



Konrad-Zuse-Zentrum
für Informationstechnik Berlin

ZIB

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ADAM NIELSEN AND KONSTANTIN FACKELDEY AND
MARCUS WEBER

On a Generalized Transfer Operator

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

On a Generalized Transfer Operator

Adam Nielsen*and Konstantin Fackeldey†and Marcus Weber‡,

November 26, 2013

Abstract We introduce a generalized operator for arbitrary stochastic processes by using a pre-kernel, which is a generalization of the Markov kernel. For deterministic processes, such an operator is already known as the Frobenius-Perron operator, which is defined for a large class of measures. For Markov processes, there exists transfer operators being only well defined for stationary measures in L^2 . Our novel generalized transfer operator is well defined for arbitrary stochastic processes, in particular also for deterministic ones. We can show that this operator is acting on L^1 . For stationary measures, this operator is also an endomorphism of L^2 and, therefore, allows for a mathematical analysis in Hilbert spaces.

*nielsen@zib.de
†fackeldey@zib.de
‡weber@zib.de

1 Introduction

In the last decades, operators which propagate probability densities of a dynamical system have been extensively studied [6, 13, 9]. In particular the transfer operator for deterministic systems, also called the Perron Frobenius operator, has been analyzed [11, 8, 7, 9]. For an exhaustive overview of transfer operators and stochastic aspects we refer to [1]. Inspired by the work of Dellnitz and Junge in [10] where the Perron-Frobenius operator for small random perturbations was analyzed with the aim to gain a numerical approximation of complicated dynamical behavior, Schütte considered a transfer operator for applications in the simulation of bio molecules [16]. More precisely, he introduced a transfer operator in a weighted L^2 space which projects the dynamics of the phase space onto the conformational state space. This operator can be interpreted as a momentum weighted Perron-Frobenius operator on the state space. A Galerkin discretization of this projection combined with the weighting allows for describing the dynamics of a bio molecule in terms of transition probabilities between subsets of the state space. Since then, much progress has been done. In the dissertation of Huisinga [14], the spectrum of a general class of Markov transfer operators was analyzed, and in the dissertation of Sarich [15], the projection of Markov transfer operators onto committor functions was investigated. In the habilitation thesis of Weber [17], a class of transfer operators that derives from averaging a Markov kernel was analyzed.

In this article, we construct a generalized transfer operator that propagates probability densities of a dynamical system. For this, we need some preparations. Firstly, we have to define in which sense we can get hold of a dynamical system. Secondly, we need to define in which space this transfer operator should act, and if it is well defined. Finally, we need to show that it has the property of propagating the according probability densities.

Operator	Process	Kernel	Ch.-Kolmog.
Frobenius [1]	deterministic	Dirac delta	in a discrete sense
Markov [5, 14]	stochastic	Markov	yes
Schütte [16]	special projection of a deterministic system	pre-kernel	no
Dellnitz [10]	deterministic system with stochastic perturbation	pre-kernel	no
Weber [17]	special projection of a Markov process	pre-kernel	not in general

Table 1: Table of different operators mentioned in literature that can be seen as a special case of the general operator defined in this article. The bold printed entries represent the special qualifications of the given operator compared to the general case.

1.1 The essence of a dynamical system

In the following, let (E, Σ_E, μ) define a σ -finite measure space on some set E , i.e. there is a sequence $(A_i)_{i \in \mathbb{N}}$, $A_i \in \Sigma_E$, satisfying

$$E = \bigcup_{i \in \mathbb{N}} A_i \text{ and } \mu(A_i) < \infty \text{ for all } i \in \mathbb{N},$$

and let us denote the corresponding Lebesgue space¹ by $L^1(E, \Sigma_E, \mu)$. Here, we define the Lebesgue spaces for $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and $1 \leq p < \infty$ by

$$L^p(E, \Sigma_E, \mu) := \{f: E \rightarrow \mathbb{K} \mid f \text{ is measurable, } \|f\|_p < \infty\}/N,$$

