pick $q_0 \geq p$ such that it satisfies the property in equation (8.3.2). Inductively, let $q_{n+1} \in \mathcal{N}$ be a semi-norm such that $q_{n+1} \geq q_n$ and $\sup_{t\leq 1} q_{n+1}(T(t)x) \leq Mq_n(x)$. Now let $t \geq 0$. Express $t = s + n$ where $n \in \mathbb{N}$ and $0 \leq s < 1$, then it follows that

$$p(T(t)x) \leq Mq_0(T(n)) \leq \cdots \leq M^{n+1} q_n(x) \leq Me^{t \log M} q_n(x).$$

Setting $\omega = \log M$, we obtain $\sup_{t\leq T} e^{-\omega t} p(T(t)x) \leq Mq_{\lceil T \rceil}(x)$ for every $x \in X$. □

This last result inspires the following definition, which is clearly analogous to the situation for semigroups in Banach spaces.

Definition 8.3.8. We say that a semigroup on a space $(X, \tau, \|\cdot\|)$ that satisfies Condition B is of type (M, ω), $M \geq 1$ and $\omega \in \mathbb{R}$, if for every $p \in \mathcal{N}$ and $T \geq 0$ there exists $q \in \mathcal{N}$ such that

$$\sup_{t\leq T} e^{-\omega t} p(T(t)x) \leq Mq(x)$$

for all $x \in X$. We say that it is of type $(M, \omega)^*$ if

$$\sup_{t\geq 0} e^{-\omega t} p(T(t)x) \leq Mq(x).$$

Furthermore, we define the growth bound ω_0 of $\{T(t)\}_{t\geq 0}$ by

$$\omega_0 := \inf \{\omega \in \mathbb{R} \mid \exists M \geq 1 \text{ such that } \{T(t)\}_{t\geq 0} \text{ is of type } (M, \omega)\}.$$

It follows that if a semigroup is of type (M, ω) for some M and ω, then it is locally equi-continuous. Furthermore, if it is of type $(M, \omega)^*$ it is quasi equi-continuous.

Condition (Convexity condition C)**.** A locally convex space (X, τ) also equipped with a norm $\|\cdot\|$, denoted by $(X, \tau, \|\cdot\|)$, satisfies *Condition C* if

(a) (X, τ) is sequentially complete.

(b) τ is weaker than the norm topology.

(c) Both topologies have the same bounded sets.

(d) $\mathcal{N}$ is countably convex.

We give some conditions that imply that $\mathcal{N}$ is countably convex. Interestingly, the same spaces that are strong Mackey, if equipped with a suitable norm, also turn out to satisfy Condition C. The countable convexity is equivalent to *property (L)*, see Theorem 2.2 in Saxon and Sánchez Ruiz [1997].

We say that a space (X, τ) is *transseparable* if for every open neighbourhood U of 0, there is a countable subset $A \subseteq X$ such that $A + U = X$. Note that a separable space is transseparable.

Proposition 8.3.9. *Let (X, τ) be a sequentially complete locally convex space that is also equipped with some norm $\|\cdot\|$ such that τ is weaker than the norm topology and such that both topologies have the same bounded sets. The set $\mathcal{N}$ is countably convex if either of the following hold*

(a) $\tau^+ = \tau$.

(b) (X, τ) *is Mackey and* $(X'_\tau, \sigma(X'_\tau, X))$ *is locally complete.*

(c) (X, τ) *is transseparable and* $(X'_\tau, \sigma(X'_\tau, X))$ *is sequentially complete.*

(d) τ *equals the weak topology* $\sigma(X, X'_\tau)$ *and* $(X'_\tau, \sigma(X'_\tau, X))$ *is locally complete.*

Furthermore, (b) holds for all three classes of spaces mentioned in Proposition 8.2.3.

Note that $\tau^+ = \tau$ is satisfied if τ is sequential. This holds for example if (X, τ) is Banach or Fréchet. Local completeness of $(X'_\tau, \sigma(X'_\tau, X))$ is implied by sequential completeness of $(X'_\tau, \sigma(X'_\tau, X))$ [Carreras and Bonet, 1987, Corollary 5.1.8]. If $(X'_\tau, \sigma(X'_\tau, X))$ is locally complete, then (X, τ) is called *dual locally complete*, Saxon and Sánchez Ruiz [1997].

Proof of Proposition 8.3.9. Pick $p_n \in \mathcal{N}$ and $\alpha_n \geq 0$, such that $\sum_n \alpha_n = 1$. Define $p(\cdot) = \sum_n \alpha_n p_n(\cdot)$. First of all, it is clear that p is a semi-norm. Thus, we need to show that p is τ continuous.

Suppose that $\tau^+ = \tau$. By Theorem 7.4 in Wilansky [1981] a sequentially continuous semi-norm is continuous. Thus it is enough to show sequential continuity of p. This follows directly from the dominated convergence theorem, as every p_n is continuous and $p_n(\cdot) \leq \|\cdot\|$.

For the proof of (b), (c) and (d), we need the explicit form of the semi-norms in $\mathcal{N}$ given in Theorem 8.3.3. Recall that $B_\tau := \{x' \in (X,\tau)' \mid \|x'\|' \leq 1\}$. If $q \in \mathcal{N}$, then there is an absolutely convex closed and equi-continuous set $\mathfrak{S} \subseteq B$ such that

$$q(\cdot) = \sup_{x' \in \mathfrak{S}} |\langle \cdot, x' \rangle|.$$

We proceed with the proof of (b). The sequence of semi-norms p_n are all of the type described above. So let $\mathfrak{S}_n$ be the equi-continuous subset of B_τ that corresponds to p_n. Define the set

$$\mathfrak{S} := \left\{ \lim_{n\to\infty} \sum_{i=1}^{n} \alpha_i u_i \;\middle|\; u_i \in \mathfrak{S}_i \right\}.$$

The dual local completeness of (E, τ) shows that these limits exists by Theorem 2.3 in Saxon and Sánchez Ruiz [1997]. Under the stronger assumption that $(X'_\tau, \sigma(X'_\tau, X))$ is sequentially complete this is obvious.

To finish the proof of case (b), we prove two statements. The first one is that $p(x) = \sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle|$, the second is that $\mathfrak{S}$ is τ equi-continuous. Together these statements imply that p is τ continuous.

We start with the first statement. For every $x \in X$, there are $x'_n \in \mathfrak{S}_n$ such that $p_n(x) = \langle x, x'_n \rangle$ by construction. Therefore,

$$p(x) = \sum_{n=1}^{\infty} \alpha_n \langle x, x'_n \rangle = \langle x, \sum_{n=1}^{\infty} \alpha_n x'_n \rangle = \sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle|.$$

On the other hand,

$$\sup_{y' \in \mathfrak{S}} |\langle x, y' \rangle| = \sup_{\substack{y'_n \in \mathfrak{S}_n \\ n \geq 1}} |\langle x, \sum_{n=1}^{\infty} \alpha_n y'_n \rangle|$$

$$\leq \sum_{n=1}^{\infty} \alpha_n \sup_{y'_n \in \mathfrak{S}_n} |\langle x, y'_n \rangle| \leq p(x).$$

Combining these statements, we see that $p(x) = \sup_{x' \in \mathfrak{S}} |\langle x, x' \rangle|$.

We prove the equi-continuity of $\mathfrak{S}$. Consider $\mathfrak{S}_n$ equipped with the restriction of the $\sigma(X'_\tau, X)$ topology. Define the product space $\mathcal{P} := \prod_{n=1}^{\infty} \mathfrak{S}_n$ and equip it with the product topology. Because every closed equi-continuous set is $\sigma(X'_\tau, X)$ compact by the Bourbaki-Alaoglu theorem [Köthe, 1969, 20.9.(4)], $\mathcal{P}$ is also compact.

Let $\phi : \mathcal{P} \to \mathfrak{S}$ be the map defined by $\phi(\{x'_n\}_{n\geq 1}) = \sum_{n\geq 1} \alpha_n x'_n$. Clearly, ϕ is surjective. We prove that ϕ is continuous. Let $\beta \mapsto \{x'_{\beta,n}\}_{n\geq 1}$ be a net converging to $\{x'_n\}_{n\geq 1}$ in $\mathcal{P}$. Fix $\varepsilon > 0$ and $f \in X$. Now let N be large enough such that $\sum_{n>N} \alpha_n < \frac{1}{4\|f\|}\varepsilon$ and pick β_0 such that for every $\beta \geq \beta_0$ we have $\sum_{n\leq N} |\langle f, x'_{\beta,n} - x'_n \rangle| \leq \frac{1}{2}\varepsilon$. Then, it follows for $\beta \geq \beta_0$ that

$$\begin{aligned}
&\left|\phi(\{x'_{\beta,n}\}_{n\geq 1}) - \phi(\{x'_n\}_{n\geq 1})\right| \\
&\leq \sum_{n\leq N} \alpha_n |\langle f, x'_{\beta,n} - x'_n \rangle| + \sum_{n>N} \alpha_n |\langle f, x'_{\beta,n} - x'_n \rangle| \\
&\leq \frac{1}{2}\varepsilon + \sum_{n>N} \alpha_n \|f\| \left\|x'_{\beta,n} - x'_n\right\|' \\
&\leq \frac{1}{2}\varepsilon + 2\|f\| \frac{1}{4\|f\|}\varepsilon \\
&= \varepsilon,
\end{aligned}$$

where we use in line four that all $x'_{\beta,n}$ and x'_n are elements of B_τ. As a consequence, $\mathfrak{S}$, as the continuous image of a compact set, is $\sigma(X'_\tau, X)$ compact. $\mathfrak{S}$ is also absolutely convex, as it is the image under an affine map of an absolutely convex set. Because (X, τ) is Mackey, this yields that $\mathfrak{S}$ is equi-continuous, which in turn implies that p is τ continuous.

The proof of (d) follows immediately from the proof of (b) as the set $\mathfrak{S}$ is a singleton and is thus weakly equi-continuous for trivial reasons.

The proof of (c) follows along the lines of the proof of (b). The proof changes slightly as we can not use that a $\sigma(X'_\tau, X)$ compact set is equi-continuous. We replace this by using transseparability. We adapt the proof of (b).

Because (X, τ) is transseparable, the $\sigma(X'_\tau, X)$ topology restricted to $\mathfrak{S}_n$ is metrisable by Lemma 1 in Pfister [1976]. This implies that the product space $\mathcal{P} := \prod_{n=1}^{\infty} \mathfrak{S}_n$ with the product topology $\mathcal{T}$ is metrisable.

By 34.11.(2) in Köthe [1979], we obtain that $\mathfrak{S}$, as the continuous image of a metrisable compact set, is metrisable. The equi-continuity of $\mathfrak{S}$ now follows from corollaries of Kalton's closed graph theorem, see Theorem 2.4

and Theorem 2.6 in Kalton [1971] or 34.11.(6) and 34.11.(9) in Köthe [1979]. Note that we need the weak sequential completeness of the dual space for the closed graph theorem.

We show that that the spaces mentioned in Proposition 8.2.3 satisfy (b). If (X, τ) is Mackey and $X' = X^+$, then $\tau^+ = \tau$ by Theorem 7.4 and Corollary 7.5 in Wilansky [1981]. A sequentially complete bornological space is barrelled, so to complete the proof, we only need to consider barrelled spaces. The topology of a barrelled space coincides with the strong topology, therefore a weak Cauchy sequence in X'_τ is equi-continuous and thus has a weak limit [Treves, 1967, Proposition 32.4]. □

8.4 INFINITESIMAL PROPERTIES OF SEMIGROUPS

We now start with studying the infinitesimal properties of a semigroup. Besides the local equi-continuity which we assumed for all results in previous section, we will now also assume strong continuity.

We directly state the following weaker form of Proposition 8.2.4 for later reference.

Lemma 8.4.1. *Let $\{T(t)\}_{t\geq 0}$ be a locally equi-continuous semigroup on a locally convex space (X, τ). Then the following are equivalent.*

(a) $\{T(t)\}_{t\geq 0}$ is strongly continuous.

(b) There is a dense subset $D \subseteq X$ such that $\lim_{t\downarrow 0} T(t)x = x$ for all $x \in X$.

The *generator* $(A, \mathcal{D}(A))$ of a SCLE semigroup $\{T(t)\}_{t\geq 0}$ on a locally convex space (X, τ) is the linear operator defined by

$$Ax := \lim_{t\downarrow 0} \frac{T(t)x - x}{t}$$

for x in the set

$$\mathcal{D}(A) := \left\{ x \in X \,\middle|\, \lim_{t\downarrow 0} \frac{T(t)x - x}{t} \text{ exists} \right\}.$$

We say that $(A, \mathcal{D}(A))$ is *closed* if $\{(x, Ax) \,|\, x \in \mathcal{D}(A)\}$ is closed in the product space $X \times X$ with the product topology. We say that $\mathcal{D}$ is a *core* for $(A, \mathcal{D}(A))$, if the closure of $\{(x, Ax) \,|\, x \in \mathcal{D}\}$ in the product space is $\{(x, Ax) \,|\, x \in \mathcal{D}(A)\}$. We say that the operator $(B, \mathcal{D}(B))$ is closable, if the closure of the graph of B is the graph of an operator. We will denote this operator by $(\overline{B}, \mathcal{D}(\overline{B}))$ and call $\overline{B}$ the closure of B.

As in the Banach space setting, see Lemma 2.1.3, the generator $(A, \mathcal{D}(A))$ satisfies the following well known properties. The proofs can be found for example as Propositions 1.2, 1.3 and 1.4 in Kōmura [1968].

Lemma 8.4.2. *Let (X, τ) be a locally convex space. For the generator $(A, \mathcal{D}(A))$ of a SCLE semigroup $\{T(t)\}_{t\geq 0}$, we have*

(a) $\mathcal{D}(A)$ is closed and dense in X.

(b) For $x \in \mathcal{D}(A)$, we have $T(t)x \in \mathcal{D}(A)$ for every $t \geq 0$ and $\frac{\mathrm{d}}{\mathrm{d}t}T(t)x = T(t)Ax = AT(t)x$.

(c) For $x \in X$ and $t \geq 0$, we have $\int_0^t T(s)x\mathrm{d}s \in \mathcal{D}(A)$.

(d) For $t \geq 0$, we have

$$\begin{aligned} T(t)x - x &= A\int_0^t T(s)x\mathrm{d}s && \textit{if } x \in X \\ &= \int_0^t T(s)Ax\mathrm{d}s && \textit{if } x \in \mathcal{D}(A). \end{aligned}$$

The integral in (d) should be understood as a τ Riemann integral. This is possible due to the strong continuity and the local-equi continuity of the semigroup. The proof of the next useful result follows by the obvious generalisation of the proofs of Propositions II.1.7 and II.1.8 in Engel and Nagel [2000].

Proposition 8.4.3. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t\geq 0}$ be a SCLE semigroup with generator $(A, \mathcal{D}(A))$. If $D \subset \mathcal{D}(A)$ is dense in (X, τ) and if D is invariant under the semigroup, then D is a core for $(A, \mathcal{D}(A))$. Consequentially, $\mathcal{D}(A^\infty)$ is dense in (X, τ).*

We will now introduce the resolvent of A. The notation in this section is slightly different to the notation in the rest of the Thesis. This is because the literature on semigroup theory has a different definition for the resolvent as the literature on Markov processes.

Define the *resolvent set* of $(A, \mathcal{D}(A))$ by $\rho(A) := \{\lambda \in \mathbb{C} \,|\, \mathbb{1} - \lambda A \textit{ is bijective}\}$ and for $\lambda \in \rho(A)$ the *resolvent* $\mathcal{R}(\lambda, A) = (\mathbb{1} - \lambda A)^{-1}$.

Proposition 8.4.4. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t\geq 0}$ be a SCLE semigroup with growth bound ω_0.*

(a) If $\lambda \in \mathbb{C}$ is such that the improper Riemann-integral

$$R(\lambda)x := \int_0^\infty e^{-\lambda t}T(t)x\mathrm{d}t$$

exists for every $x \in X$, then $\lambda \in \rho(A)$ and $R(\lambda, A) = R(\lambda)$.

(b) *Suppose that the semigroup is of type (M, ω). We have for every $\lambda \in \mathbb{C}$ such that $\operatorname{Re}\lambda > \omega$ and $x \in X$ that*

$$R(\lambda)x := \int_0^\infty e^{-\lambda t} T(t)x \,\mathrm{d}t$$

exists as an improper Riemann integral. Furthermore, $\lambda \in \rho(A)$.

(c) *If $\operatorname{Re}\lambda > \omega_0$, then $\lambda \in \rho(A)$.*

Proof. The proof of the first item is standard. We give the proof of (b) for completeness. Let λ be such that $\operatorname{Re}\lambda > \omega$. First, for every $a > 0$ the integral $R_a(\lambda)x := \int_0^a e^{-\lambda t} T(t)x \mathrm{d}t$ exists as a τ Riemann integral by the local equi-continuity of $\{T(t)\}_{t \geq 0}$ and the sequential completeness of (X, τ).

The sequence $n \mapsto R_n(\lambda)x$ is a τ Cauchy sequence for every $x \in X$, because for every semi-norm $p \in \mathcal{N}$ and $m > n \in \mathbb{N}$ there exists a semi-norm $q \in \mathcal{N}$ such that

$$\begin{aligned} p\left(R_m(\lambda)x - R_n(\lambda)x\right) &\leq p\left(\int_n^m e^{-t\lambda} T(t)x \mathrm{d}t\right) \\ &\leq p\left(\int_n^m e^{-t(\lambda-\omega)} e^{-\omega t} T(t)x \mathrm{d}t\right) \\ &\leq Mq(x) \int_n^m e^{-t(\operatorname{Re}\lambda - \omega)} \mathrm{d}t \\ &\leq M \|x\| \frac{e^{-\lambda m} - e^{-\lambda n}}{\operatorname{Re}\lambda - \omega}. \end{aligned}$$

Therefore, $n \mapsto R_n(\lambda)x$ converges by the sequential completeness of (X, τ). (c) follows directly from (a) and (b). ☐

We have shown that if $\operatorname{Re}\lambda > \omega_0$, then $\lambda \in \rho(A)$. We can say a lot more.

Theorem 8.4.5. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t \geq 0}$ be a SCLE semigroup of growth bound ω_0. For $\lambda > \omega_0$, $R(\lambda, A)$ is a τ continuous linear map. Furthermore, if $\{T(t)\}_{t \geq 0}$ is of type (M, ω), then there exists for every $\lambda_0 > \omega_0$ and semi-norm $p \in \mathcal{N}$ a semi-norm $q \in \mathcal{N}$ such that*

$$\sup_{\operatorname{Re}\lambda \geq \lambda_0} \sup_{n \geq 0} (\operatorname{Re}\lambda - \omega)^n p\left(\left(nR(n\lambda)\right)^n x\right) \leq Mq(x) \tag{8.4.1}$$

for every $x \in X$. If $\{T(t)\}_{t \geq 0}$ is of type $(M, \omega)^*$, then the last statement can be strengthened to

$$\sup_{\operatorname{Re}\lambda > \omega} \sup_{n \geq 0} (\operatorname{Re}\lambda - \omega)^n p\left(\left(nR(n\lambda)\right)^n x\right) \leq Mq(x).$$

The proof of the proposition directly yields the following corollary.

Corollary 8.4.6. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let I be some index set. Let $\{T_\alpha(t)\}_{t\geq 0, \alpha\in I}$ be a collection of SCLE semigroups with generators $(A_\alpha, \mathcal{D}(A_\alpha))$ that satisfy the following stability condition. Suppose there is some $M \geq 1$ such that for every $T \geq 0$ and $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that*

$$\sup_{t\leq T} e^{-\omega t} p\left(T_\alpha(t)x\right) \leq Mq\left(x\right) \tag{8.4.2}$$

for all $x \in X$. Then, we have

$$\sup_{\alpha\in I} \sup_{\operatorname{Re}\lambda\geq\lambda_0} \sup_{n\geq 0} (\operatorname{Re}\lambda - \omega)^n p\left(\left(nR(n\lambda, A_\alpha)\right)^n x\right) \leq Mq(x)$$

for every $x \in X$.

For the proof of Theorem 8.4.5, we will make use of Chernoff's bound and the probabilistic concept of stochastic domination. A short explanation and some basic results are given in Appendix 8.10.

Proof of Theorem 8.4.5. In the proof, we will write $\lceil s\rceil$ for the smallest integer $n \geq s$. Clearly, the τ continuity of $R(\lambda, A)$ follows directly from the result in equation (8.4.1), so we will start to prove (8.4.1). Without loss of generality, we can rescale and prove the result for a semigroup of type $(M, 0)$.

Let $\lambda_0 > 0$. Fix some semi-norm $p \in \mathcal{N}$. By the local equi-continuity of $\{T(t)\}_{t\geq 0}$, we can find semi-norms $q_n \in \mathcal{N}$, increasing in n, such that $\sup_{s\leq n} p(T(s)x) \leq Mq_n(x)$.

