4 Analysis of Transfer Operators

In the preceding section we have presented an algorithmic approach to the identification of metastable subsets under the two conditions (C1) and (C2), which are functional analytical statements on the spectrum of the propagator P_τ. In this section, we want to transform these conditions into a more probabilistic language, which will result in establishing equivalent conditions on the stochastic transition function. For general Markov processes it is natural to consider P_τ acting on $L^1(\mu)$, the Banach space that includes all probability densities w.r.t. μ. Yet, for *reversible* Markov processes it is advantageous to restrict the analysis to $L^2(\mu)$, since the propagator will then be self–adjoint. Therefore, in the first two sections we start with analyzing the conditions (C1) and (C2) in $L^1(\mu)$, while in the third section, we then concentrate on $L^2(\mu)$. For convenience we use the abbreviation $P = P_\tau$ and $p(x, C) = p(\tau, x, C)$ for some fixed time $\tau > 0$. As a consequence, $P^n = P_{n\tau}$ corresponds to the Markov process sampled at rate τ with stochastic transition function given by $p^n(\cdot, \cdot) = p(n\tau, \cdot, \cdot)$. The results presented in this section mainly follow [**?**].

4.1 The Spectrum and its Parts

Consider a complex Banach space E with norm $\|\cdot\|$ and denote the spectrum[6] of a bounded linear operator $P : E \to E$ by $\sigma(P)$. For an eigenvalue $\lambda \in \sigma(P)$, the multiplicity of λ is defined as the dimension of the generalized eigenspace; see e.g., [**?**, Chap. III.6]. Eigenvalues of multiplicity 1 are called simple. The set of all eigenvalues $\lambda \in \sigma(P)$ that are isolated and of finite multiplicity is called the **discrete spectrum**, denoted by $\sigma_{\text{discr}}(P)$. The **essential spectral radius** $r_{\text{ess}}(P)$ of P is defined as the smallest real number, such that outside the ball of radius $r_{\text{ess}}(P)$, centered at the origin, are only discrete eigenvalues, i.e.,

$$r_{\text{ess}}(P) \;=\; \inf\{r \geq 0 : \lambda \in \sigma(P) \text{ with } |\lambda| > r \text{ implies } \lambda \in \sigma_{\text{discr}}(P)\}.$$

This definition of $r_{\text{ess}}(P)$ is unusual in the sense that it does not involve any definition of the essential spectrum; yet, it is the way we will exploit $r_{\text{ess}}(P)$ and it will be justified by Theorem 4.1 below. Usually, the essential spectral radius is related to the smallest disc containing the entire essential spectrum $\sigma_{\text{ess}}(P)$ of P. Unfortunately, there are many different characterizations of essential spectra (see e.g., [**?**, **?**] or [**?**, Chapter 107]). The definition that results in the smallest set is due to Kato [**?**, Chapter IV.5.6] who defines $\sigma_{\text{ess}}^{\text{Kato}}(P)$ as the complement of $\{\lambda \in \mathbf{C} : \lambda - P \text{ is semi Fredholm}^7\}$. The

[6] For common functional analytical terminology see, e.g., [**?**, **?**, **?**, **?**].

[7] A bounded linear operator $P : E \to E$ on a Banach space E is said to be semi Fredholm, if its range $R(P) = \{y = Px : x \in E\}$ is closed and the dimension of its kernel

definition that results in the largest set is due to Browder [?] according to whom $\sigma_{\text{ess}}^{\text{Browder}}(P)$ is the complement of the discrete spectrum, as defined above. According to Lebow and Schechter [?] we get the surprising result that all other known definitions of essential spectra fall between those of Kato and Browder and lie inside the ball with radius $r_{\text{ess}}(P)$ centered at the origin:

Theorem 4.1 *For every bounded linear operator $P : E \to E$ on a complex Banach space E holds*

$$\sup\{|\lambda| : \lambda \in \sigma_{\text{ess}}^{Kato}(P)\} \;=\; r_{\text{ess}}(P) \;=\; \sup\{|\lambda| : \lambda \in \sigma_{\text{ess}}^{Browder}(P)\}.$$

Loosely speaking, Theorem 4.1 states that the essential spectral radius is invariant under the definition of the essential spectrum.

As a guiding example for our strategy to bound the essential spectral radius, consider the following semi–norm $\|\cdot\|_c$ defined by

$$\|P\|_c \;=\; \inf\{\|P - S\| : S \text{ compact}\}.$$

Then the essential spectral radius is characterized by

$$r_{\text{ess}}(P) \;=\; \lim_{n\to\infty} \|P^n\|_c^{1/n}.$$

Note the analogy to the spectral radius $r(P)$ of P, defined as the smallest upper bound for all elements of the spectrum: $r(P) = \sup\{|\lambda| : \lambda \in \sigma(P)\}$. In terms of the operator norm $\|\cdot\|_1$, the representation $r(P) = \lim \|P^n\|_1^{1/n}$ as $n \to \infty$ is well–known [?, Chap. VII.3.5]. The above characterization of $r_{\text{ess}}(P)$ is closely related to **quasi–compactness**:

Definition 4.2 ([?]) *A bounded linear operator $P : E \to E$ is called quasi–compact, if there exist some $m \in \mathbf{Z}_+$ and a compact operator $S : E \to E$ such that $\|P^m - S\| < 1$.*

Combining quasi–compactness with the characterization of $r_{\text{ess}}(P)$ yields:

Corollary 4.3 *For bounded linear operator $P : E \to E$ holds*

(i) if $r_{\text{ess}}(P) < 1$ then P is quasi–compact

(ii) if P is quasi–compact for some $m \in \mathbf{Z}_+$ and compact operator S with $\|P^m - S\| = 1 - \eta < 1$, then $r_{\text{ess}}(P) \le (1-\eta)^{1/m} < 1$.

$N(P) = \{x \in E : Px = 0\}$ or the codimension of its range, i.e., $\dim E/R(P)$, are finite [?, Chapter IV.5]. If both, the dimension of the kernel and the codimension of the range are finite, then P is called a *Fredholm operator*.

We conclude that the essential spectral radius can be bound by using compact operators:

> Find for some power P^m with $m \in \mathbf{Z}_+$ a decomposition into a compact part S and the remaining part $P^m - S$. Then, we have the upper bound: $r_{\mathrm{ess}}(P) \leq \|P^m - S\|^{1/m}$.

In other words, the "larger" the compact part of P^m is, the smaller the essential spectral radius of P will be. Our goal is to relate compactness of S to properties of the stochastic transition function that defines P. Due to the various possible definitions of essential spectra, this approach might not be restricted to compact operators. This is indeed the case, as we will see below. The crucial point will be to find the class of operators that fits best both the Banach space as well as the propagator and the Markov process. In $L^1(\mu)$ weakly compact operators are better adapted for our purpose, while in $L^2(\mu)$ the compact ones will do a good job. This is basically due to the fact that in either case we can characterize the property of being (weakly) compact in terms of the underlying probability space, which finally enables us to relate bounds on the essential spectral radius to properties of the stochastic transition function. For relations between the essential spectral radius and measures of non–compactness, see [**?**, **?**].