$$\|f\|_p := \int_E |f(x)|^p \mu(dx),$$

$$L^\infty(E, \Sigma_E, \mu) := \{f: E \rightarrow \mathbb{K} \mid f \text{ is measurable, } \|f\|_\infty < \infty\}/N,$$

$$\|f\|_\infty := \inf_{A \in \Sigma_E, \mu(A)=0} \sup_{x \in E \setminus A} |f(x)|,$$

where N is given by

$$N = \{f: E \rightarrow \mathbb{K} \mid \exists A \in \Sigma_E \text{ with } \mu(A) = 0 \text{ and } f|_{\Omega \setminus A} = 0\}.$$

In the following - if possible - we neglect E and Σ_E and write only $L^p(\mu)$.

Deterministic System

A deterministic system is given by a map $S: E \rightarrow E$. The system moves any point $x \in E$ to $S(x)$. The operator that propagates the corresponding probability densities is known as the Frobenius-Perron Operator \mathcal{T} [1]. Let now S be a μ -non singular map, i.e. for all $A \in \Sigma_E$ with $\mu(A) = 0$ it follows $\mu(S^{-1}(A)) = 0$ where $S^{-1}(A) := \{x \in E \mid S(x) \in A\}$, then the Frobenius-Perron Operator is acting on functions $f \in L^1(E, \Sigma_E, \mu)$, and is uniquely given by

$$\int_A \mathcal{T}f(x) \mu(dx) = \int_{S^{-1}(A)} f(x) \mu(dx) \quad (1)$$

for all $A \in \Sigma_E$. To see this, consider some random variable X on E which gives rise to a probability density f on E through

$$\mathbb{P}(X \in A) = \int_A f(x) \mu(dx),$$

for any $A \in \Sigma_E$. We want \mathcal{T} to propagate the probability density according to S , i.e.

$$\mathbb{P}(S(X) \in A) = \int_A \mathcal{T}f(x) \mu(dx),$$

for any $A \in \Sigma_E$, but we also have that

$$\mathbb{P}(S(X) \in A) = \mathbb{P}(X \in S^{-1}(A)) = \int_{S^{-1}(A)} f(x) \mu(dx)$$

which gives rise to (1). The Frobenius-Perron operator \mathcal{T} by (1) is well-defined, continuous and unique [1]. Later on, we show that \mathcal{T} derives as a special case from our here introduced transfer operator.

¹See section 2.2.

Markov chain

Consider a time-homogeneous Markov chain $(X_k)_{k \in \mathbb{N}}$ on a finite set $E = \{1, \dots, n\}$, i.e.

$$\mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k, \dots, X_0 = i_0) = \mathbb{P}(X_{k+1} = i_{k+1} \mid X_k = i_k)$$

for all $i_l \in E$, $l = 1, \dots, k$, $k \in \mathbb{N}$. For a given transition matrix P , i.e.

$$P_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i), \quad i, j \in \{1, \dots, n\}$$

we can say that the according Markov chain $(X_k)_{k \in \mathbb{N}}$ with starting density $f \in \mathbb{R}^n$, i.e. $f(i) = \mathbb{P}(X_0 = i)$ for $i = 1, \dots, n$, evolve as

$$\mathbb{P}(X_k = i) = (f^T P^k)(i).$$

Therefore, the propagation of probability densities of a Markov chain is determined by the transition matrix P and a start density f .