By iterating the representation of the resolvent given in Proposition 8.4.4, we see

$$\left(n\operatorname{Re}\lambda R(n\lambda)\right)^n x = \int_0^\infty \frac{(n\operatorname{Re}\lambda)^n s^{n-1}}{(n-1)!} e^{-sn\lambda} T(s)x \mathrm{d}s,$$

which implies

$$p\left(\left(n\operatorname{Re}\lambda R(n\lambda)\right)^n x\right) \leq M \int_0^\infty \frac{(n\operatorname{Re}\lambda)^n s^{n-1}}{(n-1)!} e^{-sn\operatorname{Re}\lambda} q_{\lceil s\rceil}(x) \mathrm{d}s$$

for every $x \in X$. On the right hand side, we have the semi-norm

$$q_{n,\operatorname{Re}\lambda} := \int_0^\infty \frac{(n\operatorname{Re}\lambda)^n s^{n-1}}{(n-1)!} e^{-sn\operatorname{Re}\lambda} q_{\lceil s\rceil} \mathrm{d}s$$

in $\mathcal{N}$ by the countable convexity of $\mathcal{N}$ and the fact that we integrate with respect to a probability measure. We denote this measure on $[0, \infty)$ by

$$\mu_{n,\mathrm{Re}\,\lambda}(\mathrm{d}s) = \frac{(n\mathrm{Re}\,\lambda)^n s^{n-1}}{(n-1)!} e^{-sn\mathrm{Re}\,\lambda}\mathrm{d}s,$$

and with $B_{n,\mathrm{Re}\,\lambda}$ a random variable with this distribution. As a consequence, we have the following equivalent definitions:

$$q_{n,\mathrm{Re}\,\lambda} = \int_0^\infty q_{\lceil s \rceil} \mu_{n,\mathrm{Re}\,\lambda}(\mathrm{d}s) = \mathbb{E}\left[q_{\lceil B_{n,\mathrm{Re}\,\lambda} \rceil}\right].$$

To show equi-continuity of $(n\mathrm{Re}\,\lambda)^n R(\lambda n)^n$, we need to find one seminorm $q \in \mathcal{N}$ that dominates all $q_{n,\mathrm{Re}\,\lambda}$ for $n \geq 0$ and $\mathrm{Re}\,\lambda \geq \lambda_0$. Because $s \mapsto q_{\lceil s \rceil}(x)$ is an increasing and bounded function for every $x \in X$, the result follows from Lemma 8.10.2, if we can find a random variable Y that stochastically dominates all $B_{n,\mathrm{Re}\,\lambda}$.

In other words, we need to find a random variable that dominates the tail of the distribution of all $B_{n,\mathrm{Re}\,\lambda}$. To study the tails, we use Chernoff's bound, Proposition 8.10.4.

Let $g(s, \alpha, \beta) := \frac{\beta^\alpha s^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta s}$, $s \geq 0$, $\alpha, \beta > 0$ be the density with respect to the Lebesgue measure of a *Gamma*(α, β) random variable. Thus, we see that $B_{n,\mathrm{Re}\,\lambda}$ has a *Gamma*$(n, n\mathrm{Re}\,\lambda)$ distribution. A *Gamma*$(n, n\mathrm{Re}\,\lambda)$ random variable, can be obtained as the n-fold convolution of *Gamma*$(1, n\mathrm{Re}\,\lambda)$ random variables, i.e. exponential random variables with parameter $n\mathrm{Re}\,\lambda$. Probabilistically, this means that a *Gamma*$(n, n\mathrm{Re}\,\lambda)$ can be written as the sum of n independent exponential random variables with parameter $n\mathrm{Re}\,\lambda$. An exponential random variable η that is *Exp*(β) distributed has the property that $\frac{1}{n}\eta$ is *Exp*$(n\beta)$ distributed. Therefore, we obtain that $B_{n,\mathrm{Re}\,\lambda} = \frac{1}{n}\sum_{i=1}^n X_{i,\mathrm{Re}\,\lambda}$ where $\{X_{i,\beta}\}_{i \geq 1}$ are independent copies of an *Exp*(β) random variable X_β.

This implies that we are in a position to use a Chernoff bound to control the tail probabilities of the $B_{n,\mathrm{Re}\,\lambda}$. An elementary calculation shows that for $0 < \theta < (\mathrm{Re}\,\lambda)$, we have $\mathbb{E}[e^{\theta X_{\mathrm{Re}\,\lambda}}] = \frac{\mathrm{Re}\,\lambda}{\mathrm{Re}\,\lambda - \theta}$. Evaluating the infimum in Chernoff's bound yields for $c \geq (\mathrm{Re}\,\lambda)^{-1}$ that

$$\mathbb{P}[B_{n,\mathrm{Re}\,\lambda} > c] < e^{-n(c\mathrm{Re}\,\lambda - 1 - \log c\mathrm{Re}\,\lambda)}.$$

Define the non-negative function

$$\phi : [\lambda_0^{-1}, \infty) \times [\lambda_0, \infty) \to [0, \infty)$$
$$(c, \alpha) \mapsto c\alpha - 1 - \log c\alpha$$

so that for $c \geq \lambda_0^{-1}$ and λ such that $\mathrm{Re}\,\lambda \geq \lambda_0$ we have

$$\mathbb{P}[B_{n,\mathrm{Re}\,\lambda} > c] < e^{-n\phi(c,\mathrm{Re}\,\lambda)}. \tag{8.4.3}$$

We use this result to find a random variable that stochastically dominates all $B_{n,\mathrm{Re}\,\lambda}$ for $n \in \mathbb{N}$ and $\mathrm{Re}\,\lambda \geq \lambda_0$. Define the random variable Y on $[\lambda_0^{-1}, \infty)$ by setting $\mathbb{P}[Y > c] = \exp\{-\phi(c, \lambda_0)\}$.

First note that for fixed $c \geq \lambda_0^{-1}$, the function $\alpha \mapsto \phi(c, \alpha)$ is increasing. Also note that $\phi \geq 0$. Therefore, it follows by (8.4.3) that for λ such that $\mathrm{Re}\,\lambda \geq \lambda_0$ and $c \geq \lambda_0^{-1}$, we have

$$\mathbb{P}[B_{n,\mathrm{Re}\,\lambda} > c] < e^{-n\phi(c,\mathrm{Re}\,\lambda)} \leq e^{-\phi(c,\mathrm{Re}\,\lambda)} \leq e^{-\phi(c,\lambda_0)} = \mathbb{P}[Y > c].$$

For $0 \leq c \leq \lambda_0^{-1}$, $\mathbb{P}[Y > c] = 1$ by definition, so clearly $\mathbb{P}[B_{n,\mathrm{Re}\,\lambda} > c] \leq \mathbb{P}[Y > c]$. Combining these two statements gives $Y \succeq B_{n,\mathrm{Re}\,\lambda}$ for $n \geq 1$ and λ such that $\mathrm{Re}\,\lambda \geq \lambda_0$. This implies by Lemma 8.10.2 that

$$p\left((n\mathrm{Re}\,\lambda R(n\lambda))^n x\right) \leq \mathbb{E}\left[q_{\lceil B_{n,\mathrm{Re}\,\lambda}\rceil}(x)\right] \leq \mathbb{E}\left[q_{\lceil Y\rceil}(x)\right] =: q(x).$$

By the countable convexity of $\mathcal{N}$, q is continuous and in $\mathcal{N}$, which proves the second statement of the theorem.

The strengthening to the case where the semigroup is of type $(M, \omega)^*$ is obvious, as it is sufficient to consider just one semi-norm $q \in \mathcal{N}$ for every $p \in \mathcal{N}$. □

8.5 GENERATION RESULTS

The goal of this section is to prove a Hille-Yosida result for locally equicontinuous semigroups. First, we start with a basic generation result for the semigroup generated by a continuous linear operator.

Lemma 8.5.1. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Suppose we have some τ continuous and linear operator $G : X \to X$. Then G generates a SCLE semigroup defined by*

$$S(t)x := \sum_{k\geq 0} \frac{t^k G^k x}{k!}. \tag{8.5.1}$$

Proof. First, we show that the infinite sum in equation (8.5.1) is well defined. Because τ is weaker than the norm-topology, it is sufficient to prove that the sum exists as a norm limit. By Condition C and Lemma 8.3.4, $(X, \|\cdot\|)$ is a Banach space. Therefore, we need to show for some fixed $t \geq 0$ and $x \in X$ that the sequence $y_n = \sum_{k=0}^{n} \frac{t^k G^k x}{k!}$ is Cauchy for $\|\cdot\|$. Note that as G is τ continuous, it is also norm continuous. Suppose that $n \geq m$, then we have

$$\begin{aligned} \|y_n - y_m\| &\leq \sum_{m<k\leq n} \frac{t^k}{k!} \|G\|^k \|x\| \\ &\leq \|x\| \sum_{k>m} \frac{t^k}{k!} \|G\|^k . \end{aligned}$$

which can be made arbitrarily small by choosing m large.

We proceed with showing that the τ continuous operators $S_n(t) : X \to X$ defined for $x \in X$ by $S_n(t)x := \sum_{k=0}^{n} \frac{t^k G^k x}{k!}$ are equi-continuous. As in the proof of Lemma 8.3.5, the fact that G is τ continuous implies that for every $p \in \mathcal{N}$, there exists $q \in \mathcal{N}$ such that $p(Gx) \leq \|G\| q(x)$ for all $x \in X$. Use this method to construct for a given $p \in \mathcal{N}$ an increasing sequence of semi-norms $q_n \in \mathcal{N}$, $q_0 := p$, such that $q_n(Gx) \leq \|G\| q_{n+1}(x)$ for every $n \geq 0$ and $x \in X$. As a consequence, we obtain

$$\begin{aligned} p(S_n(t)x) = p\left(\sum_{k=0}^{n} \frac{t^k G^k x}{k!}\right) &\leq \sum_{k=0}^{n} \frac{t^k}{k!} p(G^k x) \\ &\leq \sum_{k\geq 0} \frac{t^k}{k!} p(G^k x) \\ &\leq e^{t\|G\|} \sum_{k\geq 0} \frac{(\|G\| t)^k}{k!} e^{-t\|G\|} q_k(x) \\ &\leq e^{t\|G\|} q_t(x), \end{aligned}$$

where

$$q_t(x) := \sum_{k\geq 0} \frac{(\|G\| t)^k}{k!} e^{-t\|G\|} q_k(x)$$

is a continuous semi-norm in $\mathcal{N}$ by Condition C (d). The semi-norm q_t is independent of n which implies that $\{S_n(t)\}_{n\geq 1}$ is τ equi-continuous. It follows that $S(t)$ is τ continuous: pick a net x_α in X that converges to $x \in X$ with respect to τ. Let $p \in \mathcal{N}$, then

$$\begin{aligned} p\left(S(t)x_\alpha - S(t)x\right) &\\ \leq p\left(S(t)x_\alpha - S_n(t)x_\alpha\right) &+ p\left(S_n(t)x_\alpha - S_n(t)x\right) \\ &\qquad + p\left(S_n(t)x - S(t)x_\alpha\right) \\ \leq p\left(S(t)x_\alpha - S_n(t)x_\alpha\right) &+ q_t\left(x_\alpha - x\right) + p\left(S_n(t)x - S(t)x_\alpha\right). \end{aligned}$$

By first choosing α, and then n large enough, we see $p\left(S(t)x_\alpha - S(t)x\right) \to 0$.

By stochastic domination of Poisson random variables, Lemmas 8.10.2 and 8.10.3, it follows that for $t \leq T$, we have that

$$\sup_{t \leq T} e^{-t\|G\|} p(S(t)x) \leq \sup_{t \leq T} q_t(x) = q_T(x).$$

To prove strong continuity, it suffices to check that $\lim_{t \downarrow 0} S(t)x = x$ for every $x \in X$ by Lemma 8.4.1. To that end again consider $p \in \mathcal{N}$, we see

$$p(S(t)x - x) = p\left(\sum_{k \geq 0} \frac{t^k G^k x}{k!} - x\right) \leq \sum_{k \geq 0} \frac{t^k}{k!} p(G^k x - x).$$

Note that the first order term vanishes. Therefore, the Dominated convergence theorem implies that the limit is 0 as $t \downarrow 0$. □

In the proof of the Hille-Yosida theorem on Banach spaces, the semigroup is constructed as the limit of semigroups generated by the Yosida approximants. In the locally convex context, we need to take special care of equicontinuity of the approximating semigroups.

Suppose we would like to generate a locally equi-continuous semigroup e^{tA} for some operator operator $(A, \mathcal{D}(A))$.

The next lemma will yield joint local equi-continuity of the semigroups generated by the Yosida approximants by taking $H_n = nR(n, A)$. We prove a more general version, as it will also be used for the proof of Theorem 8.7.1.

We write $\lceil z \rceil$ for the smallest integer n such that $n \geq z$.

Lemma 8.5.2. *Let (X, τ) satisfy Condition C and let I be some index set. Furthermore, let $\phi : I \to \mathbb{R}^+$ be some function. Let $\{H_\alpha\}_{\alpha \in I}$ be a family of operators in $\mathcal{L}(X, \tau)$ such that for every $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that*

$$\sup_{\alpha \in I} \sup_{k \leq \lceil \phi(\alpha) \rceil} p(H_\alpha^k x) \leq q(x) \tag{8.5.2}$$

for all $x \in X$. Then, the semigroups $e^{t(\phi(\alpha)H_\alpha - \phi(\alpha))}$ are jointly locally equi-continuous.

Proof. By Lemma 8.5.1, we can define the semigroups $S_\alpha(t) := e^{t(\phi(\alpha)H_\alpha - \phi(\alpha))}$. We see that

$$S_\alpha(t)x := \sum_{k\geq 0} \frac{(t\phi(\alpha))^k H_\alpha^k x}{k!} e^{-t\phi(\alpha)},$$

which intuitively corresponds to taking the expectation of $k \mapsto H_\alpha^k x$ under the law of a Poisson random variable with parameter $t\phi(\alpha)$. We exploit this point of view, to show equi-continuity of the family $\{S_\alpha(t)\}_{t\leq T, \alpha\in I}$ for some arbitrary fixed time $T \geq 0$.

For $\mu \geq 0$, let the random variable Z_μ have a *Poisson*(μ) distribution and for $t \geq 0$ and $\alpha \in I$ let $B_{\alpha,t} := \lceil \frac{Z_{t\phi(\alpha)}}{\lceil \phi(\alpha) \rceil} \rceil$. In other words, the random variable $B_{\alpha,t}$ is obtained from $Z_{t\phi(\alpha)}$ as follows: 0 is mapped to 0, and the values $\{l\lceil\phi(\alpha)\rceil + k\}_{k=1}^{\phi(\alpha)}$ are mapped to $l+1$. Fix some α and let $n = \lceil \phi(\alpha) \rceil$. Fix a semi-norm $p \in \mathcal{N}$, and use equation (8.5.2), to construct an increasing sequence of semi-norms in $\mathcal{N}$: $q_0 = p, q_1, \dots$ such that every pair q_l, q_{l+1} satisfies the relation in (8.5.2). As a consequence, we obtain

$$\begin{aligned}
&p(S_\alpha(t)) \\
&\leq p\left(\sum_{k\geq 0} \frac{(t\phi(\alpha))^k H_\alpha^k x}{k!} e^{-t\phi(\alpha)}\right) \\
&\leq p(x)e^{-t\phi(\alpha)} + \sum_{l\geq 0}\sum_{k=1}^{n} \frac{(nt)^{nl+k}}{(nl+k)!} e^{-t\phi(\alpha)} p\left(H_\alpha^{nl+k} x\right) \\
&\leq q_0(x)e^{-t\phi(\alpha)} + \sum_{l\geq 0}\sum_{k=1}^{n} \frac{(nt)^{nl+k}}{(nl+k)!} e^{-t\phi(\alpha)} q_{l+1}(x) \\
&= \mathbb{P}[B_{\alpha,t} = 0] q_0(x) + \sum_{l\geq 0} \mathbb{P}[B_{\alpha,t} = l+1] q_{l+1}(x) \\
&= \mathbb{E}\left[q_{B_{\alpha,t}}(x)\right].
\end{aligned} \tag{8.5.3}$$

We see that, as in the proof of the second property in Theorem 8.4.5, we are done if we can find a random variable Y that stochastically dominates all $B_{\alpha,t}$ for $\alpha \in I$ and $t \leq T$.

Again for some fixed α, we calculate the tail probabilities of $B_{\alpha,t}$ in the case that $t > 0$. If $t = 0$, all tail probabilities are 0. Recall that we write $n = \lceil \phi(\alpha) \rceil$. By definition and Lemma 8.10.3,

$$\mathbb{P}[B_{\alpha,t} > k] = \mathbb{P}\left[\frac{1}{n} Z_{t\phi(\alpha)} > k\right] \leq \mathbb{P}\left[\frac{1}{n} Z_{nt} > k\right].$$

Because Z_{nt} is *Poisson*(nt) distributed, we can write it as $Z_{nt} = \sum_{i=1}^{n} X_i$ where $\{X_i\}_{i \geq 0}$ are independent and *Poisson*(t) distributed. This implies that we can apply Chernoff's bound to $\frac{1}{n} Z_{nt}$, see Proposition 8.10.4. First of all, for all $\theta \in \mathbb{R}$, we have $\mathbb{E}\left[e^{\theta X}\right] = \exp\{t(e^{\theta} - 1)\}$. Evaluating the infimum in Chernoff's bound for $k \geq \lceil T \rceil$, $T \geq t$ yields

$$\mathbb{P}[B_{\alpha,t} > k] \leq \mathbb{P}\left[\frac{1}{n} Z_{nt} > k\right] < e^{-n\left(k \log \frac{k}{t} - k + t\right)}.$$

Define the function

$$\begin{aligned} \phi &: [\lceil T \rceil, \infty) \times (0, T] \to [0, \infty) \\ (a, b) &\mapsto a \log \frac{a}{b} - a + b, \end{aligned}$$

so that for $k \geq \lceil T \rceil$, $T \geq t$, we have $\mathbb{P}[B_{\alpha,t} > k] < e^{-n\phi(k,t)}$.

We define a new random variable Y on $\{n \in \mathbb{N} \,|\, n \geq \lceil T \rceil\}$ by putting $\mathbb{P}[Y = \lceil T \rceil] = 1 - e^{-\phi(\lceil T \rceil, T)}$, and for $k \geq \lceil T \rceil$: $\mathbb{P}[Y > k] = e^{-\phi(k,T)}$, or stated equivalently $\mathbb{P}[Y = k+1] = e^{-\phi(k,T)} - e^{-\phi(k+1,T)}$. In other words, we construct Y so that the tail variables agree with $e^{-\phi(k,T)}$.

For $k < \lceil T \rceil$, we have by definition that $\mathbb{P}[Y > k] \geq \mathbb{P}[B_{\alpha,t} > k]$ as the probability on the left is 1. For $k \geq \lceil T \rceil$, an elementary computation shows that for $t \leq T$ the function $\phi_k(t) := \phi(k,t)$ is decreasing in t. This implies that

$$\mathbb{P}[B_{\alpha,t} > k] \leq e^{-n\phi(k,t)} \leq e^{-\phi(k,t)} \leq e^{-\phi(k,T)} = \mathbb{P}[Y > k].$$

In other words, as α was arbitrary, we see $Y \succeq B_{\alpha,t}$ for all $\alpha \in I$ and $0 < t \leq T$. For the remaining cases, where $t = 0$, the result is clear as $B_{\alpha,t} = 0$ with probability 1. By Lemma 8.10.2 and equation (8.5.3), we obtain that

$$p(S_\alpha(t)) \leq \mathbb{E}\left[q_{B_{\alpha,t}}(x)\right] \leq \mathbb{E}\left[q_Y(x)\right] =: q(x).$$

For the second inequality, we use that Y stochastically dominates $X_{\alpha,t}$ for all $\alpha \in I$ and $t \leq T$. The semi-norm $q(x)$ is in $\mathcal{N}$ by the countable convexity of $\mathcal{N}$. We conclude that the family $\{S_\alpha(t)\}_{t \leq T, \alpha \in I}$ is equicontinuous. □

Lemma 8.5.3. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $(A, \mathcal{D}(A))$ be a closed, densely defined operator such that there exists an $\omega \in \mathbb{R}$ such that $(\omega, \infty) \subseteq \rho(A)$ and such that for every $\lambda_0 > \omega$ and semi-norm $p \in \mathcal{N}$, there is a continuous semi-norm q such that $\sup_{\lambda \geq \lambda_0} p((\lambda - \omega)R(\lambda)x) \leq q(x)$ for every $x \in X$. As $\lambda \to \infty$, we have*

(a) $\lambda R(\lambda)x \to x$ for every $x \in X$,

(b) $\lambda AR(\lambda)x = \lambda R(\lambda)Ax \to Ax$ for every $x \in \mathcal{D}(A)$.

The lemma can be proven as in the Banach space case [Engel and Nagel, 2000, Lemma II.3.4]. We have now developed enough machinery to prove a Hille-Yosida type theorem which resembles the equivalence between (a) and (b) of Theorem 16 in Kühnemund [2003].

Theorem 8.5.4. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. For a linear operator $(A, \mathcal{D}(A))$ on (X, τ), the following are equivalent.*

(a) $(A, \mathcal{D}(A))$ generates a SCLE semigroup of type (M, ω).

(b) $(A, \mathcal{D}(A))$ is closed, densely defined and there exists $\omega \in \mathbb{R}$ and $M \geq 1$ such that for every $\lambda > \omega$ one has $\lambda \in \rho(A)$ and for every semi-norm $p \in \mathcal{N}$ and $\lambda_0 > \omega$ there exists a semi-norm $q \in \mathcal{N}$ such that for all $x \in X$ one has

$$\sup_{n \geq 1} \sup_{\lambda \geq \lambda_0} p\left((n(\lambda - \omega)R(n\lambda))^n x\right) \leq Mq(x). \tag{8.5.4}$$

(c) $(A, \mathcal{D}(A))$ is closed, densely defined and there exists $\omega \in \mathbb{R}$ and $M \geq 1$ such that for every $\lambda \in \mathbb{C}$ satisfying $\operatorname{Re} \lambda > \omega$, one has $\lambda \in \rho(A)$ and for every semi-norm $p \in \mathcal{N}$ and $\lambda_0 > \omega$ there exists a semi-norm $q \in \mathcal{N}$ such that for all $x \in X$ and $n \in \mathbb{N}$

$$\sup_{n \geq 1} \sup_{\operatorname{Re} \lambda \geq \lambda_0} p\left((n(\operatorname{Re} \lambda - \omega)R(n\lambda))^n x\right) \leq Mq(x).$$

Remark 8.5.5. We will refer to the estimate in equation (8.5.4) as the *Hille-Yosida estimate.*

By a simplification of the arguments, or arguing as in Section IX.7 in Yosida [1978], we can also give a necessary and sufficient condition for the generation of a quasi equi-continuous semigroup of type $(M, \omega)^*$, which corresponds with the result obtained in Theorem 8.4.5. Theorem 3.5 in Kunze [2009] states a similar result.