Spectral conditions can be quite sensitive to the Banach space of functions the operator is regarded to act on. This is illustrated by the following example due to Davies [**?**, Chapter 4.3].

Example 4.4 *Consider the Smoluchowski equation*

$$\dot{q} \quad = \quad -q + \dot{W} \tag{41}$$

on the state space $\mathbf{X} = \mathbf{R}$. *It corresponds to the harmonic potential* $V(q) = q^2/2$ *with* $\gamma = \sigma = 1$ *and invariant probability measure*

$$\mu_Q(\mathrm{d}q) \quad = \quad \frac{1}{Z}\exp(-q^2)\mathrm{d}q.$$

The Markov process defined by (41) is known as the Ornstein–Uhlenbeck process [?]. The evolution of densities $v = v(t,q)$ *w.r.t.* μ_Q *is governed by the Fokker–Planck equation*

$$\partial_t v \quad = \quad \underbrace{\Big(\frac{1}{2}\Delta - q \cdot \nabla_q\Big)}_{\mathcal{L}} v, \tag{42}$$

which defines a strongly continuous contraction semigroup $P_t = \exp(t\mathcal{L})$ *on* $L^r(\mu)$ *for every* $1 \leq r < \infty$. *The spectra of* $\mathcal{L}$ *and* P_t *have the following properties:*

(i) *In* $L^1(\mu)$ *it is* $\sigma(\mathcal{L}) = \{z \in \mathbf{C} : Re(z) \leq 0\}$, *with every* $z \in \sigma(\mathcal{L})$ *satisfying* $Re(z) < 0$ *being an eigenvalue of multiplicity two. This implies for the propagator that*

$$\sigma(P_t) \quad = \quad \{z \in \mathbf{C} : |z| \leq 1\},$$

with every $z \in \sigma(P_t)$ *satisfying* $|z| < 1$ *being an eigenvalue of infinite multiplicity, hence* $r_{\text{ess}}(P_t) = 1$.

(ii) *In* $L^2(\mu)$ *it is* $\sigma(\mathcal{L}) = \{z \in \mathbf{C} : z = 0, -1, -2, \dots\}$, *with the nth Hermite polynomial being the eigenfunction corresponding to* $\lambda_n = -n$. *Hence, the entire spectrum is discrete. This implies for the propagator*

$$\sigma(P_t) \quad = \quad \{z \in \mathbf{C} : z = e^{-tn} \text{ for } n = 0, 1, 2, \dots\},$$

with $r_{\text{ess}}(P_t) = 0$.

From a numerical point of view, we would like to consider the space of functions that is "generated" by the discretization procedure for finer and finer decompositions of the state space. This, however, is believed to be a very tough question.

4.2 Bounds on the Essential Spectral Radius in $L^1(\mu)$

This section analyzes the essential spectral radius of an arbitrary propagator $P : L^1(\mu) \to L^1(\mu)$ in terms of its stochastic transition function. In doing so, weakly compact operators will play an important role. The main result is stated in Theorem 4.13, which relates the essential spectral radius, uniform constrictiveness and a certain Doeblin–condition.

Definition 4.5 ([?, ?]) *A bounded linear operator* $S : L^1(\mu) \to L^1(\mu)$ *is called* ***weakly compact*** *if it maps the closed unit ball* $B_1(X)$ *onto a relatively weakly compact set, i.e., the closure of* $S(B_1(X))$ *is compact in the weak topology.*

Obviously, every compact operator is weakly compact; the converse is not true. The next theorem characterizes the essential spectral radius of an arbitrary bounded linear operator in terms of weakly compact operators.

Theorem 4.6 ([?, ?]) *Let* $P : L^1(\mu) \to L^1(\mu)$ *denote a bounded linear operator. Define the semi–norm* $\Delta(P)$ *according to*

$$\Delta(P) \quad = \quad \min\{\|P - S\|_1 : S \text{ is weakly compact}\}.$$

Then the essential spectral radius of P *is characterized by*

$$r_{\text{ess}}(P) \quad = \quad \lim_{n\to\infty} \Delta(P^n)^{1/n}. \tag{43}$$

In particular, $r_{\text{ess}}(P) \leq \Delta(P)$.

The theorem states that the larger the weakly compact part of P is, the less the essential spectral radius will be. Hence, good upper bounds on $r_{\text{ess}}(P)$ require a detailed analysis of weak compactness. It should be clear from the introductory statements of this section that we could also apply Corollary 4.3 to characterize the essential spectral radius in $L^1(\mu)$ by compact operators. The utility of weakly compact operators will become apparent by the next theorem that relates this particular class of operators to the underlying measure space $(\mathbf{X}, \mathcal{A}, \mu)$.

Theorem 4.7 ([?, ?]) *Let $P : L^1(\mu) \to L^1(\mu)$ denote a bounded linear operator. Then*

$$\Delta(P) \quad = \quad \limsup_{\mu(A)\to 0} \|\mathbf{1}_A \circ P\|_1, \tag{44}$$

where the limit is understood to be taken over all sequences of subsets whose μ–measure converges to zero, and $\mathbf{1}_A$ is interpreted as a multiplication operator: $(\mathbf{1}_A v)(x) = \mathbf{1}_A(x)v(x)$. In particular,

$$\limsup_{\mu(A)\to 0} \|\mathbf{1}_A \circ P\|_1 \quad = \quad 0,$$

if and only if P is weakly compact.

As a consequence of Theorem 4.7, we will deduce in the following that absolutely continuous stochastic transition functions may give rise to weakly compact operators, while transition functions that are singular w.r.t. μ never do so. This will finally enable us to characterize the essential spectral radius in terms of properties of the stochastic transition function.

Corollary 4.8 *Consider some propagator $S : L^1(\mu) \to L^1(\mu)$ defined by*

$$Sv(y) \quad = \quad \int_{\mathbf{X}} v(x)p(x,y)\mu(\mathrm{d}x) \tag{45}$$

associated with some absolutely continuous *stochastic transition function $p(x,\mathrm{d}y) = p(x,y)\mu(\mathrm{d}y)$. Then S is weakly compact if there exits some $s > 1$ such that $\|p(x,\cdot)\|_s \in L^\infty(\mu)$ as a function of x, i.e.,*

$$\operatorname*{ess\,sup}_{x\in\mathbf{X}} \int_{\mathbf{X}} p(x,y)^s \mu(\mathrm{d}y) \quad < \quad \infty$$

holds. In particular, S is weakly compact if $\operatorname{ess\,sup}_{x,y\in\mathbf{X}} p(x,y) < \infty$.