Markov process

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathcal{A} \subset \Sigma$ a σ -algebra. A function $\phi: \Omega \rightarrow \mathbb{R}$ is called conditional expectation of a \mathbb{P} -integrable function $Y: \Omega \rightarrow \mathbb{R}$ under condition \mathcal{A} , if

- (i) ϕ is \mathcal{A} -measurable,
- (ii) and for all $A \in \mathcal{A}$ it holds

$$\int_A Y(x) \mathbb{P}(dx) = \int_A \phi(x) \mathbb{P}(dx).$$

One can show that ϕ is almost surely unique, and we denote it in the following by $\mathbb{E}[Y \mid \mathcal{A}]$. For a stochastic process $(X_t)_{t \geq 0}$, we denote with

$$\mathcal{F}_t = \sigma(X_s^{-1} \mid s \leq t)$$

the standard filtration. A stochastic process $(X_t)_{t \geq 0}$, $X_t: \Omega \rightarrow \mathbb{R}^n$ is called a Markov process if we have for any Borel-set $A \in \mathcal{B}(\mathbb{R}^n)$:

$$\mathbb{E}[\mathbf{1}_{X_s \in A} \mid \mathcal{F}_t] = \mathbb{E}[\mathbf{1}_{X_s \in A} \mid \sigma(X_t)]$$

for all $s \leq t$. For the context of a Markov process on a infinite state space E , the transition matrix as given by a Markov chain becomes replaced by a transition function $p: [0, \infty) \times E \times \Sigma_E \rightarrow \mathbb{R}$, where for some $x \in E$, $A \in \Sigma_E$ the term $p(t, x, A)$ can be seen as the probability to move from state x to some state of A after time $t \geq 0$. A transition function is called Markov kernel if it meets:

1. $A \mapsto p(t, x, A)$ is a probability measure on Σ_E for all $t \geq 0$, $x \in E$,
2. $x \mapsto p(t, x, A)$ is measurable on E for all $A \in \Sigma_E$ and $t \geq 0$,
3. $p(0, x, A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{else,} \end{cases}$

4. $p(t+s, x, A) = \int_E p(s, y, A) p(t, x, dy)$ for all $x \in E$, $A \in \Sigma_E$ and $t, s \geq 0$.

It can be shown that a Markov process X_t in E of p exists, i.e.

$$\mathbb{E}[\mathbb{1}_{X_{t+s} \in A} \mid \mathcal{F}] = p(s, X_t, A), \quad (2)$$

for all $s, t \geq 0$, $A \in \Sigma_E$. In particular, all finite dimensional densities are uniquely determined by p (see [12] chapter 4.1 Markov processes and transition functions, Theorem 1.1). It can also be shown that for any Markov process X_t , there exists a unique Markov kernel that meets (2). Therefore, the Markov kernel can be seen as a characterization of Markov processes. It has been suggested that the transfer operator for Markov processes can be introduced as follows [5, 14]. First, define the operator

$$\begin{aligned} \mathcal{U}_t: X &\rightarrow X \\ f &\mapsto \left(x \rightarrow \int_E f(y) p(t, x, dy) \right) \end{aligned}$$

on some space $X = L^r(E, \Sigma_E, \mu)$ for $1 \leq r \leq \infty$. In [4], it has been shown that for the case of a stationary density μ of X_t and the corresponding Markov kernel p , i.e.

$$\int_A \mu(dx) = \int_E p(t, x, A) \mu(dx),$$

for all $A \in \Sigma_E$, the operator \mathcal{U}_t is well-defined for any $1 \leq r \leq \infty$. If we define the μ -weighted L^2 scalar product:

$$\langle f, g \rangle_\mu := \int_E f(x) \overline{g(x)} \mu(dx),$$

for $f \in L^q(E, \Sigma_E, \mu)$ and $g \in L^p(E, \Sigma_E, \mu)$ with $\frac{1}{p} + \frac{1}{q} = 1$, then it is known that for each $1 \leq r < \infty$ there exists an adjoint operator \mathcal{T}_t with

$$\langle \mathcal{U}_t f, g \rangle_\mu = \langle f, \mathcal{T}_t g \rangle_\mu,$$

for all $f \in L^q(E, \Sigma_E, \mu)$, $g \in L^p(E, \Sigma_E, \mu)$. Unfortunately, there are two strong restrictions:

1. We only know that \mathcal{U}_t is well-defined if μ is a stationary measure.
2. For $L^\infty(\mu)$ we do not know if the operator $\mathcal{T}_t: L^1(\mu) \rightarrow L^1(\mu)$ exists. In general, for an arbitrary operator $S: L^\infty \rightarrow L^\infty$ there does not exist an adjoint operator. This is a very strong restriction, because L^1 contains all possible probability densities, and it seems natural to define the transfer operator on such a space.