Suppose we have a semigroup of type (M, ω) and let $\omega' > \omega$. Equation (8.5.4) yields

$$\sup_{n \geq 1} \sup_{\lambda > \omega'} p\left(\left(n(\lambda - \omega')R(n\lambda)\right)^n x\right) \leq Mq(x) \tag{8.5.5}$$

which implies that the semigroup is of type $(M, \omega')^*$. We state this as a corollary.

Corollary 8.5.6. *Suppose that $(X, \tau, \|\cdot\|)$ satisfies Condition C. If a semigroup is of type (M, ω), then it is of type $(M, \omega')^*$ for all $\omega' > \omega$.*

As Equation (8.5.4) implies (8.5.5), it is sufficient, for the construction of a semigroup, to use the weaker result as in Kunze [2009]. However, one obtains that the semigroup is of type (M, ω') for $\omega' > \omega$, which does not give any control if the semigroup is rescaled by $e^{-\omega t}$. A semigroup that is of type (M', ω') for all $\omega' > \omega$ is not necessarily of type (M, ω) for any $M \geq M'$ as is shown in Example I.5.7(ii) in Engel and Nagel [2000].

The proof of the Hille-Yosida theorem stated here, however, gives explicit control on the semigroup rescaled by $e^{-\omega t}$ via the construction in Lemma 8.5.2 and gives a result as strong as the equivalence of (a) and (b) of Theorem 16 in Kühnemund [2003].

Proof of Theorem 8.5.4. (a) to (c) is the content of Theorem 8.4.5 and (c) to (b) is clear. So we need to prove (b) to (a).

First note that we can always assume that $\omega = 0$ by a suitable rescaling. We start by proving the result for $\omega = 0$ and $M = 1$. We follow the lines of the proof of the Hille-Yosida theorem for Banach spaces in [Engel and Nagel, 2000, Theorem II.3.5].

Define for every $n \in \mathbb{N} \setminus \{0\}$ the Yosida approximants

$$A_n := nAR(n) = n^2 R(n) - n\mathbb{1}.$$

These operators commute and for every n A_n satisfies the condition in Lemma 8.5.1. Furthermore, we can apply Lemma 8.5.2 to the operators $H_n = nR(n)$. Note that Equation (8.5.2) is satisfied as a consequence of Equation (8.5.4), as the latter implies

$$\sup_k \sup_{\lambda \in \{\frac{n}{k} \,|\, n \geq k\}} p\left((\lambda k R(\lambda k))^k x\right) \leq q(x)$$

for all x, which in turn can be rewritten to

$$\sup_n \sup_{k \leq n} p\left((nR(n))^k x\right) \leq q(x)$$

for all $x \in X$. Hence, we obtain that the operators A_n generate jointly locally equi-continuous strongly continuous commuting semigroups $t \mapsto T_n(t)$ of type $(1, 0)$. We show that there exists a limiting semigroup.

Let $x \in \mathcal{D}(A)$ and $t \geq 0$, the fundamental theorem of calculus applied to $s \mapsto T_m(t-s)T_n(s)x$ for $s \leq t$, yields

$$\begin{aligned} T_n(t)x - T_m(t)x &= \int_0^t T_m(t-s)\left(A_n - A_m\right)T_n(s)x\mathrm{d}s \\ &= \int_0^t T_m(t-s)T_n(s)\left(A_nx - A_mx\right)\mathrm{d}s. \end{aligned}$$

By Lemma 8.5.2, we obtain that for every semi-norm $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that

$$p(T_n(t)x - T_m(t)x) \leq tq(A_nx - A_mx). \tag{8.5.6}$$

Hence, for $x \in \mathcal{D}(A)$ the sequence $n \mapsto T_n(s)x$ is τ-Cauchy uniformly for $s \leq t$ by Lemma 8.5.3 (b). The joint local equi-continuity of $\{T_n(t)\}_{t\geq 0, n\geq 1}$ implies that this property extends to all $x \in X$.

Define the point-wise limit of this sequence by $T(s)x := \lim_n T_n(s)x$. This directly yields that the family $\{T(s)\}_{s\leq t}$ is equi-continuous, because it is contained in the closure of an equi-continuous set of operators, Proposition 32.4 in Treves [1967]. Consequently, this shows that $\{T(t)\}_{t\geq 0}$ is a locally equi-continuous set of operators of type $(1,0)$.

The family of operators $\{T(t)\}_{t\geq 0}$ is a semigroup, because it is the point-wise limit of the semigroups $\{T_n(t)\}_{t\geq 0}$. We show that it is strongly continuous by using Lemma 8.4.1. Let $p \in \mathcal{N}$ and $x \in \mathcal{D}(A)$, then for every n:

$$p(T(t)x - x) \leq p(T(t)x - T_n(t)x) + p(T_n(t)x - x).$$

Because $p(T(t)x - T_n(t)x) \to 0$, uniformly for $t \leq 1$, we can first choose n large to make the first term on the right hand side small, and then t small, to make the second term on the right hand side small.

We still need to prove that the semigroup $\{T(t)\}_{t\geq 0}$ has generator $(A, \mathcal{D}(A))$. Denote with $(B, \mathcal{D}(B))$ the generator of $\{T(t)\}_{t\geq 0}$. For $x \in \mathcal{D}(A)$, we have for a continuous semi-norm p that

$$\begin{aligned} p\left(\frac{T(t)x - x}{t} - Ax\right) \leq p&\left(\frac{T(t)x - T_n(t)x}{t}\right) \\ &+ p\left(\frac{T_n(t)x - x}{t} - A_nx\right) + p(A_nx - Ax), \end{aligned}$$

for some continuous semi-norm q. By repeating the argument that led to (8.5.6), we can rewrite the first term on the second line to obtain

$$p\left(\frac{T(t)x - x}{t} - Ax\right)$$
$$\leq q\left(Ax - A_n x\right) + p\left(\frac{T_n(t)x - x}{t} - A_n x\right) + p(A_n x - Ax).$$

By first choosing n large and then t small, we see that $x \in \mathcal{D}(B)$ and $Bx = Ax$. In other words, $(B, \mathcal{D}(B))$ extends $(A, \mathcal{D}(A))$.

For $\lambda > 0$, we know that $\lambda \in \rho(A)$, so $\lambda - A : \mathcal{D}(A) \to X$ is bijective. Because B generates a semigroup of type $(1, 0)$, we also have that $\lambda - B : \mathcal{D}(B) \to X$ is bijective. But B extends A, which implies that $(A, \mathcal{D}(A)) = (B, \mathcal{D}(B))$.

We extend the result for general $M \geq 1$. The strategy is to define a norm on X that is equivalent to $\|\cdot\|$ for which the semigroup that we want to construct is $(1, 0)$ bounded. Equations (8.3.1) and (8.5.4) imply that $\|\mu^n R(\mu)^n\| \leq M$. Define

$$\|x\|_\mu := \sup_{n \geq 0} \|\mu^n R(\mu)^n x\|$$

and then define $|||x||| := \sup_{\mu > 0} \|x\|_\mu$. This norm has the property that $\|x\| \leq |||x||| \leq M\|x\|$ and $|||\lambda R(\lambda)||| \leq 1$ for every $\lambda > 0$, see the proof of Theorem II.3.8 in Engel and Nagel [2000]. Use this norm to define a new set of continuous semi-norms as in Definition 8.3.1 by

$$\mathcal{N}^* := \{p \,|\, p \textit{ is a } \tau \textit{ continuous semi-norm such that } p(\cdot) \leq |||\cdot|||\}.$$

As a consequence of $|||\lambda R(\lambda)||| \leq 1$ and the τ continuity of $\lambda R(\lambda)$, we obtain that for every $p \in \mathcal{N}^*$ there exists $q \in \mathcal{N}^*$ such that $p(\lambda R(\lambda)x) \leq q(x)$ for every $x \in X$. Likewise, we obtain for every $\lambda_0 > 0$ that for every $p \in \mathcal{N}^*$ there exists $q \in \mathcal{N}^*$ such that

$$\sup_{\lambda \geq \lambda_0} \sup_{n \geq 1} p\left((n\lambda R(n\lambda))^n x\right) \leq q(x).$$

This means that we can use the first part of the proof to construct a SCLE semigroup $\{T(t)\}_{t \geq 0}$ that has bound $(1, 0)$ with respect to $\mathcal{N}^*$.

Let $T \geq 0$. Pick a semi-norm $p \in \mathcal{N}$. It follows that $p \in \mathcal{N}^*$, so there exists a $q \in \mathcal{N}^*$ such that $\sup_{t \leq T} p(T(t)x) \leq q(x)$ for all $x \in X$.

Because $|||\cdot||| \leq M \|\cdot\|$, it follows that $\mathcal{N}^*$ is a subset of $M\mathcal{N}$ which implies that $\hat{q} := \frac{1}{M} q \in \mathcal{N}$. We obtain $\sup_{t \leq T} p(T(t)x) \leq M\hat{q}(x)$ for all $x \in X$. In other words, A generates a SCLE semigroup $\{T(t)\}_{t \geq 0}$ of type $(M, 0)$. □

8.6 APPROXIMATION RESULTS

Our next goal is to prove the Trotter-Kato-Approximation theorems. We follow the approach taken in Engel and Nagel [2000], and specify where the methods need to be adapted to the setting of locally convex setting. We start by introducing pseudo-resolvents and state a number of well known results.

8.6.1 *Pseudo-resolvents*

Let $\Lambda \subset \mathbb{C}$, and consider operators $J(\lambda) \in \mathcal{L}(X, \tau)$ for $\lambda \in \Lambda$. The family $\{J(\lambda)\}_{\lambda \in \Lambda}$ is called a *pseudo-resolvent* if it satisfies the *resolvent equation*

$$J(\lambda) - J(\mu) = (\mu - \lambda) J(\lambda) J(\mu) \tag{8.6.1}$$

for $\lambda, \mu \in \Lambda$. We give a number of basic results for pseudo-resolvents.

Proposition 8.6.1. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T_n(t)\}_{t \geq 0, n \in \mathbb{N}}$ be SCLE semigroups with generators $(A_n, \mathcal{D}(A_n))$. Suppose the following stability condition is satisfied. For every $T \geq 0$ and $p \in \mathcal{N}$ there exists $q \in \mathcal{N}$ such that*

$$\sup_{t \leq T} p(T_n(t)x) \leq q(x) \qquad \text{for all } x \in X.$$

Assume that for some $\lambda_0 > 0$ the limit $\lim_{n \to \infty} R(\lambda_0, A_n)x$ exists for every $x \in X$. Then, the limit $R(\lambda)x := \lim_n R(\lambda, A_n)x$ exists for all $\lambda > 0$ and defines a pseudo-resolvent that satisfies the Hille-Yosida estimate in equation (8.5.4) *with with $M = 1$ and $\omega = 0$.*

Proof. The existence of the limit $R(\lambda)x := \lim_n R(\lambda, A_n)x$ for all λ such that $\operatorname{Re} \lambda > 0$ follows as in the proof of Proposition III.4.4 in Engel and Nagel [2000]. The result of Corollary 8.4.6 and the convergence of $R(\lambda, A_n)x$ to $R(\lambda)$ yields directly that $R(\lambda)$ satisfies the Hille-Yosida estimate, equation (8.5.4). □

The next three results also follow as in the Banach space case and do not need the assumption that $(X, \tau, \|\cdot\|)$ satisfies Condition C. For proofs see Section III.4.a in Engel and Nagel [2000].

Lemma 8.6.2. *Let $\{J(\lambda) \mid \lambda \in \Lambda\}$ be a pseudo-resolvent on (X, τ). For $\lambda, \mu \in \Lambda$, we have $J(\lambda)J(\mu) = J(\mu)J(\lambda)$, $\ker J(\lambda) = \ker J(\mu)$, and $\operatorname{rg} J(\lambda) = \operatorname{rg} J(\mu)$.*

Proposition 8.6.3. *Let $\{J(\lambda) \mid \lambda \in \Lambda\}$ be a pseudo-resolvent. The following are equivalent:*

(a) $\ker J(\lambda_0) = \{0\}$ *and* $\overline{\operatorname{rg} J(\lambda_0)} = X$ *for some* $\lambda_0 \in \Lambda$,

(b) $\ker J(\lambda) = \{0\}$ *and* $\overline{\operatorname{rg} J(\lambda)} = X$ *for all* $\lambda \in \Lambda$,

(c) there exists a closed densely defined operator $(A, \mathcal{D}(A))$ such that $\Lambda \subset \rho_\tau(A)$ and $J(\lambda) = R(\lambda, A)$.

The next corollary follows as Corollary III.4.7 in Engel and Nagel [2000].

Corollary 8.6.4. *Let $\{J(\lambda) \mid \lambda \in \Lambda\}$ be a pseudo-resolvent. If there is an unbounded sequence $\{\lambda_n\}_{n\geq 1} \subset \Lambda$ such that $\lim_n \lambda_n J(\lambda_n)x = x$ for all $x \in X$, then $\{J(\lambda) \mid \lambda \in \Lambda\}$ is the resolvent of a closed densely defined operator $(A, \mathcal{D}(A))$.*

In particular, this holds if the range of $J(\lambda)$ is dense and if there is some $M \geq 1$ such that for every $p \in \mathcal{N}$ there is $q \in \mathcal{N}$ such that

$$\sup_{n\geq 1} p\left(\lambda_n J(\lambda_n)x\right) \leq M q\left(x\right)$$

for all $x \in X$.

8.6.2 *The Trotter-Kato theorems*

For the proof of the first Trotter-Kato theorem, we will need the following result

Lemma 8.6.5. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t\geq 0}$ and $\{S(t)\}_{t\geq 0}$ be SCLE semigroups on (X, τ) with generators $(A, \mathcal{D}(A))$ and $(B, \mathcal{D}(B))$ with growth bounds ω_A, ω_B. Suppose that $\mathcal{D}(A) \cap \mathcal{D}(B)$ is dense in X. For $x \in X$ and $\lambda > \omega_A \vee \omega_B$, we have*

$$R(\lambda, A)\left[T(t) - S(t)\right] R(\lambda, B)x = \int_0^t T(s)\left[R(\lambda, A) - R(\lambda, B)\right] S(t-s)x\,\mathrm{d}s.$$

Note that the lemma can be proven without the assumption that $\mathcal{D}(A) \cap \mathcal{D}(B)$ is dense in X as in Lemma 3.3 in Albanese and Mangino [2004]. The next proof under this slightly stronger assumption is faster.

Proof. For $x \in \mathcal{D}(A) \cap \mathcal{D}(B)$, we have

$$\begin{aligned}
&R(\lambda, A)\,[T(t) - S(t)]\,R(\lambda, B)x \\
&= \int_0^\infty \int_0^\infty e^{-\lambda(a+b)}\,[T(a+t)S(b) - T(a)S(b+t)]\,x\mathrm{d}a\mathrm{d}b \\
&= \int_0^t \int_0^\infty \int_0^\infty e^{-\lambda(a+b)} T(a+s)\,[A - B]\,S(b+t-s)x\mathrm{d}a\mathrm{d}b\mathrm{d}s \\
&= \int_0^t \int_0^\infty \int_0^\infty e^{-\lambda(a+b)} T(a+s)\,[\lambda - B]\,S(b+t-s)x\mathrm{d}a\mathrm{d}b\mathrm{d}s \\
&\quad - \int_0^t \int_0^\infty \int_0^\infty e^{-\lambda(a+b)} T(a+s)\,[\lambda - A]\,S(b+t-s)x\mathrm{d}a\mathrm{d}b\mathrm{d}s \\
&= \int_0^t T(s)\,[R(\lambda, A) - R(\lambda, B)]\,S(t-s)x\mathrm{d}s.
\end{aligned}$$

□

Theorem 8.6.6 (First Trotter-Kato Theorem)**.** *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t\geq 0}$ and $\{T_n(t)\}_{t\geq 0}$, $n \geq 1$ be SCLE semigroups on (X, τ) with generators $(A, \mathcal{D}(A))$ and $(A_n, \mathcal{D}(A_n))$, and assume that there exist $M \geq 1$, $\omega \in \mathbb{R}$ such that for every $p \in \mathcal{N}$ and $T \geq 0$, there exists $q \in \mathcal{N}$ such that*

$$\sup_{t\leq T} e^{-\omega t}\left[p(T(t)x) \vee \sup_n p(T_n(t)x)\right] \leq Mq(x)$$

for all $x \in X$. Let D be a core for $(A, \mathcal{D}(A))$. Consider the following statements

(a) $D \subset \mathcal{D}(A_n)$ for all $n \geq 1$ and $A_n x \to Ax$ for all $x \in D$.

(b) For each $x \in D$ there exists $x_n \in \mathcal{D}(A_n)$ such that $x_n \to x$ and $Ax_n \to Ax$.

(c) $R(\lambda, A_n)x \to R(\lambda, A)x$ for all $x \in X$ some $\lambda > \omega$.

(d) $T_n(t)x \to T(t)x$ for all $x \in X$ uniformly for t in compact intervals.

The implications (a) ⇒ (b) ⇔ (c) ⇔ (d) hold.

Proof. Without loss of generality, we rescale and assume that $\omega = 0$.

The implication (a) to (b) is clear. The proof from (b) to (c) follows as in the Banach space case, the proof of Theorem III.4.8 in Engel and Nagel [2000], as the family $\{R(\lambda, A_n)\}_{n\geq 0} \cup \{R(\lambda, A)\}$ is equi-continuous by Corollary 8.4.6. Also the implication from (c) to (b) follows as in the Banach space case. For the proof of (c) to (d), the Banach space proof can be adapted by simply replacing the norm by semi-norms from $\mathcal{N}$.

For the proof from (c) to (d) we follow the strategy of Albanese and Mangino [2004]. Fix some $T \geq 0$. Let $x \in \mathcal{D}(A^2)$. Set $x_1 = (\lambda - A)x$ and $x_2 = (\lambda - A)x_1 = (\lambda - A)^2 x$. We obtain

$$\begin{aligned} p\left(T_n(t)x - T(t)x\right) \leq\ & p\left(T_n(t)\left(R(\lambda, A) - R(\lambda, A_n)\right)x_1\right) \\ & + p\left(R(\lambda, A_n)\left(T_n(t) - T(t)\right)x_1\right) \\ & + p\left(\left(R(\lambda, A_n) - R(\lambda, A)\right)T(t)x_1\right). \end{aligned} \tag{8.6.2}$$

The first term converges to 0 uniformly in $t \leq T$ as n goes to infinity by the uniform local equi-continuity of the semigroups $\{T_n\}_{n\geq 0}$.

To show that the third term of equation (8.6.2) converges to 0 uniformly for $t \leq T$, first note that $T_n(t)$ and $T(t)$ are jointly locally equi-continuous, which implies by Corollary 8.4.6 that the family $\{R(\lambda, A_n)\}_{n\geq 0} \cup \{R(\lambda, A)\}$ is equi-continuous. Let q be the semi-norm in $\mathcal{N}$ such that $\sup_n p\left(R(\lambda, A_n)y\right) \vee p\left(R(\lambda, A)y\right) \leq q(y)$ for all $y \in X$.

Fix $\varepsilon > 0$. Because $[0, T]$ is compact and $t \mapsto T(t)x_1$ is τ continuous, the set $K = \{T(t)x_1 \,|\, t \in [0, T]\}$ is τ compact. This compactness implies that we can find $y_1, \ldots, y_k \in K$ such that $K \subset \cup_{i\leq k}\{y \,|\, q(y - y_i) < \varepsilon\}$.

Pick $y \in K$. By the construction above, there is some i such that $q(y - y_i) \leq \varepsilon$. Let N be large enough such that for $n \geq N$, we have $p\left(R(\lambda, A_n)y_i - R(\lambda, A)y_i\right) \leq \varepsilon$. Therefore, for $n \geq N$,

$$\begin{aligned} p\left(R(\lambda, A_n)y - R(\lambda, A)y\right) &\leq p\left(R(\lambda, A_n)y - R(\lambda, A_n)y_i\right) \\ &\quad + p\left(R(\lambda, A_n)y_i - R(\lambda, A)y_i\right) \\ &\quad + p\left(R(\lambda, A)y_i - R(\lambda, A)y\right) \\ &\leq 2q\left(x_i - y\right) + \varepsilon \\ &\leq 3\varepsilon. \end{aligned}$$

The choice of $\varepsilon > 0$ was arbitrary, which implies that the third term of equation (8.6.2) converges to 0.

By Lemma 8.6.5 and the joint local equi-continuity of $\{T_n(t)\}_{t\geq 0, n\in\mathbb{N}}$, we get the following bound on the second term of equation (8.6.2):

$$\begin{aligned}
&p\left(R(\lambda, A_n)\left(T_n(t) - T(t)\right)x_1\right) \\
&\quad \leq \int_0^t p\left(T_n(s)\left[R(\lambda, A_n) - R(\lambda, A)\right]T(t-s)x_2\right)\mathrm{d}s \\
&\quad \leq \int_0^t q\left(\left[R(\lambda, A_n) - R(\lambda, A)\right]T(t-s)x_2\right)\mathrm{d}s.
\end{aligned}$$

As in the argument for the third term, the integrand converges to 0 uniformly. This implies that the second term of equation (8.6.2) converges to 0 uniformly for $t \leq T$.