Proof: For $A \in \mathcal{B}(\mathbf{X})$, we have

$$\|\mathbf{1}_A \circ S\|_1 \quad = \quad \sup_{\|v\|_1 \le 1} \int_A \int_{\mathbf{X}} v(x)p(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

Applying Hölder's inequality twice, we finally get

$$\|\mathbf{1}_A \circ S\|_1 \;\; \leq \;\; \operatorname*{ess\,sup}_{x \in \mathbf{X}} \int_A p(x,y)\mu(\mathrm{d}y) \;\; \leq \;\; \|\mathbf{1}_A\|_r \;\; \operatorname*{ess\,sup}_{x \in \mathbf{X}} \|p(x,\cdot)\|_s$$

with $1 \leq r, s \leq \infty$ and $1/s + 1/r = 1$. For $1 < s$, the limit of $\|\mathbf{1}_A \circ S\|_1$ as $\mu(A) \to 0$ tends to zero, since $\|\mathbf{1}_A\|_r = \sqrt[r]{\mu(A)}$. $\square$

For analyzing propagators corresponding to not necessarily absolutely continuous stochastic transition functions, consider the **Lebesgue decomposition** of $p(x, \mathrm{d}y) = p_a(x,y)\mu(\mathrm{d}y) + p_s(x, \mathrm{d}y)$, where p_a and p_s represent the absolutely continuous and the singular part w.r.t. μ, respectively [**?**]. Furthermore, define the (not necessarily stochastic) transition function

$$r_n(x,y) = \begin{cases} p_a(x,y) & \text{if } p_a(x,y) \geq n \\ 0 & \text{otherwise} \end{cases} .$$

With this notation, we are ready to state the important

Theorem 4.9 ([?]) *For an arbitrary propagator $P : L^1(\mu) \to L^1(\mu)$ the equality*

$$\Delta(P) \;\; = \;\; \inf_{n \in \mathbf{Z}_+} \; \operatorname*{ess\,sup}_{x \in \mathbf{X}} \{r_n(x, \mathbf{X}) + p_s(x, \mathbf{X})\}$$

holds.

In the particular case, where p_a gives rise to a weakly compact operator, Theorem 4.9 states that

$$\Delta(P) \;\; = \;\; \operatorname*{ess\,sup}_{x \in \mathbf{X}} p_s(x, \mathbf{X}) \;\; = \;\; 1 - \operatorname*{ess\,inf}_{x \in \mathbf{X}} \int_{\mathbf{X}} p_a(x,y)\mu(\mathrm{d}y).$$

If only some decomposition $P = R + S$ with weakly compact S is known, we may still apply Theorem 4.6 to get an upper bound on $\Delta(P)$. Assume that the stochastic transition function can be decomposed according to $p(x, \mathrm{d}y) = p_R(x, \mathrm{d}y) + p_W(x, \mathrm{d}y)$ such that S, defined via $Sv(y) = \int_{\mathbf{X}} v(x) p_W(x, \mathrm{d}y)$, is weakly compact. Then

$$\Delta(P) \;\; \leq \;\; \operatorname*{ess\,sup}_{x \in \mathbf{X}} p_R(x, \mathbf{X}) \;\; \leq \;\; 1 - \operatorname*{ess\,inf}_{x \in \mathbf{X}} p_W(x, \mathbf{X})$$

by Theorem 4.6. Using one of the inequalities involving $\Delta(P)$, we are able to bound the essential spectral radius due to Theorem 4.6. This is illustrated by the following example due to Schütte [**?**, Chapter 4.1].

Example 4.10 *Consider the Hamiltonian system with randomized momenta for the harmonic potential $V(q) = q^2/2$ on some position space $\Omega \subset \mathbf{R}$ with* *inverse temperature β and positional canonical distribution μ_Q. Choose the observation time span $\tau = 2\pi$ and decompose the stochastic transition function according to*

$$p_\tau(q, \mathrm{d}y) \quad = \quad p_a(q,y)\mu_Q(\mathrm{d}y) + p_s(q, \mathrm{d}y)$$

into an absolutely continuous and singular part w.r.t. μ_Q. Depending on the position space, we distinguish two cases

(i) Consider $\Omega = \mathbf{R}$, the bounded system case. Since $\tau = 2\pi$ is the period of the harmonic oscillator, we deduce that $P_\tau = \mathrm{Id}$ and hence $r_{\mathrm{ess}}(P_\tau) = 1$. In terms of the stochastic transition function this means that $p_a = 0$ and $p_s(q, \mathrm{d}y) = \delta_q(\mathrm{d}y)$ for every $q \in \Omega$.

(ii) Consider $\Omega = [-1, 1]$ with periodic boundary conditions. It can be shown that in this case the density p_a is bounded and satisfies

$$\inf_{q\in\Omega} \int_\Omega p_a(q,y)\mu_Q(\mathrm{d}y) \quad = \quad 2\,\Phi(-\sqrt{\beta})$$

where Φ denotes the distribution function of the standard normal distribution. Setting

$$\eta \quad = \quad 2\Phi(-\sqrt{\beta}) \;=\; 2\big(1 - \Phi(\sqrt{\beta})\big)$$

we have $0 \le \eta \le 1$ and finally $r_{\mathrm{ess}}(P_\tau) \le \Delta(P_\tau) = 1 - \eta$ due to[8] Theorem 4.9. In other words, the (upper bound on the) essential spectral radius depends on the inverse temperature and therefore on the mean energy of the ensemble. The lower the mean energy (and hence the higher the inverse temperature) is, the larger the essential spectral radius will be. This corresponds to the intuition that the periodic system behaves more and more like the bounded system for decreasing mean energy.

So far we have shown how to prove $r_{\mathrm{ess}}(P) < 1$ in terms of the stochastic transition function p. The properties imposed on p emerged from functional analytical requirements on the propagator P. We now link these results to the theory of Markov processes and Markov operators. An important property of Markov operators is constrictiveness [**?**]; it rules out the possibility that for some initial density v the iterates $P^n v$ eventually concentrate on a set of very small or vanishing measure.

[8] For the propagator regarded to act on $L^2(\mu_Q)$ the stronger statement $r_{\mathrm{ess}}(P_\tau) = 1 - \eta$ is proved in [**?**].