This shows already that even in the case of Markov processes we have some difficulties to define the transfer operator. In addition, a formal proof that \mathcal{T} does indeed propagate probability densities of the corresponding dynamics is still missing. We now show that we can abandon these restrictions, and that there is no need to confine to Markov processes, in particular with regard to item 4.

1.2 Structure of this article

We present a transfer operator for a general dynamical system in the following stages:

1. In Section 2.1, we introduce a pre-kernel which is a generalization of a Markov kernel, and show that for each stochastic process there exists a pre-kernel p , such that the density of the process is captured by p . Therefore, it always can be assumed that for a given dynamical system there exists a corresponding pre-kernel.
2. In Section 2.2, we introduce the operator \mathcal{U}_t as mentioned above, but with the difference that we replace the Markov kernel with the here defined pre-kernel without being restricted on stationary measures. After that, we prove that the adjoint \mathcal{T}_t in the case of L^∞ exists.
3. At the end of Section 2.2, we prove that \mathcal{T} propagates probability densities according to the dynamical system.
4. In Section 2.3, we show that when μ is a stationary density we have

$$\mathcal{T}(L^2(\mu)) \subseteq L^2(\mu)$$

which allows us to use the projection machinery of L^2 . The projection theory for Markov processes on committor functions has been recently investigated in [15].

5. In Section 3, we end the article by showing some basic properties of the transfer operator, and we prove that the here defined transfer operator is a generalization of the Frobenius-Perron Operator and the transfer operator considered by Schütte [16].

2 Transfer operator

Let (E, Σ_E, μ) be a σ -finite measuring space where Σ_E is the σ -algebra on some set E , and μ is a σ -finite measure defined on E . For a set $A \in \Sigma_E$ we denote with

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{else,} \end{cases}$$

the characteristic function of A . Furthermore, we define the probability space $(\Omega, \Sigma, \mathbb{P})$ with the σ -algebra Σ on a set Ω and the probability measure \mathbb{P} , i.e. $\mathbb{P}(\Omega) = 1$.

2.1 Pre-kernel

We denote

$$p: [0, \infty) \times E \times \Sigma_E \rightarrow \mathbb{R}$$

as a pre-kernel iff

1. $A \mapsto p(t, x, A)$ is a probability measure on Σ_E for all $t \geq 0$, $x \in E$,
2. $x \mapsto p(t, x, A)$ is measurable on E for all $A \in \Sigma_E$ and $t \geq 0$,

3. and $p(0, x, A) = \mathbb{1}_A(x)$ holds for all $A \in \Sigma_E$ and $x \in E$.

We say that the pre-kernel corresponds to a stochastic process $(X_t)_{t \geq 0}$, $X_t: \Omega \rightarrow E$, iff

$$\mathbb{P}(X_t \in B, X_0 \in A) = \int_A p(t, x, B) \nu(dx) \quad (3)$$

holds for all $A, B \in \Sigma_E$, $t \geq 0$, where ν denotes the probability measure $\nu(A) := \mathbb{P}(X_0 \in A)$ for $A \in \Sigma_E$. The following lemma shows, that this definition of a pre-kernel is reasonable.

Proposition 1. Let us given a stochastic process $(X_t)_{t \geq 0}$, and denote with ν the measure

$$\nu(A) := \mathbb{P}[X_0 \in A] \text{ for all } A \in \Sigma.$$

Then, there exist a ν -almost surely unique corresponding pre-kernel.