We conclude that $p\left(T_n(t)x - T(t)x\right) \to 0$ for $x \in \mathcal{D}(A^2)$. $\mathcal{D}(A^2)$ is dense in (X, τ) by Proposition 8.4.3, which together with the joint local equi-continuity of $\{T_n\}_{n\geq 0}$ and T, yields the final result. □

Theorem 8.6.7 (Second Trotter-Kato Theorem)**.** *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T_n(t)\}_{t\geq 0}$, $n \geq 1$ be SCLE semigroups on (X, τ) with generators $(A_n, \mathcal{D}(A_n))$, and assume that there exist $M \geq 1$, $\omega \in \mathbb{R}$ such that for every $p \in \mathcal{N}$ and $T \geq 0$, there exists $q \in \mathcal{N}$ such that*

$$\sup_{t\leq T}\sup_{n} e^{-\omega t} p(T_n(t)x) \leq Mq(x) \tag{8.6.3}$$

for all $x \in X$. For some $\lambda_0 > \omega$ consider the following statements.

(a) There exits a densely defined operator $(A, \mathcal{D}(A))$ such that $A_n x \to Ax$ for all x in a core D of $(A, \mathcal{D}(A))$ and such that the range $rg(\lambda_0 - A)$ is dense in X.

(b) The operators $R(\lambda_0, A_n)$, $n \geq 1$ converge strongly to a continuous operator R with dense range.

(c) The semigroups $\{T_n(t)\}_{t\geq 0}$, $n \geq 1$ converge strongly, and uniformly for t in compact intervals to a SCLE semigroup $\{T(t)\}_{t\geq 0}$, with generator B such that $R = R(\lambda_0, B)$.

The implications (a) ⇒ (b) ⇔ (c) hold. In particular, if (a) holds, then $B = \overline{A}$.

Proof. We prove the theorem for $\omega = 0$.

The proof of (a) to (b) follows as in the Banach space case, see the proof of Theorem III.4.9 in Engel and Nagel [2000], by using that the family $\{R(\lambda_0, A_n)\}_{n\geq 0}$ is equi-continuous by Corollary 8.4.6. The implication (c) to (b) follows by Theorem 8.6.6. We prove (b) to (c). Proposition 8.6.1 gives us that

$$R(\lambda)x := \lim_n R(\lambda, A_n)x, \qquad \text{for } \lambda > 0$$

is a pseudo-resolvent for which the Hille-Yosida estimate holds. By the assumption on the range of R and Lemma 8.6.2, $R(\lambda)$ has dense range for all $\lambda > 0$. As a consequence, Corollary 8.6.4 shows there is a closed operator $(B, \mathcal{D}(B))$ such that $R(\lambda) = R(\lambda, B)$. The Hille-Yosida theorem, Theorem 8.5.4 implies that $(B, \mathcal{D}(B))$ generates a semigroup $\{T(t)\}_{t\geq 0}$. The final result now follows from the first Trotter-Kato theorem.

The proof that $\overline{A} = B$ follows again as in the Banach space case. □

8.7 CONSEQUENCES OF THE TROTTER-KATO THEOREM

Theorem 8.7.1 (Chernoff product formula)**.** *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Consider a function $V : \mathbb{R}^+ \to \mathcal{L}(X, \tau)$ satisfying $V(0) = I$. Suppose that there exists $M \geq 1$ and $\omega \in \mathbb{R}$ such that for every $p \in \mathcal{N}$ and $T \geq 0$, there exists $q \in \mathcal{N}$ such that*

$$\sup_{t\leq T}\sup_{m\geq 1} e^{-tm\omega} p\left(V\left(\frac{t}{m}\right)^m x\right) \leq Mq(x) \tag{8.7.1}$$

for all $x \in X$. Assume that $Ax := \lim_{t\downarrow 0}\frac{V(t)x - x}{t}$ exists for all $x \in D$, where D and $(\lambda_0 - A)D$ are dense in X for some $\lambda_0 > \omega$. Then the closure $\overline{A}$ of A generates a SCLE semigroup $\{T(t)\}_{t\geq 0}$ of type (M, ω), which is obtained by $T(t)x = \lim_n V\left(\frac{t}{n}\right)^n x$ for all $x \in X$ and uniformly for t in compact intervals.

If the supremum over $t \leq T$ in equation (8.7.1) *can be extended to a supremum over $t \geq 0$, then $\{T(t)\}_{t\geq 0}$ is of type $(M, \omega)^*$.*

Proof. Clearly, the theorem follows by rescaling by the result for $\omega = 0$. We start with the case $M = 1$ and use this to obtain the general result afterwards. For $s > 0$ define the continuous operators

$$A_s := \frac{V(s) - I}{s}$$

so that $A_s x \to Ax$ as $s \to 0$ for $x \in D$. Every A_s generates a SCLE semigroup $\{T_s(t)\}_{t\geq 0}$ by Lemma 8.5.1. We show that these semigroups satisfy the stability condition, equation (8.6.3), of Theorem 8.6.7, for which we use Lemma 8.5.2. Set $I = (0,1]$, $H_s = V(s)$ and $\phi(s) = \frac{1}{s}$.

Equation (8.7.1), for $T = 2$, implies for $M = 1$ and $\omega = 0$ that

$$\sup_{s\le 2}\sup_{m\le \frac{2}{s}} p\left(V(s)^m x\right) \le q(x),$$

by making the substitution $s = \frac{t}{m}$. This yields

$$\sup_{s\le 1}\sup_{m\le\lceil\phi(s)\rceil} p\left(V(s)^m x\right) \le q(x).$$

Therefore, Lemma 8.5.2 yields the joint equi-continuity of the semigroups $\{e^{tA_s}\}_{t\ge 0}$ for $s \le 1$.

The second Trotter-Kato theorem, Theorem 8.6.7, gives a semigroup $\{T(t)\}_{t\ge 0}$ with generator $\overline{A}$. Furthermore, we have for every $p \in \mathcal{N}, T \ge 0$ and $x \in X$ that

$$\lim_{s\downarrow 0}\sup_{t\le T} p\left(T(t)x - e^{tA_s}x\right) = 0,$$

which directly implies

$$\lim_{n\to\infty}\sup_{t\le T} p\left(T(t)x - e^{tA_{\frac{t}{n}}}x\right) = 0 \tag{8.7.2}$$

for every $p \in \mathcal{N}, T \ge 0$ and $x \in X$. On the other hand, we have for $p \in \mathcal{N}$ that

$$\begin{aligned} p\left(e^{tA_{\frac{t}{n}}}x - \left[V\left(\frac{t}{n}\right)\right]^n x\right) &\le \left\| e^{tA_{\frac{t}{n}}}x - \left[V\left(\frac{t}{n}\right)\right]^n x\right\| \\ &= \left\| e^{n\left(V\left(\frac{t}{n}\right)-\mathbb{1}\right)}x - \left[V\left(\frac{n}{t}\right)\right]^n x\right\| \\ &\le \sqrt{n}\left\| V\left(\frac{t}{n}\right)x - x\right\| \\ &= \frac{1}{\sqrt{n}}\left\| A_{\frac{t}{n}}x\right\|, \end{aligned}$$

where we have used Lemma III.5.1 from Engel and Nagel [2000] in line three. The sequence $\left\| A_{\frac{t}{n}}x\right\|$ is bounded, so we obtain

$$p\left(e^{tA_{\frac{t}{n}}}x - \left[V\left(\frac{t}{n}\right)\right]^n x\right) \to 0$$

for every $x \in D$, uniformly for $0 < t \leq T$. Because $V(0) = \mathbb{1}$, this extends to $t = 0$. Together with equation (8.7.2), this yields for every $p \in \mathcal{N}$ and $x \in D$ that

$$\sup_{t \leq T} p\left(T(t)x - \left[V\left(\frac{t}{n}\right)\right]^n x\right) \to 0.$$

The local equi-continuity of $\{T(t)\}_{t\geq 0}$ and equation (8.7.1) extend the result to all $x \in X$.

We now extend the result to the case where $M > 1$. By (8.3.1) on page 243, Equation (8.7.1) implies that

$$\sup_{m\geq 1}\sup_{t\geq 0}\left\|V\left(\frac{t}{m}\right)^m\right\| \leq M.$$

Define the norm

$$|||x||| = \|x\| \vee \sup_{m\geq 1}\sup_{t\geq 0}\left\|V\left(\frac{t}{m}\right)^m x\right\| = \sup_{m\geq 0}\sup_{t\geq 0}\|V(t)^m x\|,$$

and let $\mathcal{N}^*$ be the τ continuous semi-norms that are bounded by $|||\cdot|||$. We clearly have $\|x\| \leq |||x||| \leq M\|x\|$, and more importantly, $|||V(t)||| \leq 1$ for all $t \geq 0$. This implies by Lemma 8.3.5 on 244 for $|||\cdot|||$ and $\mathcal{N}^*$ that for every $p \in \mathcal{N}^*$ there exists $q \in \mathcal{N}^*$ such that

$$\sup_{t\leq T}\sup_{m\geq 1} p\left(V\left(\frac{t}{m}\right)^m x\right) \leq q(x).$$

This implies by the argument above, that the semigroup $\{T(t)\}_{t\geq 0}$ is of type $(1,0)$ for $\mathcal{N}^*$. We obtain the result for $\mathcal{N}$ instead of $\mathcal{N}^*$ by noting that $\|x\| \leq |||x||| \leq M\|x\|$ implies $\mathcal{N} \subset \mathcal{N}^* \subset M\mathcal{N}$. So for $T \geq 0$ and $p \in \mathcal{N} \subset \mathcal{N}^*$, we find $q \in \mathcal{N}^*$ such that

$$\sup_{t\leq T} p\left(T(t)x\right) \leq q\left(x\right)$$

for all $x \in X$. However, as $\mathcal{N}^* \subset M\mathcal{N}$, there exists some $q' \in \mathcal{N}$ such that $q = Mq'$, which concludes the proof. □

Theorem 8.7.2 (Trotter product formula). *Let $(X, \tau, \|\cdot\|)$ satisfy Condition C. Let $\{T(t)\}_{t\geq 0}$ and $\{S(t)\}_{t\geq 0}$ be SCLE semigroups with generators $(A, \mathcal{D}(A))$ and $(B, \mathcal{D}(B))$. Suppose that there is $M \geq 1$ and $\omega \in \mathbb{R}$ such that for all $p \in \mathcal{N}$ and $T \geq 0$ there exists $q \in \mathcal{N}$ such that*

$$\sup_{t\leq T}\sup_{n\geq 1} e^{-\omega t} p\left(\left[T\left(\frac{x}{n}\right)S\left(\frac{x}{n}\right)\right]^n x\right) \leq Mq(x). \tag{8.7.3}$$

Consider $A+B$ on $D = \mathcal{D}(A)\cap\mathcal{D}(B)$ and suppose that D and $(\lambda_0-A-B)D$ are dense in X for some $\lambda_0 > \omega$. Then $\overline{A+B}$ generates a SCLE semigroup $\{U(t)\}_{t\geq 0}$ of type (M,ω) given by the Trotter product formula

$$U(t)x = \lim_{n\to\infty}\left[T\left(\frac{x}{n}\right)S\left(\frac{x}{n}\right)\right]^n x$$

for all $x \in X$, with uniform convergence for t in compact intervals.

If the supremum over $t \leq T$ in equation (8.7.3) *can be extended to a supremum over $t \geq 0$, then $\{U(t)\}_{t\geq 0}$ is of type $(M,\omega)^*$.*

Proof. Define $V(t) := T(t)S(t)$ for $t \geq 0$. Using the local equi-continuity of $\{T(t)\}_{t\geq 0}$, we obtain that $\lim_{t\downarrow 0}\frac{V(t)x-x}{t} = (A+B)x$ for $x \in D$. Therefore, the result follows from Theorem 8.7.1. □

8.8 RELATING BI-CONTINUOUS SEMIGROUPS TO SCLE SEMIGROUPS

Bi-continuous semigroups were introduced by Kühnemund [2003] to study semigroups on Banach spaces that are strongly continuous with respect to a weaker locally convex topology τ and where τ has good sequential properties on norm bounded sets. We will consider the *mixed topology* $\gamma := \gamma(\|\cdot\|,\tau)$, introduced by Wiweger [1961], also see Cooper [1987], which is the strongest locally convex topology that coincides with τ on norm bounded sets. We will show that if γ satisfies $\gamma^+ = \gamma$, then bi-continuity of a semigroup for τ is equivalent to being SCLE for γ.

We start with the assumptions underlying bi-continuous semigroups.

Condition 8.8.1. Let $(X,\|\cdot\|)$ be a Banach space with continuous dual X'_n and dual unit ball B_n. Let τ be another, coarser, locally convex topology on X, with continuous dual X'_τ and dual unit ball $B_\tau = B_n \cap X'_\tau$ that has the following two properties.

(a) The space (X,τ) is sequentially complete on norm bounded sets.

(b) X'_τ is norming for $(X,\|\cdot\|)$, i.e. $\|x\| = \sup_{x'\in B_\tau}|\langle x,x'\rangle|$.

An operator family $\{T(t)\}_{t\geq 0}$ of norm continuous operators on X is called *locally bi-continuous* if for any $t_0 \geq 0$ and for any norm bounded sequence $\{x_n\}_{n\geq 0}$ that converges to x in X with respect to τ, we have

$$\tau - \lim_{n\to\infty} T(t)(x_n - x) = 0$$

uniformly for $t \leq t_0$. Kühnemund [2003] then introduces bi-continuous semigroups.

Definition 8.8.2. A semigroup $\{T(t)\}_{t\geq 0}$ of norm continuous operators on X is called a *bi-continuous semigroup* of type (M, ω) if it satisfies the following properties.

(a) $\{T(t)\}_{t\geq 0}$ is τ strongly continuous.

(b) $\{T(t)\}_{t\geq 0}$ is locally bi-continuous as an operator family.

(c) The semigroup is exponentially bounded: $\|T(t)\| \leq Me^{\omega t}$ for all $t \geq 0$.

We will compare bi-continuous semigroups for τ to SCLE semigroups for mixed topology $\gamma := \gamma(\|\cdot\|, \tau)$.

Proposition 8.8.3. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition 8.8.1. Then, γ is sequentially complete, has the same bounded sets as the norm topology and $(X, \gamma, \|\cdot\|)$ satisfies Condition C.*

For the proof of countable convexity of $\mathcal{N}$, we introduce some notation. Pick some locally convex space (Y, τ^Y). Pick an absolutely convex absorbing subset A of Y. Define τ_A^Y to be the finest locally convex topology on Y that coincides with τ^Y on A. We say that (Y, τ^Y) satisfies *property (L)* if $\tau^Y = \tau_A^Y$ for any absolutely convex absorbing subset A, see Saxon and Sánchez Ruiz [1997].

Proof of Proposition 8.8.3. By Condition 8.8.1, τ and $\|\cdot\|$ satisfy properties (n), (o) and (d) in Wiweger [1961]. Thus, it follows by the Corollary of 2.4.1 in Wiweger [1961] that the γ bounded sets equal the norm bounded sets.

By 2.2.1 in Wiweger [1961], γ coincides with τ on norm bounded sets, which implies that γ is sequentially complete.

We are left to prove that $\mathcal{N}$ is countably convex, which is equivalent to proving that γ satisfies property (L) by [Saxon and Sánchez Ruiz, 1997, Theorem 2.2].

Pick an arbitrary absolutely convex absorbing set $A \subseteq X$. Denote by B the unit ball for $\|\cdot\|$. We start by showing that there exists $\lambda > 0$ such that $\lambda B \subseteq A$. Because A is absorbing, we find by Theorem 8-4-12 in Wilansky [1978] that A° is bounded. By Lemma 8.3.4 on page 243, the norm topology is Banach and equal to the strong topology $\beta(X, (X, \gamma)')$. Therefore, Theorem 20.11.8 in Köthe [1969] implies that A° is also bounded for the dual norm. That it, there exists a $\alpha > 0$ such that $A^\circ \subseteq \alpha B^\circ$, which implies $\alpha^{-1} B \subseteq A^{\circ\circ}$.

The strong closure of A is a barrel, and as the norm topology equals the strong topology, The semi-norm defined by $p_A(x) := \inf_{\lambda > 0}\{x \in \lambda A\}$ is strongly continuous. Thus $2^{-1}\alpha^{-1} B \subseteq A$.

Hence, we can assume that there exists $\lambda > 0$ such that $\lambda B \subseteq A$. Then the finest locally convex topology coinciding with γ on A, denoted by γ_A, is weaker then the finest locally convex topology $\gamma_{\lambda B}$ coinciding on λB with γ. In other words, $\gamma \subseteq \gamma_A \subseteq \gamma_{\lambda B}$. But γ is the mixed topology, and hence the strongest locally convex topology that coincides with τ on norm bounded sets, which implies that $\gamma = \gamma_{\lambda B}$, yielding $\gamma_A = \gamma$. □

The definition of bi-continuous semigroups is given using the convergence of sequences. Therefore, we expect a connection to SCLE semigroups if $\gamma^+ = \gamma$.

Theorem 8.8.4. *Let $(X, \tau, \|\cdot\|)$ satisfy Condition 8.8.1 and let γ be such that $\gamma^+ = \gamma$. $\{T(t)\}_{t\geq 0}$ is bi-continuous for τ if and only if it is SCLE for γ.*

This theorem is an extension of Theorem 3.4 in Farkas [2011], which proves the next result for the strict topology on $X = C_b(E)$ for a Polish space E, see also Section 8.9.1.

Proof. Let $\{T(t)\}_{t\geq 0}$ be bi-continuous for τ. Fix $t_0 > 0$. Because $\sup_{t\leq t_0} \|T(t)\| < \infty$, it follows from 2.2.1 in Wiweger [1961] and the τ strong continuity of $\{T(t)\}_{t\geq 0}$ that the semigroup is also γ strongly continuous.

Because a γ converging sequence is norm bounded, it converges for τ. Thus $\{T(t)\}_{t\leq t_0}$ is sequentially equi-continuous for γ by the local bi-continuity of $\{T(t)\}_{t\geq 0}$. It follows that for a γ continuous semi-norm p there exists a sequentially continuous semi-norm q such that

$$\sup_{t\leq t_0} p\left(T(t)x\right) \leq q(x)$$

for all $x \in X$. However, using that $\gamma^+ = \gamma$ and Theorem 7.4 in Wilansky [1981], q is γ continuous. In other words, $\{T(t)\}_{t\geq 0}$ is locally equi-continuous.

Now let Let $\{T(t)\}_{t\geq 0}$ be SCLE for γ. The semigroup is exponentially bounded by Corollary 8.3.7. Thus, 2.2.1 in Wiweger [1961] implies that $t \mapsto T(t)x$ is τ continuous for every $x \in X$ and that $\{T(t)\}_{t\geq 0}$ is τ locally bi-continuous. □

8.9 THE STRICT TOPOLOGY

We give two examples where a strict topology can be defined. In both cases, this topology is strongly Mackey and satisfies Condition C.

For the first example, let E be a Polish space. We will define the *strict topology* β on $C_b(E)$ which is a particularly nice topology as the continuous dual of $(C_b(E), \beta)$ is the space of Radon measures on E of finite total variation. Therefore, this topology is useful for, for example, the study of transition semigroups of Markov processes.

For the second example, we take a Hilbert space $\mathfrak{H}$ and consider the *strict topology* β on $\mathcal{B}(\mathfrak{H})$, the space of bounded operators on $\mathfrak{H}$. The dual of $(\mathcal{B}(\mathfrak{H}), \beta)$ is the space of normal linear functionals, which are at the basis of non-commutative measure theory, see Takesaki [1979], Kadison and Ringrose [1986], Bratelli and Robinson [1979]. As a consequence, the space $(\mathcal{B}(\mathfrak{H}), \beta)$ is suitable for the study of quantum dynamical semigroups.

8.9.1 *Definition and basic properties of the strict topology on* $C_b(E)$

Recall the strict topology β introduced in Section 2.5.2. Let E be a Polish space. We repeat the construction of β.

For every compact set $K \subseteq E$, define the semi-norm $p_K(f) := \sup_{x \in K} |f(x)|$. Pick a non-negative sequence a_n in $\mathbb{R}$ such that $a_n \to 0$ and pick an arbitrary sequence of compact sets $K_n \subseteq E$. Define

$$p_{(K_n),(a_n)}(f) := \sup_n a_n p_{K_n}(f). \tag{8.9.1}$$

The *strict* topology $\beta = \gamma(\|\cdot\|, \kappa)$ defined on $C_b(E)$ is generated by the semi-norms

$$\left\{ p_{(K_n),(a_n)} \,\middle|\, K_n \text{ compact}, a_n \geq 0, a_n \to 0 \right\}.$$

Obviously, $C_b(E)$ can also equipped with the sup norm topology. In this situation, the set $\mathcal{N}$ contains all semi-norms of the type given in Equation (8.9.1) such that $\sup_n a_n \leq 1$.

Sentilles [1972] studied the strict topology and gives, amongst many others, the following results.

Theorem 8.9.1. *The space* $(C_b(E), \beta)$ *is complete, Mackey, satisfies* $\beta^+ = \beta$ *and* $\beta = \mu(C_b(E), \mathcal{M}(E))$.

Additionally, we have that β *has the same bounded sets as the norm topology and that the norm topology equals* $\beta(C_b(E), \mathcal{M}(E))$.

Proof. Most of these properties have been stated before in Theorem 2.5.8. Equality of the norm and the strong topology follows from Lemma 8.3.4. □

The next result follows directly from Propositions 8.2.3 and 8.3.9.

Corollary 8.9.2. *The locally convex space $(C_b(E), \beta)$ together with the sup norm is strong Mackey and satisfies Condition C.*

Note that (d) implies that $f_n \xrightarrow{\beta} f$ if and only if $\sup_n \|f_n\| < \infty$ and $f_n \xrightarrow{\kappa} f$. This reminds us somewhat of the bounded and pointwise convergence concept defined on $M_b(E)$.

8.9.2 *Definition and properties of the bounded pointwise topology on $M_b(E)$*

A second definition of convergence of functions that is often used in the study of generators of Markov processes is the notion bounded pointwise(bp) on $M_b(E)$, where E is Polish. We say that $f_n \to f$ bounded and pointwise if

(a) $\sup_n \|f_n\| < \infty$,

(b) $f_n(x) \to f(x)$ for all $x \in E$.