Definition 4.11 *A propagator $P : L^1(\mu) \to L^1(\mu)$ is called constrictive if there exist constants $\epsilon, \delta > 0$ such that for every* density $v \in L^1(\mu)$ *there exists $m = m(v) \in \mathbf{Z}_+$ with*

$$\mu(A) \le \epsilon \quad \Rightarrow \quad \int_A P^n v(y)\mu(\mathrm{d}y) \le 1 - \delta, \tag{46}$$

*for every $n \ge m$. We call a propagator **uniformly constrictive** if there exists $m \in \mathbf{Z}_+$ such that (46) holds for $n \ge m$ uniformly in $L^1(\mu)$.*

For arbitrary $v \in L^1(\mu)$, uniform constrictiveness can be restated as $\mu(A) \le \epsilon \Rightarrow \|\mathbf{1}_A \circ P^n\|_1 \le 1 - \delta$ for every $n \ge m$. Moreover, it is sufficient to assume that the condition holds for $n = m$ only, since due to $\|P^k\|_1 = 1$ for $k \in \mathbf{Z}_+$ this already implies (46) for all $n \ge m$. In view of the characterization of $\Delta(P)$ in (44), uniform constrictiveness seems to be closely related to $\Delta(P) < 1$ and thus to some bound on the essential spectral radius; this is indeed the case, as we will see below. Furthermore, there should exist a similar condition involving the backward transfer operator T. This, in turn, is closely related to the *Doeblin–condition*, which is well–known in the theory of Markov processes [**?**, **?**, **?**]. It states that there exists a probability measure ν, constants $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$ such that $\nu(A) \le \epsilon \Rightarrow \sup_{x \in \mathbf{X}} \; p^m(x, A) \le 1 - \delta$. To suit our context, we slightly adapt the Doeblin–condition in the way that we require $\nu = \mu$ and that the implication holds for μ–a.e. points only:

Definition 4.12 *The stochastic transition function p is said to fulfill the **μ-a.e. Doeblin–condition** if there exist constants $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$ such that*

$$\mu(A) \le \epsilon \quad \Rightarrow \quad p^m(x, A) \le 1 - \delta \tag{47}$$

for μ–a.e. $x \in \mathbf{X}$ and every $A \in \mathcal{B}(\mathbf{X})$.

Using the backward transfer operator, we deduce that (47) is equivalent to $\mu(A) < \epsilon \Rightarrow \|T^m \mathbf{1}_A\|_\infty = \operatorname{ess\,sup}_{x \in \mathbf{X}} p^m(x, A) \le 1 - \delta$. In fact, the condition is true for all $n \ge m$, since $\|T^{m+k}\mathbf{1}_A\|_\infty \le \|T^k\|_\infty \|T^m \mathbf{1}_A\|_\infty$ and $\|T^k\|_\infty = 1$ holds for $k \ge 1$. The next theorem states the main result of this section. It relates the functional–analytical, the Markov operator theoretical and the Markov process theoretical point of view.

Theorem 4.13 *Let $P : L^1(\mu) \to L^1(\mu)$ denote the propagator corresponding to a stochastic transition function $p : \mathbf{X} \times \mathcal{B}(\mathbf{X}) \to [0, 1]$. Then, the following statements are equivalent:*

(i) The essential spectral radius of P is less than one: $r_{\mathrm{ess}}(P) < 1$.

(ii) The propagator P is uniformly constrictive.

(iii) The stochastic transition functions fulfills the μ-a.e. Doeblin–condition.

If conditions (ii) or (iii) are satisfied for some $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$, then condition (i) holds with $r_{\text{ess}}(P) \leq (1-\delta)^{1/m}$.

Proof: Assume (i) holds, i.e., $r_{\text{ess}}(P) < 1$. Due to Eqs. (43) and (44), there exists $m \in \mathbf{Z}_+$ such that $\Delta(P^m) < 1$, which implies the μ-a.e. Doeblin–condition (4.12) due to $\|\mathbf{1}_A \circ P^n\|_1 = \|T^n \mathbf{1}_A\|_\infty$ (see Lemma 4.1 in [**?**]). As just stated, (iii) is equivalent to (ii). Using the note following Def. 4.11, it is obvious that (ii) and (i) are equivalent. The bound on $r_{\text{ess}}(P)$ follows from (43) and (44). □

In view of the established equivalence, the essential spectral radius is related to the possibility of the system to eventually concentrate on a set of small or vanishing measure. In other words, the more the dynamics is smeared over the entire state space, the less is the essential spectral radius, while irregular or singular behavior may give rise to a large essential spectral radius.

4.3 Peripherical Spectrum and Properties in $L^1(\mu)$

This section analyzes the peripherical spectrum and its relation to properties of propagators P acting on $L^1(\mu)$. Due to our particular interest—cf. condition (C1)—we restrict the analysis to uniformly constrictive propagators, i.e., we assume that $r_{\text{ess}}(P) < 1$. We will see that under this assumption the peripherical spectrum completely characterizes the asymptotic properties of P, as it is known from the finite dimensional case.

Recall that we require throughout this thesis that the probability measure μ is invariant w.r.t. the Markov process. This is equivalent to the condition $P\mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$. A subset $E \subset X$ is called non–null if $\mu(E) > 0$. A non–null subset $E \subset X$ is called **invariant** if $P\mathbf{1}_E = \mathbf{1}_E$. Parts of the following two theorems are scattered over the literature see, e.g., [**?**, **?**, **?**].

Theorem 4.14 (Invariant Decomposition) *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator. Then*

(i) there are only finitely many eigenvalues $\lambda \in \sigma_{\text{discr}}(P)$ with $|\lambda| = 1$, each being a root of unity. The dimension of each eigenspace is finite and equal to the multiplicity of the corresponding eigenvalue;

(ii) the eigenvalue $\lambda = 1$ is of multiplicity d, if and only if there exists a decomposition of the state space

$$\mathbf{X} \;\; = \;\; E_1 \cup \cdots \cup E_d \cup F$$

into d mutually disjoint invariant subsets E_j and a set $F = \mathbf{X} \setminus \bigcup_j E_j$ of μ–measure zero.

Proof: Direct application of Thm. 4.13 of this thesis and Thm. 3 of [**?**, VIII.8] proves the first part. For the second statement, we exploit the fact that $Pv = v$ implies $Pv^+ = v^+$ and $Pv^- = v^-$, where $v^{+/-}$ denotes the positive or negative part of v, respectively [**?**]. Assume that the multiplicity of $\lambda = 1$ is d. Then, as a consequence of the first part, there exist d linear independent eigenfunctions $v_1, \dots, v_d$. Due to the decomposition result for v, we can also choose d linear independent *densities*, which we again denote by $v_1, \dots, v_d$. We now show that the densities can be chosen in such a way that their supports $E_j = \mathrm{supp}(v_j)$ are mutually disjoint, i.e., $\mu(E_j \cap E_k) = 0$ for $j \neq k$. If for some choice of linear independent densities $v_1, \dots, v_d$ there exist v_j, v_k such that $\mu(E_j \cap E_k) > 0$, we simply substitute v_j, v_k by $(v_j - v_k)^+, (v_j - v_k)^-$. This is possible, since $\mathrm{span}\{(v_j - v_k)^+, (v_j - v_k)^-\} = \mathrm{span}\{v_j, v_k\}$ and $\mathrm{span}\{(v_j - v_k)^+, (v_j - v_k)^-, v_j, v_k\} > 2$ would be in contradiction to the fact that the multiplicity of $\lambda = 1$ is d. Due to $P\mathbf{1}_\mathbf{X} = \mathbf{1}_\mathbf{X}$, we have $v_j = \mathbf{1}_{E_j}/\mu(E_j)$ and $\sum_j \mu(E_j) = 1$. Finally, define $F = \mathbf{X} \setminus \bigcup_j E_j$. Since any decomposition into d mutually disjoint invariant subsets results in a multiplicity of $\lambda = 1$ of at least d, the second statement is proved. □