Clearly, this notion of convergence is well suited in the context of measure theory because of the dominated convergence theorem.

We will show that this type of convergence can be embedded into the weak topology $\sigma(M_b(E), \mathcal{M}(E))$ and show that this topology is related to the strict topology.

Lemma 8.9.3. *The space $\sigma(M_b(E), \mathcal{M}(E))$ is sequentially complete, the supremum norm topology equals $\beta(M_b(E), \mathcal{M}(E))$ and weakly bounded sets equal norm bounded sets.*

Finally, $(M_b(E), \sigma(M_b(E), \mathcal{M}(E)))$ satisfies Convexity condition C.

Proof. We start by proving the sequential completeness of $(M_b(E), \sigma(M_b(E), \mathcal{M}(E)))$. Suppose we have a Cauchy sequence f_n that is not norm bounded.

Without loss of generality we can find $x_n \in E$ such that $|f_n(x_n)| > n$. We can also assume that all x_n are distinct, because if some point y appears infinitely often in the sequence $\{x_n\}_{n\geq 1}$, we see that $f_n(y)$ diverges.

Now consider the map $\phi : M_b(E) \to l^\infty$, defined by $\phi(f) = \{f(x_n)\}_{n\geq 1}$. We show that this map is continuous from $\sigma(M_b(E), \mathcal{M}(E))$ to $\sigma(l^\infty, l^1)$.

For any sequence $a = \{a_n\}_{n \geq 1}$ in l^1, we define a measure $\mu_a \in \mathcal{M}(E)$ by setting

$$\mu(A) = \sum_n a_n \mathbb{1}_{\{x_n \in A\}}.$$

It follows that $\langle \phi(f), a \rangle = \langle f, \mu_a \rangle$ which implies the continuity of ϕ. Because $l^\infty = (l^1, \|\cdot\|)'$, the norm and weakly bounded sets coincide in l^∞ by the Principle of Uniform boundedness, see for example Theorem V.1.10 in Conway [2007]. Because $\phi(f_n)$ is Cauchy in $(l^\infty, \sigma(l^\infty, l^1))$, it is norm bounded, which contradicts the fact that $|f_n(x_n)| > n$.

We conclude that $\sup_n \|f_n\| < \infty$. Clearly, as $f_n(x)$ is Cauchy for all $x \in E$, we can define a pointwise limit $f \in M_b(E)$. By the dominated convergence theorem we obtain $\langle f_n, \mu \rangle \to \langle f, \mu \rangle$ for all $\mu \in \mathcal{M}(E)$. In other words, $(M_b(E), \sigma(M_b(E), \mathcal{M}(E))$ is sequentially complete.

For the second and the third claim, note that $\|f\| = \sup_{x \in E} |\langle f, \delta_x \rangle|$, therefore $\|\cdot\|$ is lower semi-continuous for $\sigma(M_b(E), \mathcal{M}(E))$. It follows by Lemma 8.3.4 that the norm topology equals $\beta(M_b(E), \mathcal{M}(E))$ and that the norm and $\sigma(M_b(E), \mathcal{M}(E))$ bounded sets coincide.

Convexity property C follows from Proposition 8.3.9 (d). □

We proceed with the result that connects *bp* convergence to weak convergence. This Lemma can also be found as Proposition 3.1 in the Appendix of Ethier and Kurtz [1986].

Lemma 8.9.4. *Consider the weak topology $\sigma(M_b(E), \mathcal{M}(E))$. A sequence f_n in $M_b(E)$ converges weakly to $f \in M_b(E)$ if and only if $f_n \overset{bp}{\to} f$.*

Proof. Clearly, by the dominated convergence theorem, if $f_n \overset{bp}{\to} f$, then $f_n \to f$ for $\sigma(M_b(E), \mathcal{M}(E))$. On the other hand, if $f_n \to f$ for $\sigma(M_b(E), \mathcal{M}(E))$, then we obtain $f_n(x) \to f(x)$ for all x by considering the Dirac measures $\delta_x \in \mathcal{M}(E)$. Boundedness follows from Lemma 8.9.3. □

As a final question, we consider the relation between $M_b(E)$ and $C_b(E)$. First note that $M_b(E)$ is a weakly sequentially closed subspace of the $\sigma(C_b(E), \mathcal{M}(E))$ completion of $C_b(E)$ by a theorem of Grothendieck, 21.2.(2) Köthe [1969].

Not every element in $M_b(E)$ can be weakly approximated by a sequence of elements from $C_b(E)$ as can be seen from the indicator function of $\mathbb{Q}$.

However, writing $\mathbb{Q} = \{q_1, q_2, \dots\}$, it is clear we can approximate the function $f_k = \mathbb{1}_{\{q_1,\dots,q_k\}}$ by a sequence of continuous functions for the weak topology. Additionally, $f_k \to \mathbb{1}_{\mathbb{Q}}$ for the weak topology.

This raises the question whether $M_b(E)$ is the smallest sequentially closed subspace of the weak completion of $C_b(E)$ that is sequentially closed. In some sense, the following result is the functional analytic counterpart of the fact that the smallest σ algebra containing all closed balls is the Borel σ algebra.

Proposition 8.9.5. *The space $M_b(E)$ is the smallest sequentially closed subspace in the $\sigma(C_b(E), \mathcal{M}(E))$ completion of $C_b(E)$.*

Proof. If E is metric, let $B_r(x)$ be the closed metric ball of radius r centred at x. Clearly, the sequence of functions

$$f_n(y) = (1 - nd(y, B_r(x))) \vee 0$$

decreases pointwise to $\mathbb{1}_{B_r(x)}$. Therefore, the sequence converges in the $\sigma(M_b(E), \mathcal{M}(E))$ topology.

The result follows by a monotone class argument, see Theorem 4.3 in the Appendix of Ethier and Kurtz [1986]. □

8.9.3 *Definition and basic properties of the strict topology on $\mathcal{B}(H)$*

Let $\mathfrak{H}$ be a Hilbert space and let $(\mathcal{B}(\mathfrak{H}), \|\cdot\|)$ be the Banach space of bounded linear operators on $\mathfrak{H}$. Furthermore, let $\mathcal{K}(\mathfrak{H})$ and $\mathcal{T}(\mathfrak{H})$ be the subspace of compact and trace class operators on $\mathfrak{H}$. Note that $\mathcal{B}(\mathfrak{H}) = \mathcal{T}(\mathfrak{H})' = \mathcal{K}(\mathfrak{H})''$ as Banach spaces by Theorems II.1.6 and II.1.8 in Takesaki [1979].

We define four additional topologies on $\mathcal{B}(\mathfrak{H})$.

(a) The *strong* (operator) topology* generated by the semi-norms $\{p_\xi \mid \xi \in \mathfrak{H}\}$, where $p_\xi(A) := \sqrt{\|A\xi\|^2 + \|A^*\xi\|^2}$.

(b) The *ultraweak (operator) topology* generated by the family of semi-norms $\{p_T \mid T \in \mathcal{T}(\mathfrak{H})\}$, where $p_T(A) := |\operatorname{Tr}(AT)|$.

(c) The *ultrastrong* (operator) topology* generated by the family of semi-norms $\{p_T \mid T \in \mathcal{T}(\mathfrak{H}), T \geq 0\}$, where $p_T(A) := \sqrt{\operatorname{Tr}(TA^*A)}$.

(d) The *strict topology* β defined by the set of semi-norms $p_B(A) := \|AB\|$ and $q_B(A) := \|BA\|$ for compact operators $B \in \mathcal{K}(\mathfrak{H})$.

The ultraweak topology is the weak topology of the dual pair $(\mathcal{B}(\mathfrak{H}), \mathcal{T}(\mathfrak{H}))$ and also the ultrastrong* topology is a topology of this pair, see for example

[Takesaki, 1979, Lemma II.2.4]. The strict topology is the Mackey topology of this dual pair by Theorem 3.9 in Busby [1968] and Corollary 2.8 in Taylor [1970].

The linear functionals on $\mathcal{B}(\mathfrak{H})$ that are continuous with respect to any topology of the dual pair $(\mathcal{B}(\mathfrak{H}), \mathcal{T}(\mathfrak{H}))$ are called *normal*, to distinguish them from the larger class of linear functionals on $\mathcal{B}(\mathfrak{H})$ that are continuous for the norm, see also the reference that were mentioned before Takesaki [1979], Kadison and Ringrose [1986], Bratelli and Robinson [1979]. The distinction between the two classes of functionals is analogous to the difference between Radon measures on $C_b(E)$, E non-compact and Polish, and the linear functionals on $C_b(E)$ that are norm continuous.

Proposition 8.9.6. *The space $(\mathcal{B}(\mathfrak{H}), \beta)$ is complete, strong Mackey, the bounded sets equal the operator norm bounded sets, and $(\mathcal{T}(\mathfrak{H}), \sigma(\mathcal{T}(\mathfrak{H}), \mathcal{B}(\mathfrak{H}))$ is sequentially complete.*

Proof. Proposition 3.6 in Busby [1968] gives completeness. The principle of uniform boundedness gives equality of the bounded sets. To show that β is strong Mackey, we need to verify that the absolutely convex hull of a $\sigma(\mathcal{T}(\mathfrak{H}), \mathcal{B}(\mathfrak{H}))$ compact set is also compact. This follows directly from Krein's theorem, 24.5.(4) in Köthe [1969] as the Mackey topology $\mu(\mathcal{T}(\mathfrak{H}), \mathcal{B}(\mathfrak{H}))$ is the Banach topology generated by the Trace norm. The final statement follows from Corollary III.5.2 in Takesaki [1979]. □

Corollary 8.9.7. *The space $(\mathcal{B}(\mathfrak{H}), \beta)$ together with the operator norm is strong Mackey and satisfies Condition C.*

If $\mathfrak{H}$ is separable, we additionally have the following result.

Proposition 8.9.8. *If $\mathfrak{H}$ is separable, then $(\mathcal{B}(\mathfrak{H}), \beta)$ is separable and $\beta^+ = \beta$*

Proof. Suppose $\mathfrak{H}$ is separable. Then $\mathcal{K}(\mathfrak{H})$ is norm separable by Lemma 1 in Goldberg [1959] which implies that it is separable for β. By Proposition 3.5 in Busby [1968], $\mathcal{K}(\mathfrak{H})$ is β dense in $\mathcal{B}(\mathfrak{H})$, which implies the first statement.

By Theorem II.2.6 and Proposition II.2.7 in Takesaki [1979] $(\mathcal{B}(\mathfrak{H}), ultrastrong^*)$ is Mazur. Consider the topology $(ultrastrong^*)^+$. By Theorem 7.5 in Wilansky [1981], $(ultrastrong^*)^+$ is a topology of the dual pair $(\mathcal{B}(\mathfrak{H}), \mathcal{T}(\mathfrak{H}))$, hence must be coarser than the strict topology. By Theorem III.5.7 in Takesaki [1979] the strict topology coincides on bounded sets with the $ultrastrong^*$ topology. Hence, both have the same convergent sequences, which implies that $(ultrastrong^*)^+$ is finer than the strict topology. Therefore, they coincide. This also implies that $\beta^+ = \beta$. □

Let $\{P_t\}_{t\geq 0}$ be a strongly continuous semigroup on $\mathfrak{H}$. The semigroup $\{T(t)\}_{t\geq 0}$ defined on $\mathcal{B}(\mathfrak{H})$ by $T(t)A = P^*(t)AP(t)$ is a basic example in the study of quantum dynamical semigroups, which are normally defined to be merely continuous for the ultraweak topology Fagnola [1999].

Proposition 8.9.9. *The semigroup $\{T(t)\}_{t\geq 0}$ is a SCLE semigroup for the strict topology.*

It is of interest to see whether more quantum dynamical semigroups are in fact continuous for the strict topology. This, however, goes beyond the scope of this thesis.

Proof. Fix $A \in \mathcal{B}(\mathfrak{H})$. The strong continuity of $\{P(t)\}_{t\geq 0}$ implies the operator strong* continuity of $t \mapsto T(t)A$. Therefore, the trajectory $t \mapsto T(t)A$ is locally bounded for the strong* topology, and hence, by the principle of uniform boundedness for the norm topology. Because the strict topology coincides with the strong* topology on bounded sets [Takesaki, 1979, Lemma II.2.5 and Theorem III.5.7] $t \mapsto T(t)A$ is continuous for the strict topology. The semigroup is locally equi-continuous by Lemma 8.2.2. □

8.10 APPENDIX: STOCHASTIC DOMINATION AND THE CHERNOFF BOUND

In this appendix, we recall the definition of stochastic domination [Lindvall, 1992, Section IV.1] and give a number of useful results.

Definition 8.10.1. Suppose that we have two random variables η_1 and η_2 taking values in $\mathbb{R}$. We say that η_1 stochastically dominates η_2, denoted by $\eta_1 \succeq \eta_2$ if for every $r \in \mathbb{R}$ we have $\mathbb{P}[\eta_1 > r] \geq \mathbb{P}[\eta_2 > r]$.

Lemma 8.10.2. *For two real valued random variables η_1, η_2, we have that $\eta_1 \succeq \eta_2$ if and only if for every bounded and increasing function ϕ, we have $\mathbb{E}[\phi(\eta_1)] \geq \mathbb{E}[\phi(\eta_2)]$.*

We say that a random variable η is Poisson(γ) distributed, $\gamma \geq 0$, denoted by $\eta \sim$ Poisson(γ) if $\mathbb{P}[\eta = k] = \frac{\gamma^k}{k!}e^{-\gamma}$.

Lemma 8.10.3. *If $\eta_1 \sim$ Poisson(γ_1) and $\eta_2 \sim$ Poisson(γ_2) and $\gamma_1 \geq \gamma_2$, then $\eta_1 \succeq \eta_2$.*

Using the theory of couplings [Lindvall, 1992, Section IV.2], a proof follows directly from the fact that if $\gamma_1 \geq \gamma_2$, then η_1 is in distribution equal to $\eta_2 + \zeta$, where $\zeta \sim \text{Poisson}(\gamma_1 - \gamma_2)$.

The next result, introduced by Chernoff [1952], is useful in the context of stochastic domination.

Proposition 8.10.4. *Let X be a random variable on $\mathbb{R}$ for which there exists $\theta_0 > 0$, such that for $\theta < \theta_0$, the Laplace transform $\mathbb{E}[e^{\theta X}]$ exists. Let $\{X_i\}_{i\geq 1}$ be independent and distributed as X. Then for $c \geq \mathbb{E}[X]$, we have*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > c\right] < \exp\left\{-n \inf_{0<\theta<\theta_0} \left\{c\theta - \log \mathbb{E}[e^{\theta X}]\right\}\right\}.$$

We give a proof for completeness.

Proof. For all $0 < \theta < \theta_0$, we have

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > c\right] &= \mathbb{P}\left[e^{\theta \sum_{i=1}^{n} X_i} > e^{n\theta c}\right] \\ &< \exp\left\{-\left(n\theta c - \log \mathbb{E}\left[e^{\theta \sum_{i=1}^{n} X_i}\right]\right)\right\},\end{aligned}$$

where we used Markov's inequality in line 2. Because the X_i are independent, $\log \mathbb{E}\left[e^{\theta \sum_{i=1}^{n} X_i}\right] = n \log \mathbb{E}\left[e^{\theta X}\right]$, which yields the final result. □

9 MARTINGALE PROBLEM AND THE TRANSITION SEMIGROUP

Equipped with the knowledge of Chapter 8, we consider the setting of the martingale problem of Section 2.3.4. In particular, we will establish that, under a compact containment condition, the transition semigroup of the solution of a well-posed martingale problem gives rise to a strongly continuous and locally equi-continuous semigroup on the space of bounded continuous functions equipped with the strict topology.

9.1 PRELIMINARIES

We shortly recall some notions. We work with a complete separable metric space (E, d).

Definition 9.1.1 (The martingale problem)**.** Let $A : \mathcal{D}(A) \subseteq C_b(E) \to C_b(E)$ be a linear operator. For $(A, \mathcal{D}(A))$ and a measure $\nu \in \mathcal{P}(E)$, we say that $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ solves the *martingale problem* for (A, ν) if for all $f \in \mathcal{D}(A)$

$$M_f(t) := f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s$$

is a mean 0 $\mathbb{F}^X = \{\mathcal{F}_t^X\}_{t\geq 0}$ martingale under $\mathbb{P}$, and if $\mathbb{P}X(0)^{-1} = \nu$.

We denote the set of all solutions to the martingale problem, for varying initial measures ν, by $\mathcal{M}_A$. We say that *uniqueness* holds for the martingale problem if for every $\nu \in \mathcal{P}(E)$ the set $\{\mathbb{P} \in \mathcal{M}_A \,|\, \mathbb{P}X(0)^{-1} = \nu\}$ is empty or a singleton. Furthermore, we say that the martingale problem is *well-posed* if this set is a singleton.

Additionally, we will consider restricted martingale problem. Let $\Gamma \subseteq \mathcal{M}_A$ and denote

$$\Gamma_\nu := \left\{\mathbb{P} \in \Gamma \,\middle|\, \mathbb{P}X(0)^{-1} = \nu\right\}.$$

We write Γ_x for Γ_{δ_x}. We say that the Γ-restricted martingale problem is well posed if Γ_ν is a singleton for all $\nu \in \mathcal{P}(E)$.

An example of where this latter construction occurs is in the construction of measure valued diffusion processes. See for example Section II.5 Perkins [2002] or Chapter 1 in Etheridge [2000], where Γ is of the form

$$\Gamma := \{\mathbb{P} \in \mathcal{M}_A \,|\, \forall f \in \mathcal{D}(A) : < M_f >_t = \phi_f(t)\},$$

where $< M_f >$ denotes the quadratic variation process of M_f and where $\{\phi_f(t)\}_{f\in\mathcal{D}(A)}$ is some collection of increasing deterministic function.

The construction of solutions to the martingale problem can often be done via approximating processes. Classically, uniqueness for the martingale problem is proven via duality. Costantini and Kurtz [2015], however, introduced a method based on viscosity solutions. In the light of other results based on viscosity solutions in this thesis, we will sketch shortly how this works in the case that $\Gamma = \mathcal{M}_A$.

For $\lambda > 0$ and $h \in C_b(E)$, consider the equation

$$f(x) - \lambda Af(x) - h(x) = 0. \tag{9.1.1}$$

For a function $h \in C_b(E)$, $\lambda > 0$ and $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ define

$$R(\lambda, h, \mathbb{P}) := \mathbb{E}^{\mathbb{P}}\left[\int_0^\infty \frac{1}{\lambda} e^{-\lambda^{-1}t} h(X(t))\mathrm{d}t\right],$$

and use this to define for $h \in C_b(E)$ and $x \in E$ and $\lambda > 0$

$$R^+(\lambda)h(x) = \sup_{\mathbb{P}\in\Gamma_x} R(\lambda, h, \mathbb{P})$$
$$R^-(\lambda)h(x) = \inf_{\mathbb{P}\in\Gamma_x} R(\lambda, h, \mathbb{P}).$$

The following lemma is Lemma 3.5 in Costantini and Kurtz [2015].

Lemma 9.1.2. *For $\lambda > 0$ and $h \in C_b(E)$, $R^+(\lambda)h$ is a viscosity subsolution to* (9.1.1) *and $R^-(\lambda)h$ is a viscosity supersolution to* (9.1.1).

Clearly, if the comparison principle is satisfied for $h \in C_b(E)$ and $\lambda > 0$, we obtain $R^+(\lambda)h = R^-(\lambda)h$. In fact, if this holds for sufficiently many $h \in C_b(E)$ then we obtain uniqueness for martingale problem.

Theorem 9.1.3 (Theorem 3.7 in Costantini and Kurtz [2015])**.** *Let $D \subseteq C_b(E)$ be β dense. If the comparison principle holds for $f - \lambda Af - h = 0$ for all $\lambda > 0$ and $h \in D$, then uniqueness holds for the martingale problem.*

Note that the results in Costantini and Kurtz [2015] also hold for alternative definitions of viscosity solutions. The important meta-statement is that any definition of viscosity solution such that $R^+(\lambda)h$ and $R^-(\lambda)h$ are viscosity sub- and supersolutions to (9.1.1), for which one can also show the comparison principle, suffices to prove uniqueness to the martingale problem.

9.2 THE TRANSITION SEMIGROUP IS STRONGLY CONTINUOUS AND LOCALLY EQUI-CONTINUOUS WITH RESPECT TO THE STRICT TOPOLOGY

In the setting that uniqueness holds for the martingale problem, e.g. if the comparison principle holds as above, we obtain a strengthened version of Theorem 4.5.11, Ethier and Kurtz [1986].

Theorem 9.2.1. *Let $A \subseteq C_b(E) \times C_b(E)$ and let Γ be a set of solutions to the martingale problem for A. Suppose that the closed convex hull of $\mathcal{D}(A)$ is β dense in $C_b(E)$. Suppose that for all $\nu \in \mathcal{P}(E)$ $\Gamma_\nu \neq \emptyset$ and that for all compact $K \subseteq \mathcal{P}(E)$, $\varepsilon > 0$ and $T > 0$, there exists a compact set $K' = K'(K, \varepsilon, T)$ such that for all $\mathbb{P} \in \Gamma$, we have*

$$\mathbb{P}\left[X(t) \in K' \text{ for all } t < T, X(0) \in K\right] \geq (1-\varepsilon)\mathbb{P}\left[X(0) \in K\right]. \tag{9.2.1}$$

Suppose in addition that uniqueness holds for the Γ-martingale problem for A, then the solutions correspond to strong Markov processes with a β-SCLE semigroup $\{S(t)\}_{t\geq 0}$ on $C_b(E)$ defined by $S(t)f(x) = \mathbb{E}[f(X(t)) \mid X(0) = x]$.

In particular, this holds if there is a β dense set $D \subseteq C_b(E)$ such that the comparison principle is satisfied for $f - \lambda Af - h = 0$ for all $\lambda > 0$ and $h \in D$.

Proof of Theorem 9.2.1. The proof that the solutions are strong Markov and correspond to a semigroup

$$S(t)f(x) = \mathbb{E}[f(X(t)) \mid X(0) = x]$$

that maps $C_b(E)$ into $C_b(E)$ follows as in the proof of (b) and (c) of Theorem 4.5.11 Ethier and Kurtz [1986]. We are left to show that $\{S(t)\}_{t\geq 0}$ is SCLE for β, which we do in Lemma 9.2.2 and Proposition 9.2.3. □

Lemma 9.2.2. *Let $\{S(t)\}_{t\geq 0}$ be the semigroup introduced in Theorem 9.2.1. The family $\{S(t)\}_{t\geq 0}$ is locally equi-continuous for β.*

Proof. Fix $T \geq 0$. We will prove that $\{S(t)\}_{t \leq T}$ is β equi-continuous by using Theorem 2.5.8 (c) and (d). Pick a sequence f_n converging to f with respect to β. It follows that $\sup_n \|f_n\| \leq \infty$, which directly implies that $\sup_n \sup_{t \leq T} \|S(t) f_n\| < \infty$.

We also know that $f_n \to f$ uniformly on compact sets. We prove that this implies the same for $S(t)f_n$ and $S(t)f$ uniformly in $t \leq T$. Fix $\varepsilon > 0$ and a compact set $K \subseteq E$, and let $\hat{K}$ be the set introduced in Equation (9.2.1) for T. Then we obtain that

$$
\begin{aligned}
&\sup_{t \leq T} \sup_{x \in K} |S(t)f(x) - S(t)f_n(x)| \\
&\leq \sup_{t \leq T} \sup_{x \in K} \mathbb{E}_x \left| f(X(t)) - f_n(X(t)) \right| \\
&\leq \sup_{t \leq T} \sup_{x \in K} \mathbb{E}_x \Big| (f(X(t)) - f_n(X(t))) \mathbb{1}_{\{X(t) \in \hat{K}\}} \\
&\qquad\qquad + (f(X(t)) - f_n(X(t))) \mathbb{1}_{\{X(t) \in \hat{K}^c\}} \Big| \\
&\leq \sup_{t \leq T} \sup_{y \in \hat{K}} |f(y) - f_n(y)| + \sup_n \|f_n - f\| \varepsilon.
\end{aligned}
$$

As $n \to \infty$ this quantity is bounded by $\sup_n \|f_n - f\| \varepsilon$ as f_n converges to f uniformly on compacts. As ε was arbitrary, we are done. □

Proposition 9.2.3. *Let $\{S(t)\}_{t \geq 0}$ be the semigroup introduced in Theorem 9.2.1. Then $\{S(t)\}_{t \geq 0}$ is β strongly continuous.*

For the proof, we recall one definition from Section 8.2. We say that a semigroup is *weakly continuous* if for all $f \in C_b(E)$ and $\mu \in \mathcal{M}(E)$ the trajectory $t \mapsto \langle S(t)f, \mu \rangle$ is continuous in $\mathbb{R}$.

Proof of Proposition 9.2.3. First, recall that $(C_b(E), \beta)$ is strong Mackey and complete. By Lemma 9.2.2 the semigroup $\{S(t)\}_{t \geq 0}$ is locally equi-continuous. Therefore, Proposition 8.2.5 implies that we only need to prove weak continuity. So let $f \in C_b(E)$ and $\mu \in \mathcal{M}(E)$. Write μ as the Hahn-Jordan decomposition: $\mu = c^+ \mu^+ - c^- \mu^-$, where $c^+, c^- \geq 0$ such that $\mu^+, \mu^- \in \mathcal{P}(E)$.

We show that $t \mapsto \langle S(t)f, \mu \rangle$ is continuous, by showing that $t \mapsto \langle S(t)f, \mu^+ \rangle$ and $t \mapsto \langle S(t)f, \mu^- \rangle$ are continuous. Clearly, it suffices to do this for either of the two.

Let $\mathbb{P}$ be the unique measure in Γ_{μ^+}. It follows by Theorem 4.3.12 in Ethier and Kurtz [1986] that $\mathbb{P}[X(t) = X(t-)] = 1$ for all $t > 0$, so $t \mapsto X(t)$ is continuous $\mathbb{P}$ almost surely. Fix some $t > 0$, we show that our trajectory is continuous for this specific t.

$$\left|\langle S(t)f, \mu^+\rangle - \langle S(t+h)f, \mu^+\rangle\right| \leq \mathbb{E}^{\mathbb{P}} \left|f(X(t) - f(X(t+h))\right|.$$

By the almost sure convergence of $X(t + h) \to X(t)$ as $h \to 0$, and the boundedness of f, we obtain by the dominated convergence theorem that this difference converges to 0 as $h \to 0$. As $t > 0$ was arbitrary, the trajectory is continuous for all $t > 0$. Continuity at 0 follows by the fact that all trajectories in $D_E(\mathbb{R}^+)$ are continuous at 0. □

10 A BANACH-DIEUDONNÉ THEOREM FOR THE SPACE OF BOUNDED CONTINUOUS FUNCTIONS

The identification of the strict topology on the space of bounded continuous functions as the correct space in relation to measure theory in Section 8.9.1 and Chapter 9 causes the question to arise which properties of the space $(C_b(X), \|\cdot\|)$ carry over to $(C_b(X), \beta)$ if we replace a compact metric space X by a Polish space X. A list of properties for which this is the case has been given in Theorem 2.5.8.

Specific to the theory of Banach spaces, i.e. when X is compact, are the closed graph, inverse- and open mapping theorems. Given that measure theory, i.e. the study of the continuous dual of $(C_b(X), \beta)$ is in some sense not dependent on the fact that X is compact, leads to the conjecture that these results can be obtained for non-compact Polish X as well. These results have been obtained in:

Richard Kraaij. A Banach-Dieudonné theorem for the space of bounded continuous functions on a separable metric space with the strict topology. *Topology and its Applications*, 2016c. ISSN 0166-8641. doi: 10.1016/j.topol.2016.06.003.

10.1 INTRODUCTION AND RESULTS

Let (E, t) be a locally convex space. Denote by E' the continuous dual space of (E, t) and denote by $\sigma = \sigma(E', E)$ the weak topology on E'. We consider the following additional topologies on E':

- σ^f, the finest topology coinciding with σ on all t-equi-continuous sets in E'.
- σ^{lf}, the finest locally convex topology coinciding with σ on all t-equi-continuous sets in E'.
- t° the polar topology of t defined on E'. t° is defined in the following way. Let $\mathcal{N}$ be the collection of all t pre-compact sets in E. A pre-

compact set, is a set that is compact in the completion of (E, t). Then the topology t° on E' is generated by all semi-norms of the type

$$p_N(\mu) := \sup_{f \in N} |\langle f, \mu \rangle| \qquad N \in \mathcal{N}.$$

The Banach-Dieudonné theorem for locally convex spaces is the following, see Theorems 21.10.1 and 21.9.8 in Köthe [1969].

Theorem 10.1.1 (Banach-Dieudonné)**.** *Let (E, t) be a metrizable locally convex space, then the topologies σ^f and t° on $(E, t)'$ coincide. If (E, t) is complete, these topologies also coincide with σ^{lf}.*

The Banach-Dieudonné theorem is of interest in combination with the closed graph theorem. For the discussion of closed graph theorems, we need some additional definitions. Considering a locally convex space (E, t), we say that

(a) E' satisfies the *Krein-Smulian* property if every σ^f closed absolutely convex subset of E' is σ closed;

(b) (E, t) is a *Pták space* if every σ^f closed linear subspace of E' is σ closed;

(c) (E, t) is a *infra Pták space* if every σ dense σ^f closed linear subspace of E' equals E'.

Infra-Pták spaces are sometimes also called B_r *complete* and Pták spaces are also known as B *complete* or *fully complete*. Finally, Theorem 1 of Kelley [1958] shows that the Krein-Smulian property for E' is equivalent to *hypercompleteness* of E: the completeness of the space of absolutely convex closed neighbourhoods of 0 in (E, t) equipped with the Hausdorff uniformity.

Clearly, we have that E hypercomplete implies E Pták implies E infra Pták. Additionally, if E is a infra Pták space, then it is complete by 34.2.1 in Köthe [1979]. See also Chapter 7 in Carreras and Bonet [1987] for more properties of Pták spaces.

We have the following straightforward result, using that the absolutely convex closed sets agree for all locally convex topologies that give the same dual.

Proposition 10.1.2. *If σ^{lf} and σ^f coincide on E', then E is hypercomplete.*

Theorem 10.1.3 (Closed graph theorem, cf. 34.6.9 in Köthe [1979])**.** *Every closed linear map of a barrelled space E to an infra-Pták space F is continuous.*

Because Banach and Fréchet spaces are metrizable, they are infra-Pták space by Theorem 10.1.1 and Proposition 10.1.2. Additionally, they are both barrelled spaces, which implies that the closed graph theorem holds for closed linear maps from a Fréchet space to a Fréchet space. As a consequence, also the inverse and open mapping theorems hold.

In this chapter, we study the space of bounded and continuous functions on a separable metric space X equipped with the *strict* topology β. For the definition and a study of the properties of β, see Sentilles [1972]. Note that this setting is slightly more general than in previous chapters where we considered spaces that were metrizable by a complete separable metric. A difference is that for the strict topology on a separably metrizable space, the dual space equals the space $\mathcal{M}_\tau(X)$ of τ-additive Borel measures on X. A Borel measure μ is called τ-additive if for any increasing net $\{U_\alpha\}_\alpha$ of open sets, we have

$$\lim_{\alpha} |\mu|(U_\alpha) = |\mu| \left(\cup_\alpha U_\alpha\right).$$

As in the previous chapter, if X is metrizable by a complete separable metric, the space of τ additive Borel measures equals the space of Radon measures.

The space $(C_b(X), \beta)$ is not barrelled unless X is compact, Theorem 4.8 of Sentilles [1972] so Theorem 10.1.3 does not apply for this class of spaces. Thus, the following closed graph theorem by Kalton is of interest, as it puts more restrictions on the spaces serving as a range, relaxing the conditions on the spaces allowed as a domain.

Theorem 10.1.4 (Kalton's closed graph theorem, Theorem 2.4 in Kalton [1971], Theorem 34.11.6 in Köthe [1979])**.** *Every closed linear map from a Mackey space E with weakly sequentially complete dual E' into a transseparable infra-Pták space F is continuous.*

Remark 10.1.5. Note that this result is normally stated for separable infra-Pták space F. In the proof of Kalton's closed graph theorem 34.11.6 in Köthe [1979], separability is only used to obtain that weakly compact sets of the dual E' are metrizable. For this transseparability suffices by Lemma 1 in Pfister [1976].

A class of spaces, more general than the class of Fréchet spaces, satisfying the conditions for both the range and the domain space in Kalton's closed graph theorem, would be an interesting class of spaces to study. In this chapter, we show that $(C_b(X), \beta)$, for a separable metric space X belongs to this class. In particular, the main result in this chapter is that $(C_b(X), \beta)$ satisfies the conclusions of the Banach-Dieudonné theorem.

First, we introduce an auxiliary result and the definition of a k-space, which are relevant in view of the defining properties of σ^f.

Proposition 10.1.6. *$(C_b(X), \beta)$ is a strong Mackey space. In other words, β is a Mackey topology and the weakly compact sets in $\mathcal{M}_\tau(X)$ and the weakly closed β equi-continuous sets coincide.*

Proof. This follows by Theorem 5.6 in Sentilles [1972], Corollary 6.3.5 and Proposition 7.2.2(iv) in Bogachev [2007]. □

We say that a topological space (Y, t) is a k-space if a set $A \subseteq Y$ is t-closed if and only if $A \cap K$ is t-closed for all t-compact sets $K \subseteq Y$. The strongest topology on Y coinciding on t-compact sets with the original topology t is denoted by kt and is called the k-ification of t. The closed sets of kt are the sets A in Y such that $A \cap K$ is t-closed in Y for all t-compact sets $K \subseteq Y$. We see that for a strong Mackey space E, $\sigma^f = k\sigma$ on E'.

Theorem 10.1.7. *Let X be a separably metrizable space. Consider the space $(C_b(X), \beta)$, where β is the strict topology. Then σ^{lf}, σ^f, $k\sigma$ and β° coincide on $\mathcal{M}_\tau(X)$.*

In the process of proving the theorem, we will obtain various auxiliary results, we will mention a result that is relevant in view of Kalton's closed graph theorem.

Lemma 10.1.8. *Let X be a separably metrizable space. Then $(C_b(X), \beta)$ is transseparable.*

Additionally, we have the following known result, Theorem 8.7.1 in Bogachev [2007].

Lemma 10.1.9. *Let X be separably metrizable, then the dual $\mathcal{M}_\tau(X)$ of $(C_b(X), \beta)$ is weakly sequentially complete.*

We immediately note that a second proof of this lemma can be given using the theory of Mazur spaces.

Remark 10.1.10 (Second proof). β is the Mackey topology on $(C_b(X), \beta)$ by Proposition 10.1.6, we find $\mathcal{M}_\tau(X)$ is weakly sequentially complete by Theorem 8.1 in Sentilles [1972], Theorem 7.4 in Wilansky [1981] and Propositions 4.3 and 4.4 in Webb [1968].

As a consequence of Theorem 10.1.7 and Lemma's 10.1.8 and 10.1.9, $(C_b(X), \beta)$ satisfies both the conditions to serve as a range, and as a domain in Kalton's closed graph theorem. We have the following important corollaries.

Corollary 10.1.11 (Closed graph theorem). *Let X, Y be separably metrizable spaces, then a closed linear map from $(C_b(X), \beta)$ to $(C_b(Y), \beta)$ is continuous.*

Corollary 10.1.12 (Inverse mapping theorem). *Let X, Y be separably metrizable spaces. Let $T : (C_b(X), \beta) \to (C_b(Y), \beta)$ be a bijective continuous linear map. Then $T^{-1} : (C_b(Y), \beta) \to (C_b(X), \beta)$ is continuous.*

Corollary 10.1.13 (Open mapping theorem). *Let X, Y be separably metrizable spaces. Let $T : (C_b(X), \beta) \to (C_b(Y), \beta)$ be a surjective continuous linear map. Then T is open.*

10.2 IDENTIFYING THE FINEST TOPOLOGY COINCIDING WITH σ ON ALL β EQUI-CONTINUOUS SETS

Denote by $\mathcal{M}_{\tau,+}(X)$ the subset of non-negative τ-additive Borel measures on X and denote by σ_+ the restriction of σ to $\mathcal{M}_{\tau,+}(X)$. Consider the map

$$\begin{cases} q : \mathcal{M}_{\tau,+}(X) \times \mathcal{M}_{\tau,+}(X) \to \mathcal{M}_\tau(X) \\ q(\mu, \nu) = \mu - \nu. \end{cases}$$

Note that by the Hahn-Jordan theorem the map q is surjective.

Definition 10.2.1. Let $\mathcal{T}$ denote the quotient topology on $\mathcal{M}_\tau(X)$ of the map q with respect to $\sigma_+ \times \sigma_+$ on $\mathcal{M}_{\tau,+}(X) \times \mathcal{M}_{\tau,+}(X)$.

The next few lemma's will provide some key properties of $\mathcal{T}$, which will lead to the proof that $\mathcal{T} = \sigma^f$.

Lemma 10.2.2. *$(\mathcal{M}_\tau(X), \mathcal{T})$ is a k-space.*

Proof. First of all, the topology σ_+ is metrizable by Theorem 8.3.2 in Bogachev [2007]. This implies that σ_+^2 is metrizable. Metrizable spaces are k-spaces by Theorem 3.3.20 in Engelking [1989]. Thus $(\mathcal{M}_\tau(X), \mathcal{T})$ is the quotient of a k-space which implies that $(\mathcal{M}_\tau(X), \mathcal{T})$ is a k-space by Theorem 3.3.23 in Engelking [1989]. □

Lemma 10.2.3. *The topology $\mathcal{T}$ is stronger than σ. Both topologies have the same compact sets and on the compact sets the topologies agree.*

Proof. For $f \in C_b(X)$, denote by $i_f : \mathcal{M}_\tau(X) \to \mathbb{R}$ the map defined by $i_f(\mu) = \int f \mathrm{d}\mu$. Because $\mathcal{T}$ is the final topology under the map q, i_f is continuous if and only if $i_f \circ q : \mathcal{M}_{\tau,+}(X) \times \mathcal{M}_{\tau,+}(X) \to \mathbb{R}$ is continuous. This, however, is clear as $i_f \circ q(\mu, \nu) = \int f \mathrm{d}(\mu - \nu)$ and the definition of the weak topology on $\mathcal{M}_{\tau,+}(X)$.

σ is the weakest topology making all i_f continuous, which implies that $\sigma \subseteq \mathcal{T}$.

For the second statement, note first that as $\sigma \subseteq \mathcal{T}$, the first has more compact sets. Thus, suppose that $K \subseteq \mathcal{M}_\tau(X)$ is σ compact. By Proposition 10.1.6 K is β equi-continuous, so by Theorem 6.1 (c) in Sentilles [1972], $K \subseteq K_1 - K_2$, where $K_1, K_2 \subseteq \mathcal{M}_{\tau,+}(X)$ and where K_1, K_2 are σ_+ and hence σ compact. It follows that $q(K_1, K_2)$ is $\mathcal{T}$ compact. Because K is a closed subset of $q(K_1, K_2)$, it is $\mathcal{T}$ compact. We conclude that the σ and $\mathcal{T}$ compact sets coincide.

Let K be a $\mathcal{T}$ and σ compact set. The identity map $i : K \to K$ is $\mathcal{T}$ to σ continuous, so it maps compacts to compacts. Because all closed sets are compact, i is homeomorphic, which implies that σ and $\mathcal{T}$ coincide on the compact sets. □

Proposition 10.2.4. *$\mathcal{T}$ is the k-ification of σ. In other words, $\mathcal{T}$ is the finest topology that coincides with σ on all σ compact sets. In particular, we find that $\mathcal{T} = \sigma^f$.*

Proof. By Lemma 10.2.2, $\mathcal{T}$ is a k-space. By Lemma 10.2.3 the compact sets for σ and $\mathcal{T}$ coincide. It follows that $\mathcal{T} = k\sigma = \sigma^f$. □

We prove an additional lemma that will yield transseparability of $(C_b(X), \beta)$, before moving on to the study of the quotient topology $\mathcal{T}$.

Lemma 10.2.5. *The σ, or equivalently, $\mathcal{T}$ compact sets in $\mathcal{M}_\tau(X)$ are metrizable.*

Proof. Let K be a σ compact set in $\mathcal{M}_\tau(X)$. In the proof of Lemma 10.2.3, we saw that $K \subseteq q(K_1, K_2)$, where K_1, K_2 are compact sets of the metrizable space $\mathcal{M}_{\tau,+}(X)$. Because q is a continuous map, we find that $q(K_1, K_2)$ and hence K is metrizable by Lemma 1.2 in Kalton [1971] or 34.11.2 in Köthe [1979]. □

Proof of Lemma 10.1.8. The σ compact sets are metrizable by Lemma 10.2.5, which implies that $(C_b(X), \beta)$ is transseparable by Lemma 1 in Pfister [1976]. $\square$

10.2.1 $(\mathcal{M}_\tau(X), \mathcal{T})$ *is a locally convex space.*

This section will focus on proving that the topology $\mathcal{T}$ on $\mathcal{M}_\tau(X)$ turns $\mathcal{M}_\tau(X)$ into a locally convex space. Given the identification $\mathcal{T} = k\sigma = \sigma^f$ obtained in Propositions 10.1.6 and 10.2.4, this is the main ingredient for the proof of Theorem 10.1.7. Indeed, for a general locally convex space the topology σ^f is in general not a vector space topology, cf. Section 2 in Kōmura [1964].

Proposition 10.2.6. *$(\mathcal{M}_\tau(X), \mathcal{T})$ is a locally convex space.*

The proof of the proposition relies on two lemma's.

Lemma 10.2.7. *The map $q : (\mathcal{M}_{\tau,+}(X)^2, \sigma_+^2) \to (\mathcal{M}_\tau(X), \mathcal{T})$ is an open map.*

Proof. Before we start proving that the map q is open, we start with two auxiliary steps.

Step 1. We first prove that the map $\oplus : (\mathcal{M}^2_{\tau,+}(X) \times \mathcal{M}_\tau(X), \sigma_+^2 \times \sigma) \to (\mathcal{M}_\tau^2(X), \sigma^2)$, defined by $\oplus(\mu, \nu, \rho) = (\mu + \rho, \nu + \rho)$ is open.

It suffices to show that $\oplus(V)$ is open for V in a basis for $\sigma_+^2 \times \sigma$ by Theorem 1.1.14 in Engelking [1989]. Hence, choose A and B be open for σ_+ and C open for σ. Set $U := \oplus(A \times B \times C)$. Choose $(\mu, \nu) \in U$. We prove that there exists an open neighbourhood of (μ, ν) contained in U. Because $(\mu, \nu) \in U = \oplus(A \times B \times C)$, we find $\mu_0 \in A, \nu_0 \in B$ and $\rho_0 \in C$ such that $\mu = \mu_0 + \rho_0$ and $\nu = \nu_0 + \rho_0$.

Because σ is the topology of a topological vector space, the sets $\mu_0 + C$ and $\nu_0 + C$ are open for σ. Thus, the set $H := (\mu_0 + C) \times (\nu_0 + C)$ is open for σ^2. By construction $(\mu, \nu) \in H$, and additionally, $H \subseteq U = \oplus(A \times B \times C)$.

We conclude that $\oplus$ is an open map.