The decomposition of the state space given by the theorem is unique up to μ–equivalence. There is an analogous decomposition result for the stochastic transition function p, since for every invariant subset E the identity

$$\mu(E) \;=\; \int_E \mathbf{1}_E(y)\mu(\mathrm{d}y) \;=\; \int_E P\mathbf{1}_E(y)\mu(\mathrm{d}y) \;=\; \int_E p(x,E)\mu(\mathrm{d}x)$$

implies $p(x,E) = 1$ for μ–a.e. $x \in E$. Thus, the decomposition of Theorem 4.14 induces a decomposition of the stochastic transition function, which again is unique up to μ–equivalence. For a "strong" decomposition holding everywhere see, e.g., [**?**]. For some root of unity $\omega = \exp(2\pi i/m)$ with $m \in \mathbf{Z}_+$, we call $\sigma_{\mathrm{cycle}}(\omega) = \{\omega, \omega^2, \dots, \omega^m\}$ an **eigenvalue cycle** associated with ω. A further subdecomposition of an invariant subset E into m mutually disjoint, non–null subsets $\{E_1, \dots, E_m\}$ is called a **subset cycle** of length m if $P\mathbf{1}_{E_j} = \mathbf{1}_{E_{j+1}}$ for $j = 1, \dots, m$ with the convention $E_{m+1} = E_1$. For the next theorem, an eigenvalue of multiplicity ν is interpreted as ν equal eigenvalues $\lambda_1, \dots, \lambda_\nu$ of multiplicity 1.

Theorem 4.15 (Cycle Decomposition) *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator. Then*

(i) each discrete eigenvalue $\lambda \in \sigma_{\mathrm{discr}}(P)$ of unit modulus is part of some eigenvalue cycle, i.e., there exists $m \in \mathbf{Z}_+$ such that $\lambda \in \sigma_{\mathrm{cycle}}(\omega)$ with $\omega = \exp(2\pi i/m)$;

(ii) there is a one–to–one correspondence between eigenvalue cycles and subset cycles. More precisely, let d denote the multiplicity of $\lambda = 1$.

Then the set of all eigenvalues of unit modulus can be decomposed into d eigenvalue cycles $\sigma_{\text{cycle}}(\omega_j)$ with $\omega_j = \exp(2\pi i/m_j)$, $m_j \in \mathbf{Z}_+$ and $j = 1, \ldots, d$, if and only if the state space $\mathbf{X}$ can be decomposed into d subset cycles $\{E_{j\,1}, \ldots, E_{j\,m_j}\}$ of length m_j for $j = 1, \ldots, d$.

Proof: Use Theorem 4.14 of this thesis and Theorem 11 in [**?**], which also holds in our case, to show that each invariant subset E can be decomposed into a subset cycles $\{E_1, \ldots, E_m\}$ of length m. Consider the restricted propagator $P_E = \mathbf{1}_E \circ P \circ \mathbf{1}_E$, which is well–defined by Theorem 4.14. Then, the length m is equal to the multiplicity of $\lambda = 1$ of ${P_E}^m$. Thus, it remains to show that $\sigma(P_E) \cap \{|\lambda| = 1\} = \sigma_{\text{cycle}}(\omega)$ with $\omega = \exp(2\pi i/m)$. But every subset cycle $\{E_1, \ldots, E_m\}$ of P is also a subset cycle of P_E and allows us to define m linear independent eigenfunctions $v_{k+1} = \sum_{j=0}^{m-1} \omega^{-kj} P_E^j \mathbf{1}_{E_1}$, see e.g. [**?**], which correspond to the eigenvalues ω^k for $k = 1, \ldots, m$. This completes the proof. □

From a functional analytical point of view, the decomposition results are related to a partial spectral decomposition of P, as we will see in the next result due to Dunford and Schwartz [**?**, Chapter VIII]. It exploits the fact that uniform constrictiveness is equivalent to quasi–compactness of the propagator (Thm. 4.13 and Cor. 4.3).

Theorem 4.16 (Spectral Decomposition) *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator and let Π_λ denote the spectral projection corresponding to the discrete eigenvalue λ. Then, for every $n \in \mathbf{Z}_+$,*

$$P^n \quad = \sum_{\lambda \in \sigma(P), |\lambda|=1} \lambda^n \Pi_\lambda + D^n$$

with some strict contraction $D : L^1(\mu) \to L^1(\mu)$ satisfying $\|D^n\|_1 \leq M q^n$ for some $M > 0$ and $0 < q < 1$. Furthermore, the projections fulfill

$$\Pi_\lambda \quad = \quad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\lambda^n} P^n, \tag{48}$$

where the limit is understood to be uniform.

Now, we exploit the above results to analyze properties of the propagator P and the underlying Markov process given by its stochastic transition function p.

Definition 4.17 *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator.*

*(i) P is said to be **ergodic** if every invariant subset E is of μ–measure 1. Equivalently, $P\mathbf{1}_E = \mathbf{1}_E$ implies $\mu(E) = 0$ or $\mu(E) = 1$.*

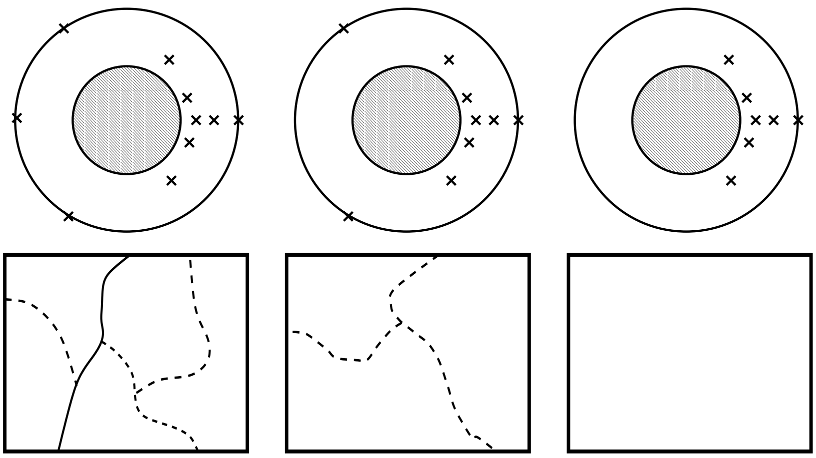

Figure 4: Top: idealized spectra of uniformly constrictive propagators. All eigenvalues are assumed to be simple except for $\lambda = 1$ in the left graphic, which must be at least two fold. Outer disc of radius $r = 1$ containing the entire spectrum and inner disc with radius $r_{\text{ess}} < 1$ containing the essential spectrum. Bottom: decomposition of the state space (rectangle) into invariant sets (separated by solid lines) and subset cycles (separated by dashed lines) corresponding to the spectra above. Left: two eigenvalues cycles with $m = 3$ and $m = 2$, respectively resulting in a decomposition of the state space into two invariant subsets that can be further decomposed into subset cycles of length $m = 3$ and $m = 2$, respectively. Middle: one eigenvalue cycle with $m = 3$ resulting in a decomposition of the state space into a subset cycle of length $m = 3$. Right: The eigenvalue 1 is simple and dominant. Hence, there neither exists a decomposition of the state space into invariant subsets nor subset cylcles.

*(ii) P is called **periodic** with period p if it is ergodic and p is the largest integer for which a subset cycle of length p occurs. If $p = 1$, then P is called **aperiodic**.*

According to [**?**], a Markov operator $P : L^1(\mu) \to L^1(\mu)$ satisfying $P\mathbf{1_X} = \mathbf{1_X}$ is said to be ergodic if $P^n v$ converges in the sense of Cesàro for every density $v \in L^1(\mu)$ weakly to $\mathbf{1_X}$. Anticipating the results of the next corollary and using Thm. 5.5.1 from [**?**, Sec. 5.5], it can easily be shown that for uniformly constrictive propagators this definition is equivalent to Def. **??** (i). In the theory of Markov processes, the term ergodicity is used slightly different, since it requires aperiodicity. Corollary **??** may be used to establish the relation. The next corollary states how these properties are related to the decomposition results previously obtained.

Corollary 4.18 *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator. Then*

(i) P is ergodic if and only if the eigenvalue $\lambda = 1$ is simple.

(ii) P is aperiodic if and only if the eigenvalue $\lambda = 1$ is simple and dominant, i.e., $\eta \in \sigma(P)$ satisfying $|\eta| = 1$ implies $\eta = 1$.

Ergodicity is related to the fact that it is impossible to decompose the state space into independent parts. The analogue in the theory of Markov processes is *irreducibility* expressing that it is possible to move from (almost) every state to every "relevant" subset within a finite time:

Definition 4.19 ([?, ?]) *A stochastic transition function p is said to be* ***μ-a.e. irreducible*** *if*

$$\mu(A) > 0 \quad \Rightarrow \quad p^m(x, A) > 0 \tag{49}$$

*for μ–a.e. $x \in \mathbf{X}$, every $A \in \mathcal{B}(\mathbf{X})$ and some $m = m(x, A) \in \mathbf{Z}_+$. If (**??**) holds for every $x \in \mathbf{X}$ then p is called μ–irreducible.*

The next theorem relates the two statements about indecomposability:

Theorem 4.20 *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator corresponding to the stochastic transition function p. Then P is ergodic if and only if p is μ-a.e. irreducible.*

Proof: Due to the remark following Def. **??**, P is ergodic if and only if $P(\mathbf{1}_B/\mu(B))$ converges to $\mathbf{1_X}$ in the sense of Cesàro for every $B \in \mathcal{B}(\mathbf{X})$ with $\mu(B) > 0$. For arbitrary $A \in \mathcal{B}(\mathbf{X})$ with $\mu(A) > 0$ this is equivalent to

$$\begin{aligned}
& \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \int_{\mathbf{X}} P^k \mathbf{1}_B(y) \mathbf{1}_A(y)\, \mu(\mathrm{d}y) = \mu(A)\, \mu(B) \\
\Leftrightarrow \quad & \lim_{n\to\infty} \int_B \frac{1}{n}\sum_{k=1}^{n} p^k(y, A)\mu(\mathrm{d}y) = \int_B \mu(A)\mu(\mathrm{d}y) \\
\Leftrightarrow \quad & \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p^k(y, A) = \mu(A); \qquad \mu\text{–a.e.},
\end{aligned}$$

where we used Lebesgue's dominated convergence theorem. Since by assumption $\mu(A) > 0$, this is equivalent to μ-a.e. irreducibility according to Def. **??**. □

Often, we are interested in dynamical systems—deterministic or stochastic—that exhibit a *unique* invariant density and guarantee that for every initial density v the iterates $P^n v$ converge to the invariant density. In view of Corollary **??**, these systems are necessarily connected to ergodic propagators, but due to possible cyclic behavior, ergodicity is not sufficient.

Definition 4.21 ([?, Chap. 5.6]) *A propagator $P : L^1(\mu) \to L^1(\mu)$ is called* ***asymptotically stable*** *if*

$$\lim_{n\to\infty} \|P^n v - \mathbf{1_X}\|_1 \quad = \quad 0 \tag{50}$$

for every density $v \in L^1(\mu)$.

Define the **limit propagator** $P_\infty : L^1(\mu) \to L^1(\mu)$ by

$$P_\infty v(y) \equiv \int_{\mathbf{X}} v(x)\mu(\mathrm{d}x) \tag{51}$$

for arbitrary $v \in L^1(\mu)$, which corresponds to the projection onto the eigenspace spanned by $\mathbf{1_X}$. In terms of P_∞ we can restate (**??**) in the equivalent form: $\lim_{n\to\infty} \|P^n v - P_\infty v\|_1 = 0$ for $v \in L^1(\mu)$. Finally, we get [**?**]

Corollary 4.22 *Let $P : L^1(\mu) \to L^1(\mu)$ denote a uniformly constrictive propagator. Then P is asymptotically stable if and only if P is ergodic and aperiodic. In either case,*

$$\|P^n - P_\infty\|_1 \leq Mq^n \qquad n \in \mathbf{Z}_+$$

for some constants $q < 1$ and $M < \infty$.

An analogous result to Cor. **??** for the backward transfer operator is well established in the theory of Markov chains. It is related to a property of the stochastic transition function called uniform ergodicity [**?**]. To state it, we introduce the **total variation norm** on measures:

$$\|\nu\|_{\mathrm{TV}} = \sup_{|u|\leq 1} \int_{\mathbf{X}} u(y)\nu(\mathrm{d}y).$$

Definition 4.23 *A stochastic transition function p is said to be **μ-a.e. uniformly ergodic** if*

$$\|p^n(x,\cdot) - \mu\|_{\mathrm{TV}} \leq Mq^n \qquad n \in \mathbf{Z}_+ \tag{52}$$

*for μ–a.e. $x \in \mathbf{X}$ and some constants $q < 1$ and $M < \infty$. If (**??**) holds for every $x \in \mathbf{X}$ then p is called uniformly ergodic.*

In terms of the backward transfer operator and its **limit backward transfer operator** $T_\infty : L^\infty(\mu) \to L^\infty(\mu)$ defined by

$$T_\infty u(x) \equiv \int_{\mathbf{X}} u(y)\mu(\mathrm{d}y),$$

we can restate (**??**) in the equivalent form $\lim_{n\to\infty} \|T^n - T_\infty\|_\infty = 0$. Exploiting the duality $P_\infty^* = T_\infty$, we can relate asymptotically stable propagators and μ-a.e. uniformly ergodic stochastic transition functions as follows:

Theorem 4.24 *Let $P : L^1(\mu) \to L^1(\mu)$ denote some propagator. Then P is uniformly constrictive and asymptotically stable if and only if its corresponding stochastic transition function p is μ-a.e. uniformly ergodic.*

Proof: The result follows from the fact that μ-a.e. uniform ergodicity is equivalent to $\lim_{n\to\infty} \|T^n - T_\infty\|_\infty = 0$, which due to duality is equivalent to uniform constrictiveness and asymptotic stability due to Cor. **??**. $\square$

As a result, we can reformulate the two conditions (C1) and (C2) imposed on the propagator P_τ regarded to act on $L^1(\mu)$ in the equivalent form:

(C1) The propagator P_τ is uniformly constrictive. Equivalently, the stochastic transition function $p(x, A) = p(\tau, x, A)$ fulfills the μ–a.e. Doeblin–condition.

(C2) Condition (C1) holds and P_τ is asymptotically stable.

Moreover, the propagator P_τ satisfies conditions (C1) and (C2) if the stochastic transition function is μ–a.e. uniformly ergodic. Since the reformulated conditions are stated in the language of Markov operators and Markov processes, we can exploit the rich literature on these topics (see [**?**, **?**] and cited reference therein) to verify the conditions (C1) and (C2) for different model systems in Section **??**.

4.4 Reversibility and Properties in $L^2(\mu)$

The basic idea in analyzing reversible propagators on $L^2(\mu)$ will be to follow along the lines of the $L^1(\mu)$ approach. In doing so, compact operators will replace the role previously played by weakly compact operators. Both cases are special situations of a much more general Δ–calculus introduced by Schechter [**?**] in 1972. His aim was to study strictly singular operators[9], which play an important role as admissible perturbations of Fredholm operators[10] [**?**, **?**]. These, moreover, are closely related to essential spectra and in particular to the essential spectral radius [**?**]. Schechter introduced his quantity for an arbitrary bounded linear operator on some Banach space. For the $L^1(\mu)$ case, Weis proved in [**?**] the very useful identity $\Delta(P) = \limsup_{\mu(A)\to\infty} \|\mathbf{1}_A \circ P\|_1$, which played the key role for the subsequent analysis in Section 4.2. As we will see, this characterization of Δ does unfortunately not carry over to $L^2(\mu)$ in general—but it remains true for integral operators [**?**].

Before we start studying propagators on $L^2(\mu)$, we want to recall that due to Hölder's inequality we have $\|v\|_1 \leq \|v\|_2$ for every $v \in L^2(\mu)$. Hence,

[9] A closed bounded linear operator $P : E \to E$ on some Banach space E is called *strictly singular*, if it does not possess a bounded inverses on any infinite dimensional subspace M of E [**?**]. Equivalently, the existence of some constant $\gamma > 0$ such that $\|Px\| \geq \gamma\|x\|$ for every $x \in M \subset E$ implies that M is finite dimensional [**?**, Chapter 4.5].

[10] For a definition see footnote on page 31.

any convergence rate obtained in $L^2(\mu)$ will imply the same rate in the $L^1(\mu)$ norm, when restricted to square integrable functions, i.e., whenever $\|P^n v - P_\infty v\|_2 \leq M q^n$ holds, then also

$$\|P^n v - P_\infty v\|_1 \quad \leq \quad M q^n$$

for every $v \in L^2(\mu)$. This way we obtain probabilistic interpretations of results established in $L^2(\mu)$.

Theorem 4.25 ([?]) *Let $P : L^2(\mu) \to L^2(\mu)$ denote a bounded linear operator. Define the semi–norm $\Delta(P)$ according to*

$$\Delta(P) \quad = \quad \min\left\{\|P - S\|_2 : S \text{ is compact}\right\}.$$

Then the essential spectral radius of P is characterized by

$$r_{\mathrm{ess}}(P) \quad = \quad \lim_{n\to\infty} \Delta(P^n)^{1/n}. \tag{53}$$

In particular, $r_{\mathrm{ess}}(P) \leq \Delta(P)$. If additionally P is positive[11] *and self–adjoint, then $r_{\mathrm{ess}}(P) = \Delta(P)$.*

Note that Corollary 4.3 applies to our situation, hence $r_{\mathrm{ess}}(P) < 1$ if and only if P is quasi–compact. This was the path followed in [**?**] by Schütte to prove that the essential spectral radius is less than 1. Our aim in the following is to relate the property of quasi–compactness and hence $r_{\mathrm{ess}}(P) < 1$ to properties of the stochastic transition function and the corresponding Markov process. We start by giving a characterization of compact operators comparable to Theorem 4.7. To do so, we have to introduce the notion of compactness in measure.

Definition 4.26 ([?, Chapter 1.3.3]) *Let $S : L^2(\mu) \to L^2(\mu)$ denote a bounded linear operator. Then S is called **compact in measure** if it maps weakly convergent sequences to sequences converging in measure. More precisely, if $\{f_n\}_{n\in\mathbf{Z}_+} \subset L^2(\mu)$ is weakly convergent, then for every $\epsilon > 0$ there is $n_0 \in \mathbf{Z}_+$ such that $\mu(\{|Sf_n - Sf_m| \geq \epsilon\}) < \epsilon$ for every $n, m > n_0$.*

An important class of operators being compact in measure are positive integral operators [**?**], and hence all propagators corresponding to absolutely continuous transition functions. We are now able to give a characterization of compact operators in terms of the probability measure μ.

[11] Here, positivity is understood in the Markov operator sense: $Pv \geq 0$ if $v \geq 0$ as stated on page 10. This is different from positivity of self–adjoint operators on a Hilbert space: $\langle v, P_t v\rangle_\mu \geq 0$ for every v.

Lemma 4.27 ([?, Thm. 3.1]) *Let $S : L^2(\mu) \to L^2(\mu)$ denote a bounded linear operator. Then S is compact, if and only if it is compact in measure and satisfies*

$$\limsup_{\mu(A)\to 0} \|\mathbf{1}_A \circ P\|_2 \quad = \quad 0, \tag{54}$$

where the limit is understood to be taken over all sequences of subsets whose μ–measure converges to zero and $\mathbf{1}_A$ is interpreted as a multiplication operator: $(\mathbf{1}_A v)(x) = \mathbf{1}_A(x)v(x)$.

Weis proved that for an arbitrary *integral* operators, the expression of the left hand side of (**??**) is identical to the Δ semi–norm and therefore allows to bound the essential spectral radius.

Theorem 4.28 ([?]) *Let $P : L^2(\mu) \to L^2(\mu)$ denote a bounded linear* integral *operator. Then*

$$\Delta(P) \quad = \quad \limsup_{\mu(A)\to 0} \|\mathbf{1}_A \circ P\|_2. \tag{55}$$

In particular,

$$\limsup_{\mu(A)\to 0} \|\mathbf{1}_A \circ P\|_2 \quad = \quad 0,$$

if and only if P is compact.