Step 2. Denote $G := \oplus^{-1}(\mathcal{M}_{\tau,+}(X)^2)$ and by $\oplus_r : G \to \mathcal{M}_{\tau,+}(X)^2$ the restriction of $\oplus$ to the inverse image of $\mathcal{M}_{\tau,+}(X)^2$. If we equip G with the subspace topology inherited from $(\mathcal{M}^2_{\tau,+}(X) \times \mathcal{M}_\tau(X), \sigma_+^2 \times \sigma)$, the map $\oplus_r$ is open by Proposition 2.1.4 in Engelking [1989] and the openness of $\oplus$.

Step 3: The proof that q is open.

Let V be an arbitrary open set in $(\mathcal{M}_{\tau,+}(X)^2, \sigma_+^2)$. As a consequence, $V \times \mathcal{M}_\tau(X)$ is open in $(\mathcal{M}_{\tau,+}^2(X) \times \mathcal{M}_\tau(X), \sigma_+^2 \times \sigma)$. By definition of the subspace topology, $(V \times \mathcal{M}_\tau(X)) \cap G$ is open for the subspace topology on G. By the openness of $\oplus_r$, we conclude that $\hat{V} := \oplus_r((V \times \mathcal{M}_\tau(X)) \cap G)$ is open in $(\mathcal{M}_{\tau,+}(X)^2, \sigma_+^2)$.

Because $\oplus_r((V \times \mathcal{M}_\tau(X)) \cap G) = \oplus(V \times \mathcal{M}_\tau(X)) \cap \mathcal{M}_{\tau,+}(X)^2$, we find that

$$\begin{aligned} \hat{V} &= \left\{(\mu,\nu) \in \mathcal{M}_{\tau,+}(X)^2 \,\middle|\, \exists \rho \in \mathcal{M}_\tau(X) :\ (\mu - \rho, \nu - \rho) \in V\right\} \\ &= \left\{(\mu,\nu) \in \mathcal{M}_{\tau,+}(X)^2 \,\middle|\, \exists \rho \in \mathcal{M}_\tau(X) :\ (\mu + \rho, \nu + \rho) \in V\right\}. \end{aligned}$$

Thus, we see that $\hat{V} = q^{-1}(q(V))$. Because $\hat{V}$ is open and q is a quotient map, we obtain that $q(V)$ is open. □

Lemma 10.2.8. *The map $q^2 : (\mathcal{M}_{\tau,+}(X)^4, \sigma_+^4) \to (\mathcal{M}_\tau(X)^2, \mathcal{T}^2)$, defined as the product of q times q, i.e.*

$$q^2(\nu_1^+, \nu_1^-, \nu_2^+, \nu_2^-) = (\nu_1^+ - \nu_1^-, \nu_2^+ - \nu_2^-),$$

is an open map. As a consequence, $\mathcal{T}^2$ is the quotient topology of σ_+^4 under q^2.

Proof. By Proposition 2.3.29 in Engelking [1989] the product of open surjective maps is open. Thus, q^2 is open as a consequence of Lemma 10.2.7. An open surjective map is always a quotient map by Corollary 2.4.8 in Engelking [1989]. □

Proof of Proposition 10.2.6. We start by proving that $(\mathcal{M}_\tau(X) \times \mathcal{M}_\tau(X), \mathcal{T}^2) \to (\mathcal{M}_\tau(X), \mathcal{T})$ defined by $+(\nu_1, \nu_2) = \nu_1 + \nu_2$ is continuous. Consider the following spaces and maps:

$$\begin{array}{ccc} (\mathcal{M}_\tau(X) \times \mathcal{M}_\tau(X), \mathcal{T}^2) & \xrightarrow[+]{\qquad} & (\mathcal{M}_\tau(X), \mathcal{T}) \\ {\scriptstyle q^2}\big\uparrow & & \big\uparrow{\scriptstyle q} \\ (\mathcal{M}_{\tau,+}(X)^4, \sigma^4) & \xrightarrow[+_2]{\qquad} & (\mathcal{M}_{\tau,+}(X)^2, \sigma_+^2) \end{array}$$

q and $+$ are the quotient and sum maps defined above. q^2 was introduced in Lemma 10.2.8 and $+_2$ is defined as

$$+_2(\nu_1^+, \nu_1^-, \nu_2^+, \nu_2^-) = (\nu_1^+ + \nu_2^+, \nu_1^- + \nu_2^-).$$

Note that the diagram commutes, i.e. $q \circ +_2 = + \circ q^2$.

Fix an open set U in $(\mathcal{M}_\tau(X), \mathcal{T})$, we prove that $+^{-1}(U)$ is $\mathcal{T}^2$ open in $\mathcal{M}_\tau(X) \times \mathcal{M}_\tau(X)$. By construction, q is continuous. Also, $+_2$ is continuous as it is the restriction of the addition map on a locally convex space. We obtain that $V := +_2^{-1}(q^{-1}(U)) = (q^2)^{-1}(+^{-1}(U))$ is σ_+^4 open. By Lemma 10.2.8 q^2 is a quotient map, which implies that $+^{-1}(U)$ is $\mathcal{T}^2$ open. We conclude that $+ : (\mathcal{M}_\tau(X)^2, \mathcal{T}^2) \to (\mathcal{M}_\tau(X), \mathcal{T})$ is continuous.

We proceed by proving that the scalar multiplication map $m : (\mathcal{M}_\tau(X) \times \mathbb{R}, \mathcal{T} \times t) \to (\mathcal{M}_\tau(X), \mathcal{T})$ defined by $m(\mu, \alpha) = \alpha\mu$ is continuous. Here, t denotes the usual topology on $\mathbb{R}$. Consider the following diagram:

$$\begin{array}{ccc}
(\mathcal{M}_\tau(X) \times \mathbb{R}, \mathcal{T} \times t) & \xrightarrow[m]{} & (\mathcal{M}_\tau(X), \mathcal{T}) \\
{\scriptstyle q \times I}\big\uparrow & & \big\uparrow{\scriptstyle q} \\
(\mathcal{M}_{\tau,+}(X)^2 \times \mathbb{R}, \sigma_+^2 \times t) & \xrightarrow[m_2]{} & (\mathcal{M}_\tau(X)^2, \sigma_+^2)
\end{array}$$

Here, $I : \mathbb{R} \to \mathbb{R}$ denotes the identity map and $m_2 : \mathcal{M}_{\tau,+}(X)^2 \times \mathbb{R} \to \mathcal{M}_{\tau,+}^2(X)$ is defined by

$$m_2(\mu_1, \mu_2, \alpha) \begin{cases} (-\alpha\mu_2, -\alpha\mu_1) & \text{if } \alpha < 0 \\ (0, 0) & \text{if } \alpha = 0 \\ (\alpha\mu_1, \alpha\mu_2) & \text{if } \alpha > 0. \end{cases}$$

Note that, using this definition of m_2, the diagram above commutes. It is straightforward to verify that m_2 is a $\sigma_+^2 \times t$ to σ_+^2 continuous map as σ is the restriction of the topology of a topological vector space. By the Whitehead theorem, Theorem 3.3.7 in Engelking [1989], the map $q \times I$ is a quotient map. We obtain, as above, that m is continuous.

The continuity of $+$ and m yield that $(\mathcal{M}_\tau(X), \mathcal{T})$ is a topological vector space. To prove that the space is locally convex, we prove that $\mathcal{T}$ has a basis of open convex sets for 0.

Let $U \subseteq \mathcal{M}_\tau(X)$ be open and such that $0 \in U$, we prove that there is an open convex subset $U_0 \subseteq U$ such that $0 \in U_0$.

Because $q : (\mathcal{M}_{\tau,+}(X)^2, \sigma_+^2) \to (\mathcal{M}_\tau(X), \mathcal{T})$ is continuous, the set $q^{-1}(U)$ is σ_+^2 open. Additionally, $q^{-1}(U)$ contains $(0, 0)$. By construction of σ_+, there exists a σ^2 open set $V \subseteq \mathcal{M}_\tau(X)^2$ that contains $(0,0)$ and such that

$$V \cap \mathcal{M}_{\tau,+}(X)^2 = q^{-1}(U).$$

Because $(\mathcal{M}_\tau(X)^2, \sigma^2)$ is locally convex, we can find a σ^2 open convex neighbourhood $V_0 \subseteq V$ of 0. By Lemma 10.2.7 q is open, additionally it is linear on its domain, thus we find that

$$U_0 := q(V_0 \cap \mathcal{M}_{\tau,+}(X)^2) \subseteq U$$

is $\mathcal{T}$ open and convex. By construction, U_0 contains 0.

We conclude that $(\mathcal{M}_\tau(X), \mathcal{T})$ is a locally convex space. □

10.2.2 *The proof of Theorem 10.1.7 and its corollaries*

We finalize with the proof of our main result and its consequences.

Proof of Theorem 10.1.7. We already noted that $k\sigma = \sigma^f$ by Proposition 10.1.6.

By Proposition 10.2.4, we find $\mathcal{T} = \sigma^f$. By Proposition 10.2.6 $\mathcal{T}$ is locally convex. Because σ^{lf} is the strongest locally convex topology coinciding with σ on all weakly compact sets, we conclude by Proposition 10.2.6 that $\sigma^{lf} = \sigma^l$.

By Proposition 10.1.2 the space $(C_b(X), \beta)$ is hypercomplete, and thus, complete. It follows by 21.9.8 in Köthe [1969] that $\sigma^{lf} = \beta^\circ$. □

Proof of Corollary 10.1.11. By Theorem 10.1.7 and 10.1.2, we obtain that $(C_b(Y), \beta)$ is an infra-Pták space. By Lemma 10.1.8 $(C_b(X), \beta)$ is transseparable and by Lemma 10.1.9 $\mathcal{M}_\tau(X)$ is weakly sequentially complete.

The result, thus, follows from Kalton's closed graph theorem 10.1.4. □

Proof of Corollary 10.1.12. Let X, Y be separably metrizable spaces. Let $T : (C_b(X), \beta) \to (C_b(Y), \beta)$ be a bijective continuous linear map. We prove that $T^{-1} : (C_b(Y), \beta) \to (C_b(X), \beta)$ is continuous.

The graph of a continuous map is always closed. Therefore, the graph of T^{-1} is also closed. The result follows now from the closed graph theorem. □

Proof of Corollary 10.1.13. Let X, Y be separably metrizable spaces. Let $T : (C_b(X), \beta) \to (C_b(Y), \beta)$ be a surjective continuous linear map. We prove that T is open.

First, note that the quotient map $\pi : (C_b(X), \beta) \to (C_b(X)/ker\, T, \beta_\pi)$ is open, where β_π is the quotient topology obtained from β, see 15.4.2 Köthe [1969]. The map T factors into $T_\pi \circ \pi$, where T_π is a bijective continuous linear map from $(C_b(X)/ker\, T, \beta_\pi)$ to $(C_b(Y), \beta)$.

We show that T_π is an open map. We can apply the inverse mapping theorem to T_π as $(C_b(X)/ker\, T, \beta_\pi)$ is a Pták space by 34.3.2 in Köthe [1979]. Additionally, it is transseparable as it is the uniformly continuous image of a transseparable space. It follows that T_π^{-1} is continuous and that T_π is open.

We find that the composition $T = T_\pi \circ \pi$ is open as it is the composition of two open maps. □

BIBLIOGRAPHY

Stefan Adams, Nicolas Dirr, Mark Peletier, and Johannes Zimmer. Large deviations and gradient flows. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371 (2005), 2013. ISSN 1364-503X. doi: 10.1098/rsta.2012.0341.

Angela Albanese and Franziska Kühnemund. Trotter-Kato approximation theorems for locally equicontinuous semigroups. *Riv. Mat. Univ. Parma (7)*, 1:19–53, 2002.

Angela Albanese and Elisabetta Mangino. Trotter-Kato theorems for bi-continuous semigroups and applications to Feller semigroups. *Journal of Mathematical Analysis and Applications*, 289(2):477 – 492, 2004. doi: 10.1016/j.jmaa.2003.08.032.

Angela Albanese, José Bonet, and Werner. Ricker. Mean ergodic semigroups of operators. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 106(2):299–319, 2012. ISSN 1578-7303. doi: 10.1007/s13398-011-0054-2.

Rami Atar and Paul Dupuis. Large deviations and queueing networks: Methods for rate function identification. *Stochastic Processes Appl.*, 84 (2):255–296, 1999. ISSN 0304-4149. doi: 10.1016/S0304-4149(99)00051-4.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. ISBN 978-3-319-00226-2; 978-3-319-00227-9. doi: 10.1007/978-3-319-00227-9.

T.A. Beatty and H.H. Schaefer. Semi-bornological spaces. *Mathematische Zeitschrift*, 221(1):337–351, 1996. ISSN 0025-5874. doi: 10.1007/BF02622119.

Sergey. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006. ISSN 0894-9840. doi: 10.1007/s10959-006-0016-3.

Vladimir I. Bogachev. *Measure Theory*. Springer-Verlag, 2007.

Vivek S. Borkar and Rajesh Sundaresan. Asymptotics of the invariant measure in mean field models with jumps. *Stoch. Syst.*, 2(2):322–380, 2012. doi: 10.1214/12-SSY064.

Djehiche Boualem and Kaj Ingemar. The rate function for some measure-valued jump processes. *The Annals of Probability*, 23(3):1414–1438, 1995. ISSN 00911798.

Ola Bratelli and Derek W. Robinson. *Operator Algebras and Quantum Statistical Mechanics I.* Springer-Verlag, 1979.

Amarjit Budhiraja, Paul Dupuis, and Vasileios Maroulas. Variational representations for continuous time processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 47(3):725–747, 08 2011. doi: 10.1214/10-AIHP382.

Amarjit Budhiraja, Paul Dupuis, and Markus Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.*, 40(1):74–102, 2012. doi: 10.1214/10-AOP616.

Amarjit Budhiraja, Paul Dupuis, Markus Fischer, and Kavita Ramanan. Limits of relative entropies associated with weakly interacting particle systems. *Electron. J. Probab.*, 20:no. 80, 1–22, 2015a. ISSN 1083-6489. doi: 10.1214/EJP.v20-4003.

Amarjit Budhiraja, Paul Dupuis, Markus Fischer, and Kavita Ramanan. Local stability of Kolmogorov forward equations for finite state nonlinear Markov processes. *Electron. J. Probab.*, 20:no. 81, 1–30, 2015b. ISSN 1083-6489. doi: 10.1214/EJP.v20-4004.

Robert C. Busby. Double centralizers and extensions of C^*-algebras. *Trans. Amer. Math. Soc.*, 132:79–99, 1968. ISSN 0002-9947.

Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, volume 58. Springer Science & Business Media, 2004.

Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 45(3):734–753, 2009. ISSN 0246-0203. doi: 10.1214/08-AIHP183.

P. Pérez Carreras and J. Bonet. *Barrelled Locally Convex Spaces.* North-Holland Publishing Co., Amsterdam, 1987.

Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 12 1952. doi: 10.1214/aoms/1177729330.

F. Comets. Large deviation estimates for a conditional probability distribution. applications to random interaction Gibbs measures. *Probability Theory and Related Fields*, 80(3):407–432, 1989.

Francis Comets. Nucleation for a long range magnetic model. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 23(2):135–178, 1987.

John B. Conway. *A Course in Functional Analysis.* Springer, second edition, 2007.

J.B. Cooper. *Saks spaces and applications to functional analysis.* North-Holland Publishing Co., Amsterdam, second edition, 1987. ISBN 0-444-70219-9.

Cristina Costantini and Thomas Kurtz. Viscosity methods giving uniqueness for martingale problems. *Electron. J. Probab.*, 20:no. 67, 1–27, 2015. ISSN 1083-6489. doi: 10.1214/EJP.v20-3624.

M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):pp. 265–298, 1971.

M. G. Crandall, L. C. Evans, and P.-L. Lions. Some properties of viscosity solutions of Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 282(2): 487–502, 1984. ISSN 0002-9947. doi: 10.2307/1999247.

Michael G. Crandall. A generalization of Peano's existence theorem and flow invariance. *Proc. Amer. Math. Soc.*, 36:151–155, 1972. ISSN 0002-9939.

Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc., New Ser.*, 27(1):1–67, 1992. ISSN 0273-0979; 1088-9485/e. doi: 10.1090/S0273-0979-1992-00266-5.

Paolo Dai Pra and Frank den Hollander. Mckean-Vlasov limit for interacting random processes in random media. *Journal of Statistical Physics*, 84 (3-4):735–772, 1996. ISSN 0022-4715. doi: 10.1007/BF02179656.

Donald A. Dawson and Shui Feng. Large deviations for the fleming-viot process with neutral mutation and selection 1. *Stochastic Processes and their Applications*, 77(2):207 – 232, 1998. ISSN 0304-4149. doi: 10.1016/S0304-4149(98)00035-0.

Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

Benjamin Dembart. On the theory of semigroups of operators on locally convex spaces. *Journal of Functional Analysis*, 16(2):123 – 160, 1974. doi: 10.1016/0022-1236(74)90061-5.

Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications.* Springer-Verlag, second edition, 1998.

Frank den Hollander, Frank Redig, and Willem van Zuijlen. Gibbs-non-Gibbs dynamical transitions for mean-field interacting Brownian motions. *Stochastic Processes and their Applications*, 125(1):371 – 400, 2015. ISSN 0304-4149. doi: http://dx.doi.org/10.1016/j.spa.2014.09.011.

Xiaoxue Deng, Jin Feng, and Yong Liu. A singular 1-D Hamilton-Jacobi equation, with application to large deviation of diffusions. *Communications in Mathematical Sciences*, 9(1), 2011.

Jean-Dominique Deuschel and Daniel W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics.* Academic Press, Inc., Boston, MA, 1989. ISBN 0-12-213150-9.

J. Diestel and J. J. Uhl, Jr. *Vector measures.* American Mathematical Society, 1977.

J. L. Doob. *Classical potential theory and its probabilistic counterpart*, volume 262 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, New York, 1984. ISBN 0-387-90881-1. doi: 10.1007/978-1-4612-5208-5.

Johannes J Duistermaat. On global action-angle coordinates. *Communications on pure and applied mathematics*, 33(6):687–706, 1980.

Paul Dupuis and Richard S. Ellis. The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc.*, 347(8):2689–2751, 1995. ISSN 0002-9947. doi: 10.2307/2154753.

Paul Dupuis, Hitoshi Ishii, and H. Mete Soner. A viscosity solution approach to the asymptotic analysis of queueing systems. *The Annals of Probability*, 18(1):pp. 226–255, 1990. ISSN 00911798.

Paul Dupuis, Kavita Ramanan, and Wei Wu. Large deviation principle for finite-state mean field interacting particle systems. *preprint; ArXiv:1601.06219*, 2016.

Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations.* Springer-Verlag, 2000.

Ryszard Engelking. *General topology.* Heldermann Verlag, Berlin, second edition, 1989.

Aernout C.D. Enter, Roberto Fernández, and Alan D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory. *Journal of Statistical Physics*, 72(5-6):879–1167, 1993.

Matthias Erbar and Jan Maas. Ricci curvature of finite markov chains via convexity of the entropy. *Archive for Rational Mechanics and Analysis*, 206(3):997–1038, 2012. ISSN 0003-9527. doi: 10.1007/s00205-012-0554-z.

Victor Ermolaev and Christof Külske. Low-temperature dynamics of the Curie-Weiss model: Periodic orbits, multiple histories, and loss of Gibbsianness. *Journal of Statistical Physics*, 141(5):727–756, 2010. ISSN 0022-4715. doi: 10.1007/s10955-010-0074-x.

Alison M. Etheridge. *An introduction to superprocesses.* Providence, RI: AMS, American Mathematical Society, 2000. ISBN 0-8218-2706-5.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and Convergence.* Wiley, 1986.

Franco Fagnola. Quantum Markov semigroups and quantum flows. *Proyecciones*, 18(3):144, 1999. ISSN 0716-0917.

Bálint Farkas. Perturbations of bi-continuous semigroups with applications to transition semigroups on Cb(H). *Semigroup Forum*, 68(1):87–107, 2004. ISSN 0037-1912. doi: 10.1007/s00233-002-0024-2.

Bálint Farkas. Adjoint bi-continuous semigroups and semigroups on the space of measures. *Czechoslovak Mathematical Journal*, 61(2):309–322, 2011. ISSN 0011-4642. doi: 10.1007/s10587-011-0076-0.

Jin Feng and Thomas G. Kurtz. *Large Deviations for Stochastic Processes.* American Mathematical Society, 2006.

Shui Feng. Large deviations for empirical process of mean-field interacting particle system with unbounded jumps. *Annals of Probability*, 22(4):2122–2151, 1994.

R. Fernández, F. den Hollander, and J. Martínez. Variational description of Gibbs-non-Gibbs dynamical transitions for the Curie-Weiss model. *Communications in Mathematical Physics*, 319(3):703–730, 2013. ISSN 0010-3616. doi: 10.1007/s00220-012-1646-1.

Hans Föllmer and Nina Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Probab.*, 25(2):901–926, 04 1997. doi: 10.1214/aop/1024404423.

M.I. Freidlin and A.D. Wentzell. *Random perturbations of dynamical systems.* Springer-Verlag, second edition, 1998.

Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics.* Walter de Gruyter & Co., Berlin, second edition, 2011. ISBN 978-3-11-025029-9. doi: 10.1515/9783110250329.

Seymour Goldberg. Some properties of the space of compact operators on a Hilbert space. *Math. Ann.*, 138:329–331, 1959.

R. G. Haydon. On the Stone-Weierstrass theorem for the strict and superstrict topologies. *Proc. Amer. Math. Soc.*, 59(2):273–278, 1976.

Benton Jamison. The Markov processes of Schrödinger. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(4):323–331, 1975. doi: 10.1007/BF00535844.

Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. II.* Academic Press, Inc., Orlando, FL, 1986. doi: 10.1016/S0079-8169(08)60611-X.