As in the $L^1(\mu)$ case, we now want to link the results concerning the Δ semi–norm to properties of the stochastic transition function, in terms of which the propagator is defined. The next lemma is comparable to Cor. 4.8.

Lemma 4.29 *Consider the* reversible *propagator $S : L^2(\mu) \to L^2(\mu)$ defined by*

$$Sv(y) \quad = \quad \int_{\mathbf{X}} v(x)p(x,y)\mu(\mathrm{d}x) \tag{56}$$

associated with some absolutely continuous stochastic transition function $p(x, \mathrm{d}y) = p(x, y)\mu(\mathrm{d}y)$, and assume that p is jointly measurable in x and y. Then, S is compact, if the stochastic transition function satisfies the ***Kontorovic condition****:*

> *there exist $1 \le r, s \le \infty$ with $1/r + 1/s = 1$ such that $\|p(x, \cdot)\|_s \in L^r(\mu)$ as a function of x, i.e.,*
>
> $$\int_{\mathbf{X}} \int_{\mathbf{X}} p(x,y)^s \mu(\mathrm{d}y)^{r/s} \mu(\mathrm{d}x) \quad < \quad \infty. \tag{57}$$

In addition, S is compact if the stochastic transition function satisfies the condition $p(\cdot,\cdot) \in L^r(\mu \times \mu)$ for some $2 \leq r \leq \infty$.

Proof: The first statement is due to Theorem 7.2 of Krasnoseslkii et al. [**?**, Chapter 2], where we have to choose $\tau = (1/2 - 1/r)/(1/s - 1/r)$ for $r \neq s$ and $\tau = 1/2$ for $r = s$. The second statement is a consequence of the first and Hölder's inequality, since $L^r(\mu \times \mu) \subset L^2(\mu \times \mu)$ for $2 \leq r \leq \infty$. □

For $s = r = 2$ the Kontorovic condition is equivalent to the statement that S is a Hilbert–Schmidt operator, which is known to be compact [**?**].

Due to investigations initiated by Roberts and Rosenthal [**?**], quasi–compactness of P is related to certain stability properties of Markov processes. To state them, let $M : \mathbf{X} \to \mathbf{R}_+$ denote an integrable function, i.e., $M \in L^1(\mu)$ and define the induced M–**norm** on measures by

$$\|\nu\|_M \quad = \quad \sup_{|v|\leq M} \left| \int_{\mathbf{X}} v(x)\nu(\mathrm{d}x) \right|,$$

where $|v| \leq M$ is understood to hold pointwise for every $x \in \mathbf{X}$. For the special case $M \equiv 1$, the M–norm coincides with the total variation norm.

Definition 4.30 *Let p denote some stochastic transition function. Then*

*(i) p is called μ-**a.e. geometrically ergodic** if*

$$\|p^n(x,\cdot) - \mu\|_{\mathrm{TV}} \quad \leq \quad M(x)q^n; \qquad n \in \mathbf{Z}_+ \tag{58}$$

for μ–a.e. $x \in \mathbf{X}$, some constant $q < 1$, and some function $M : \mathbf{X} \to \mathbf{R}$ satisfying $M < \infty$ pointwise.

*If inequality (**??**) holds for every $x \in \mathbf{X}$ and some function $M \in L^1(\mu)$, then p is called geometrically ergodic.*

*(ii) p is called V–**uniformly ergodic**[12] if*

$$\|p^n(x,\cdot) - \mu\|_M \quad \leq \quad CM(x)q^n; \qquad n \in \mathbf{Z}_+$$

for every $x \in \mathbf{X}$, constants $q < 1$ and $C \leq \infty$, and some function $M \in L^1(\mu)$ satisfying $1 \leq M$ pointwise.

The relation between the stability properties defined above is as follows: By definition, V–uniform ergodicity implies geometric ergodicity, which in turn implies μ-a.e. geometric ergodicity. On the other hand, for irreducible and aperiodic stochastic transition functions μ-a.e. geometric ergodicity implies V–uniform ergodicity according to [**?**, Prop. 2.1]. We now get the following important result:

[12] The notion V–uniform ergodicity is due to the fact that the function M involved in its definition is usually called V. However, in this thesis we already used V to denote the potential energy function.

Theorem 4.31 *Let $P : L^2(\mu) \to L^2(\mu)$ denote a reversible propagator. Then P satisfies conditions (C1) and (C2) in $L^2(\mu)$, if and only if its stochastic transition function is μ–irreducible and μ-a.e. geometrically ergodic. The latter two conditions on the stochastic transition function p are particularly satisfied, if p is geometrically or V–uniformly ergodic.*

Proof: If P satisfies the two conditions (C1) and (C2), then p is μ-a.e. geometrically ergodic due to Theorem 1 of [**?**]. On the other hand if p is reversible, μ–irreducible and μ-a.e. geometrically ergodic, then P satisfies the conditions (C1) and (C2) as an immediate result of Theorem 2 of [**?**] and Theorem 2.1 of [**?**]. The second statement follows directly from the remark preceding the theorem. □

The assumption of μ–irreducibility of the stochastic transition function in Theorem **??** seems to be artificial. One would rather expect μ–a.e. irreducibility, which furthermore would be a consequence of μ-a.e. geometric ergodicity. Hence, we expect Theorem **??** to hold without the assumption of μ–irreducibility. For reversible propagators we finally get the following relation between the conditions (C1) and (C2) in $L^1(\mu)$ and those in $L^2(\mu)$:

Theorem 4.32 *Let $P : L^1(\mu) \to L^1(\mu)$ denote some propagator satisfying conditions (C1) and (C2) in $L^1(\mu)$. If P is reversible and its stochastic transition function is μ–irreducible then $P : L^2(\mu) \subset L^1(\mu) \to L^2(\mu)$ also satisfies the conditions (C1) and (C2) in $L^2(\mu)$.*

Proof: In $L^1(\mu)$ the conditions (C1) and (C2) are equivalent to μ–a.e. uniform ergodicity of the associated Markov process (see Theorem **??**). Since μ–a.e. uniform ergodicity implies μ–a.e. geometric ergodicity, P satisfies (C1) and (C2) in $L^2(\mu)$ due to Theorem **??**. □

We finally obtain the useful

Corollary 4.33 *If $P : L^r(\mu) \to L^r(\mu)$ with $r = 1, 2$ is reversible and its stochastic transition function p is uniformly ergodic, then P satisfies the conditions (C1) and (C2) both in $L^1(\mu)$ and $L^2(\mu)$.*

As a result of this section, we can state the conditions (C1) and (C2) in a more probabilistic language. Particularly, Theorem **??** will be very useful when verifying conditions (C1) and (C2) for new model systems.