N. J. Kalton. Some forms of the Closed graph theorem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 70:401–408, 1971. ISSN 1469-8064. doi: 10.1017/S0305004100050039.

J. L. Kelley. Hypercomplete linear topological spaces. *Michigan Math. J.*, 5 (2):235–246, 1958. doi: 10.1307/mmj/1028998069.

Takako Kōmura. Semigroups of operators in locally convex spaces. *J. Functional Analysis*, 2(3):258–296, 1968.

Yukio Kōmura. Some examples on linear topological spaces. *Mathematische Annalen*, 153(2):150–162, 1964. ISSN 0025-5831. doi: 10.1007/BF01361183.

Gottfried Köthe. *Topological Vector Spaces I.* Springer-Verlag, 1969.

Gottfried Köthe. *Topological Vector Spaces II.* Springer-Verlag, 1979.

Richard Kraaij. Stationary product measures for conservative particle systems and ergodicity criteria. *Electron. J. Probab.*, 18:no. 88, 1–33, 2013. ISSN 1083-6489. doi: 10.1214/EJP.v18-2513.

Richard Kraaij. Large deviations of the trajectory of empirical distributions of Feller processes on locally compact spaces. *preprint; ArXiv:1401.2802*, 2014.

Richard Kraaij. Strongly continuous and locally equi-continuous semigroups on locally convex spaces. *Semigroup Forum*, 92(1):158–185, 2016a. ISSN 1432-2137. doi: 10.1007/s00233-015-9689-1.

Richard Kraaij. Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton–Jacobi equation. *Journal of Statistical Physics*, 164(2): 321–345, 2016b. ISSN 1572-9613. doi: 10.1007/s10955-016-1542-8.

Richard Kraaij. A Banach-Dieudonné theorem for the space of bounded continuous functions on a separable metric space with the strict topology. *Topology and its Applications*, 2016c. ISSN 0166-8641. doi: 10.1016/j.topol.2016.06.003.

Franziska Kühnemund. A Hille-Yosida theorem for bi-continuous semigroups. *Semigroup Forum*, 67(2):205–225, 2003.

Christof Külske and Arnaud Le Ny. Spin-flip dynamics of the Curie-Weiss model: Loss of Gibbsianness with possibly broken symmetry. *Communications in Mathematical Physics*, 271(2):431–454, 2007. ISSN 0010-3616. doi: 10.1007/s00220-007-0201-y.

Markus Kunze. Continuity and equicontinuity of semigroups on norming dual pairs. *Semigroup Forum*, 79(3):540–560, 2009.

Markus Kunze. A Pettis-type integral and applications to transition semigroups. *Czechoslovak Mathematical Journal*, 61(2):437–459, 2011. ISSN 0011-4642. doi: 10.1007/s10587-011-0065-3.

Thomas G. Kurtz. Convergence of sequences of semigroups of nonlinear operators with an application to gas kinetics. *Trans. Am. Math. Soc.*, 186: 259–272, 1974. ISSN 0002-9947; 1088-6850/e. doi: 10.2307/1996565.

A. Le Ny and F. Redig. Large deviation principle at fixed time in Glauber evolutions. *Markov Process. Related Fields*, 10(1):65–74, 2004.

Christian Léonard. Large deviations for long range interacting particle systems with jumps. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 31(2):289–323, 1995.

Christian Léonard. On the convexity of the entropy along entropic interpolations. *preprint; ArXiv:1310.1274*, 2013.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4): 1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533.

Thomas M. Liggett. *Interacting Particle Systems.* Springer-Verlag, 1985.

Torgny Lindvall. *Lectures on the coupling method.* Dover Publications, Inc., Mineola, NY, 1992.

A. Mielke, M.A. Peletier, and D.R.M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. *Potential Analysis*, 41(4):1293–1327, 2014. ISSN 0926-2601. doi: 10.1007/s11118-014-9418-5.

Sunao Ōuchi. Semi-groups of operators in locally convex spaces. *Journal of the Mathematical Society of Japan*, 25(2):265–276, 04 1973. doi: 10.2969/jmsj/02520265.

Zbigniew Palmowski and Tomasz Rolski. A technique for exponential change of measure for Markov processes. *Bernoulli*, 8(6):767–785, 2002.

Etienne Pardoux and Aurel Răşcanu. *Stochastic differential equations, backward SDEs, partial differential equations*, volume 69 of *Stochastic Modelling and Applied Probability.* Springer, Cham, 2014. ISBN 978-3-319-05713-2; 978-3-319-05714-9. doi: 10.1007/978-3-319-05714-9.

Edwin Perkins. Dawson-Watanabe superprocesses and measure-valued diffusions. In *Lectures on probability theory and statistics. Ecole d'eté de probabilités de Saint-Flour XXIX - 1999, Saint-Flour, France, July 8–24, 1999*, pages 125–329. Berlin: Springer, 2002. ISBN 3-540-43736-3.

Charles-Edouard Pfister. Thermodynamical aspects of classical lattice systems. In Vladas Sidoravicius, editor, *In and Out of Equilibrium*, volume 51 of *Progress in Probability*, pages 393–472. Birkhäuser Boston, 2002. doi: 10.1007/978-1-4612-0063-5_18.

Helmut Pfister. Bemerkungen zum satz über die separabilität der Fréchet-Montel-räume. *Archiv der Mathematik*, 27(1):86–92, 1976. ISSN 0003-889X. doi: 10.1007/BF01224645.

Firas Rassoul-Agha and Timo Seppäläinen. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2015. ISBN 978-0-8218-7578-0.

F. Redig and F. Wang. Transformations of one-dimensional Gibbs measures with infinite range interaction. *Markov Process. Relat. Fields*, 16(4):737–752, 2010. ISSN 1024-2953.

A.P. Robertson and W. Robertson. *Topological Vector Spaces*. Cambridge University Press, second edition, 1973.

Wojciech De Roeck, Christian Maes, and Karel Netočný. H-theorems from macroscopic autonomous equations. *Journal of Statistical Physics*, 123 (3):571–584, 2006. ISSN 0022-4715. doi: 10.1007/s10955-006-9079-x.

Stephen A. Saxon and L. M. Sánchez Ruiz. Dual local completeness. *Proc. Amer. Math. Soc.*, 125(4):1063–1070, 1997. ISSN 0002-9939. doi: 10.1090/S0002-9939-97-03864-1.

M. Schilder. Some asymptotic formulas for Wiener integrals. *Transactions of the American Mathematical Society*, 125(1):63–85, 1966. ISSN 00029947.

George R. Sell. Smooth linearization near a fixed point. *American Journal of Mathematics*, 107(5):1035–1091, 1985.

F. Dennis Sentilles. Bounded continuous functions on a completely regular space. *Trans. Amer. Math. Soc.*, 168:311–336, 1972.

Timo Seppäläinen. Large deviations for lattice systems i. parametrized independent fields. *Probability Theory and Related Fields*, 96(2):241–260, 1993. ISSN 0178-8051. doi: 10.1007/BF01192135.

Daniel W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients, i. *Communications on Pure and Applied Mathematics*, 22 (3):345–400, 1969a. ISSN 1097-0312. doi: 10.1002/cpa.3160220304.

Daniel W. Stroock and S. R. S. Varadhan. Diffusion processes with continuous coefficients, ii. *Communications on Pure and Applied Mathematics*, 22(4):479–530, 1969b. ISSN 1097-0312. doi: 10.1002/cpa.3160220404.

Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin-New York, 1979. ISBN 3-540-90353-4.

W. H. Summers. Separability in the strict and substrict topologies. *Proc. Amer. Math. Soc.*, 35:507–514, 1972. ISSN 0002-9939.

Masamichi Takesaki. *Theory of Operators Algebras I*. Springer-Verlag, New York-Heidelberg, 1979.

Donald Curtis Taylor. The strict topology for double centralizer algebras. *Transactions of the American Mathematical Society*, 150(2):pp. 633–643, 1970.

François Treves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press, New York-London, 1967.

H. F. Trotter. Approximation of semi-groups of operators. *Pacific J. Math.*, 8:887–919, 1958. ISSN 0030-8730.

A.C.D. van Enter, R. Fernández, F. den Hollander, and F. Redig. A large-deviation view on dynamical Gibbs-non-Gibbs transitions. *Moscow Mathematical Journal*, 10:687–711, 2010.

Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. ISBN 978-3-540-71049-3. doi: 10.1007/978-3-540-71050-9. Old and new.

Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy and ricci curvature. *Communications on Pure* *and Applied Mathematics*, 58(7):923–940, 2005. ISSN 1097-0312. doi: 10.1002/cpa.20060.

J. H. Webb. Sequential convergence in locally convex spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 64:341–364, 4 1968.

Albert Wilansky. *Modern methods in topological vector spaces.* McGraw-Hill International Book Co., New York, 1978. ISBN 0-07-070180-6.

Albert Wilansky. Mazur spaces. *Int. J. Math. Math. Sci.*, 4:39–53, 1981. doi: 10.1155/S0161171281000021.

A. Wiweger. Linear spaces with mixed topology. *Studia Math.*, 20:47–68, 1961.

K Yosida. *Functional Analysis.* Springer-Verlag, Berlin-New York, fifth edition, 1978.

SUMMARY

This thesis is dedicated to the study of large deviations of trajectories of Markov processes and functional analytic methods that facilitate the study of the associated semigroups of a Markov process.

In the first chapters, we focus on large deviation principles for the trajectories of averages of mean-field interacting Markov processes and applications thereof. In Chapter 3, we consider variants of Glauber dynamics for the Curie-Weiss model. As an example, we consider n processes $(\sigma^1(t), \ldots, \sigma^n(t))$ on $\{-1, 1\}$ that interact via their mean. Denote $x_n(t) = \frac{1}{n}\sum_{i\leq n}\sigma^i(t)$. If the mean $x_n(0)$ satisfies the large deviation principle

$$\mathbb{P}\left[x_n(0) \approx \alpha\right] \approx e^{-nI_0(\alpha)},$$

for some rate function I_0, then we prove under appropriate conditions that the same holds for the whole trajectory of averages. In particular, we prove that

$$\mathbb{P}\left[\{x_n(t)\}_{t\geq 0} \approx \{\gamma(t)\}_{t\geq 0}\right] \approx e^{-nI(\gamma)},$$

for $\gamma : [0, \infty) \to [-1, 1]$. I takes the form

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{L} : [-1, 1] \times \mathbb{R} \to [0, \infty]$ is some non-negative lower semi-continuous function and $\mathcal{AC}$ is the space of absolutely continuous trajectories in $[-1, 1]$.

We use this large deviation framework to in Chapter 4 to find a natural Lyapunov function for the associated McKean-Vlasov dynamics and study the exponential decay of this Lyapunov function.

In Chapter 5, we use the associated Hamilton equations to study the trajectories that optimize

$$I_t(a) = \inf_{\substack{\gamma\in\mathcal{AC} \\ \gamma(t)=a}} I_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\mathrm{d}s,$$

which is of importance in the study of Gibbs-non-Gibbs transitions.

In Chapter 6, we prove the large deviation principle with a rate function in Lagrangian form for the trajectories of the empirical distribution

$$\frac{1}{n}\sum_{i\leq n}\delta_{X^i(t)}$$

of independent copies $X^1, X^2, \dots$ of a Feller process X that takes its values in a locally compact metric space.

In Chapter 7, we consider a translation invariant Markov process $\sigma(t)$ on $\{-1,1\}^{\mathbb{Z}^d}$ that evolves by spin-flip dynamics. Set $\Lambda_n = \mathbb{Z}^d \cap [-n,n]^d$. In this setting, we prove the large deviation principle for trajectories of the empirical measure

$$\frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n}\delta_{\theta_i\sigma(t)},$$

where θ_i is the operation that shifts a configuration $(\theta_i\eta)_j = \eta_{i+j}$. We do not prove a Lagrangian representation for the rate function, but make a conjecture on such a form.

In Chapter 8, we consider strongly continuous and locally equi-continuous semigroups on a special class of locally convex spaces. We prove a Hille-Yosida and Trotter-Kato approximation theorem. We show that this theory can be applied to the strict topology on the space of bounded continuous functions on a Polish space.

In Chapter 9, we show that the transition semigroup of a Markov process on a Polish space that is the solution to a well-posed martingale process is strongly continuous and locally equicontinuous for the strict topology on the space of bounded and continuous functions.

Finally, we show in Chapter 10 that the strict topology on the space of bounded and continuous functions on a separable metric space satisfies the conclusion of the Banach-Dieudonné theorem. In particular, this implies that the closed graph, open mapping and inverse mapping theorems hold for maps between spaces of this kind.

SAMENVATTING (DUTCH SUMMARY)

Dit proefschrift bestudeert de grote afwijkingen van trajecten van Markovprocessen en de functionaal analytische methoden die gebruikt worden voor de studie van de semigroepen die corresponderen met een Markovproces.

In de eerste hoofdstukken behandelen we de grote afwijkening van de trajecten van gemiddeldes van 'mean-field' interagerende Markov processen en kijken naar de toepassingen van de resultaten. In hoofdstuk 3 bestuderen we varianten van Glauber dynamica voor het Curie-Weiss model. Als voorbeeld bekijken we n processen $(\sigma^1(t), \dots, \sigma^n(t))$ op $\{-1, 1\}$ die interageren via hun gemiddelde. We schrijven $x_n(t) = \frac{1}{n}\sum_{i \leq n} \sigma^i(t)$ voor het gemiddelde van deze variabelen. Als het gemiddelde $x_n(0)$ van de variabelen op tijd 0 aan het grote afwijkingen principe voldoet, i.e.,

$$\mathbb{P}\left[x_n(0) \approx \alpha\right] \approx e^{-nI_0(\alpha)},$$

voor een gegeven 'rate' functie I_0, dan bewijzen we, onder geschikte voorwaarden, dat ook het hele traject van de gemiddeldes aan het grote afwijkingen principe voldoet. In het bijzonder bewijzen we dat

$$\mathbb{P}\left[\{x_n(t)\}_{t\geq 0} \approx \{\gamma(t)\}_{t \geq 0}\right] \approx e^{-nI(\gamma)},$$

waar $\gamma : [0,\infty) \to [-1,1]$. De rate functie I is van de vorm

$$I(\gamma) = \begin{cases} I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s & \text{if } \gamma \in \mathcal{AC}, \\ \infty & \text{otherwise,} \end{cases}$$

waar $\mathcal{L} : [-1,1] \times \mathbb{R} \to [0,\infty)$ een beneden half-continue functie is.

We gebruiken het grote afwijkingen principe in hoofdstuk 4 om een natuurlijke Lyapunov functie te vinden voor de geassocieerde McKean-Vlasov dynamica. We geven condities waaronder deze Lyapunov functie exponentieel snel afvalt onder deze dynamica.

In hoofdstuk 5 gebuiken we de Hamilton vergelijkingen om trajecten te bestuderen die optimal zijn in de volgende variationele uitdrukking:

$$I_t(a) = \inf_{\substack{\gamma \in \mathcal{AC} \\ \gamma(t) = a}} I_0(\gamma(0)) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) \mathrm{d}s. \qquad (\star)$$

Deze optimale trajecten zijn van belang in de studie van Gibbs-niet-Gibbs overgangen in de zin dat 'slechte' configuraties a, i.e. essentië discontinuiteiten van de conditionele waarschijnlijkheden, overeenkomen met het bestaan van meerdere optimale trajecten in ($\star$).

In hoofdstuk 6 bewijzen we het grote afwijkingen principe met een rate-functie in Lagrangiaanse vorm voor de trajecten van de empirische distributie

$$\frac{1}{n}\sum_{i\leq n}\delta_{X^i(t)}$$

van onafhankelijke kopieën $X^1, X^2, \ldots$ van een Feller proces X dat zijn waarden aanneemt in een locaal compacte metrische ruimte.

In hoofdstuk 7 bestuderen we een translatie invariant proces $\sigma(t)$ op $\{-1,1\}^{\mathbb{Z}^d}$ dat evolueert onder de invloed van spin-flip dynamica. We definiëren $\Lambda_n = \mathbb{Z}^d \cap [-n,n]^d$. In deze context bewijzen we het grote afwijkingen principe voor de trajecten van de emprische maat

$$\frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n}\delta_{\theta_i\sigma(t)}.$$

Hier is θ_i de operatie die een configuratie $\eta \in \{-1,1\}^{\mathbb{Z}^d}$ i plaatsen opschuift: $(\theta_i\eta)_j = \eta_{i+j}$. We bewijzen geen Lagrangiaanse representatie van de rate functie, maar formuleren een vermoeden over wat deze represenatie zou moeten zijn.

In hoofdstuk 8 bestuderen we sterk continue en locaal equicontinue semigroepen op een speciale classe van locaal convexe ruimten. We bewijzen in deze context een Hille-Yosida en Trotter-Katto approximatiestelling. We laten zien dat deze theorie gebruikt kan worden voor de strikte topologie op de ruimte van begrensde continue functies op een Poolse ruimte.

In hoofdstuk 9 laten we zien dat de semigroep van conditionele waarschijnlijkheden van een Markovproces op een Poolse ruimte dat de oplossing is van een martingaal probleem sterk continu en locaal equicontinu is op de ruimte van continue begrensde functies.

Tot slot bewijzen we in hoofdstuk 10 dat de strikte topologie op de ruimte van continue begrensde functies op een separabele metrische ruimte voldoet aan de conclusies van de Banach-Dieudonné stelling. Als gevolg verkrijgen we dat de gesloten graaf, open afbeelding en inverse afbeelding stelling gelden voor lineare afbeeldingen tussen ruimtes van dit type.

ACKNOWLEDGEMENTS

My first and foremost thanks go to Frank for his support over the last years. Frank, from the moment you taught me the first course on interacting particle systems, I learned a tremendous amount of mathematics. Also outside the realm of math, you set an example by the way you reason and challenge established ideas. But most of all, I want to thank you for supporting my wandering mind and the opportunity to find my own path.

I want to thank Christian Maes, Christof Külske, Jin Feng, André Schlichting and Michiel Renger for supporting visits to Leuven, Bochum, Lyon, Bonn and Berlin. I want to specifically thank Jin for explaining many subtleties and techniques regarding solving Hamilton-Jacobi equations and proving large deviation principles which makes me understand so much more.

Closer to home, I want to thank the support staff of the department for helping me out with various thinks like organizing reading groups or arranging trips abroad. Of course, I want to thank my fellow PhD students, but also the post-docs and research staff of the probability and analysis departments for stimulating discussions and their support over the past few years. Chiara, Nick, Ivan and Jan, thanks for the many hours spent discussing the various connections between probability and functional analysis. Henk, Christos, Martina, Nenad and Moritz, I enjoyed the time and good atmosphere in our HB06.260 and the countless number of hours spent on discussing various topics in and outside of mathematics. In particular, I want to thank Moritz for discussions over an almost encyclopedic range of topics: algorithms for rock-paper-scissors, transformations of diffusion and Lévy bridges, the time-evolution of stochastic orderings and the pros and cons of various path-space topologies.

Dear Lun-Yi, thank you for allowing me to use the beautiful art that features on the cover of this thesis.

Fleur, Baer en Marissa, bedankt voor een onwankelbaar vertrouwen in een goede afloop van de promotie en de niet aflatende strijd om de geest scherp te houden. Pap, mam, jullie wil ik bedanken voor de steun en het vertrouwen bij het inslaan van dit pad, en de basis waarop altijd teruggevallen kan worden. Thom, het is een waar genoegen om het met jou over wiskunde te hebben. De combinatie van je achtergrond en kennis met je positie buiten de academische wereld geeft je een unieke blik waar ik op meerdere vlakken veel van leer. Bedankt dat je deze jongen met twee benen op de grond houdt.

Of het nou het kopje thee of een bakje fruit is bij werk in het weekend, de aansporing om te gaan sporten na een stressvolle dag, of de blik vanuit een ander academisch perspectief, het heeft mij deze jaren geholpen. Jorine, bedankt voor je steun.

CURRICULUM VITAE

Richard Kraaij was born on the 23rd of March 1989 in Tönistvorst, Germany. He completed his pre-university education in 2007 at 'Valuascollege' in Venlo. From 2007 to 2010, he studied for his B.Sc. degree in Mathematics at the Radboud University in Nijmegen. From 2010 to 2012, he studied at the Free University of Amsterdam and wrote a thesis on 'stationary measures for conservative particle systems' under the supervision of Prof. dr. R.W.J. Meester and Prof. dr. F.H.J. Redig. He obtained his M.Sc. degree 'cum laude' in May 2012. For a couple of months in 2012, he worked on a PhD project under the supervision of Prof. dr. R.W. van der Hofstad at the Technical University of Eindhoven. In October 2012, he started his Phd research funded by the Dutch Science Foundation NWO under the supervision of Prof. dr. F.H.J. Redig at the Technical University of Delft.

PUBLICATIONS

Publications:

Richard Kraaij. Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton–Jacobi equation. *Journal of Statistical Physics*, 164(2):321–345, 2016b. ISSN 1572-9613. doi: 10.1007/s10955-016-1542-8

Richard Kraaij. A Banach-Dieudonné theorem for the space of bounded continuous functions on a separable metric space with the strict topology. *Topology and its Applications*, 2016c. ISSN 0166-8641. doi: 10.1016/j.topol.2016.06.003

Richard Kraaij. Strongly continuous and locally equi-continuous semigroups on locally convex spaces. *Semigroup Forum*, 92(1):158–185, 2016a. ISSN 1432-2137. doi: 10.1007/s00233-015-9689-1

Richard Kraaij. Stationary product measures for conservative particle systems and ergodicity criteria. *Electron. J. Probab.*, 18:no. 88, 1–33, 2013. ISSN 1083-6489. doi: 10.1214/EJP.v18-2513

Preprint:

Richard Kraaij. Large deviations of the trajectory of empirical distributions of Feller processes on locally compact spaces. *preprint; ArXiv:1401.2802*, 2014