Proof. To see uniqueness according to ν , let us assume that we have two pre-kernel p_1 and p_2 corresponding to $(X_t)_{t \geq 0}$. It follows, that

$$\int_A p_1(t, x, B) \nu(dx) = \mathbb{P}(X_t \in B, X_0 \in A) = \int_A p_2(t, x, B) \nu(dx)$$

holds for all $A, B \in \Sigma$ and, therefore, $p_1 = p_2$ ν -almost surely.

To see existence, take $t \geq 0$ and $B \in \Sigma$ arbitrarily. Let us consider the following measure

$$\begin{aligned} p_{B,t}: \Sigma &\rightarrow \mathbb{R} \\ A &\mapsto \mathbb{P}(X_t \in B, X_0 \in A). \end{aligned}$$

Note that for $A \in \Sigma$ with $\nu(A) = 0$ it follows $p_{B,t}(A) = 0$. Therefore, due to the Radon-Nikodym theorem, this measure has a density according to ν that we denote by $p(t, \cdot, B)$. By definition, it corresponds to $(X_t)_{t \geq 0}$. We now show that p is actually a pre-kernel ν -almost surely. Therefore, we have to show that

- (i) $p(t, x, E) = 1$,
- (ii) $p(t, x, B) \geq 0$ for all $B \in \Sigma_E$,
- (iii) $p(t, x, \bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} p(t, x, A_i)$ for each sequence $(A_i)_{i \in \mathbb{N}}$ of disjoint, measurable sets,
- (iv) $p(0, x, A) = \mathbb{1}_A(x)$,

holds ν -almost surely.

We show these items by making use of the fact that for any $f, g \in L^1(\nu)$ we can derive from

$$\int_A f(x) \nu(dx) = \int_A g(x) \nu(dx) \text{ for all } A \in \Sigma_E$$

that $f = g$ holds ν -almost surely. The point (i) and (ii) follow from

$$\int_A p(t, x, E) \nu(dx) = \mathbb{P}(X_t \in E, X_0 \in A) = \mathbb{P}(X_0 \in A) = \int_A 1 \nu(dx) \text{ for all } A \in \Sigma_E$$

and

$$\int_A p(t, x, B) \nu(dx) = \mathbb{P}(X_t \in B, X_0 \in A) \geq 0 \text{ for all } A, B \in \Sigma_E.$$

To meet (iii), let us given some sequence $(A_i)_{i \in \mathbb{N}}$ of disjoint, measurable sets. From

$$\mathbb{P}(X_t \in \bigcup_{i \in \mathbb{N}} A_i, X_0 \in A) = \sum_{i \in \mathbb{N}} \mathbb{P}(X_t \in A_i, X_0 \in A) \text{ for all } A \in \Sigma_E$$

we obtain

$$\int_A p(t, x, \bigcup_{i \in \mathbb{N}} A_i) \nu(dx) = \sum_{i \in \mathbb{N}} \int_A p(t, x, A_i) \nu(dx) \text{ for all } A \in \Sigma_E,$$

and due to the lemma from Beppo Levi we obtain

$$\int_A p(t, x, \bigcup_{i \in \mathbb{N}} A_i) \nu(dx) = \int_A \sum_{i \in \mathbb{N}} p(t, x, A_i) \nu(dx) \text{ for all } A \in \Sigma_E.$$

It remains to show (iv):

$$p(0, x, A) = \mathbb{1}_A(x)$$

ν -almost surely for all $A \in \Sigma_E$ and $x \in E$. This follows from

$$\int_B p(0, x, A) \nu(dx) = \mathbb{P}(X_0 \in B, X_0 \in A) = \int_{A \cap B} \nu(dx) = \int_B \mathbb{1}_A(x) \nu(dx)$$

for all $A, B \in \Sigma_E$ and $x \in E$. \square

The main statement of Theorem 1 is that for each stochastic process we find a corresponding pre-kernel. We show vice versa that for each pre-kernel a stochastic process exists. To do so, some preparations are necessary:

For $n \in \mathbb{N}$ let $I = \{1, \dots, n\}$ be an index set and $\alpha(I)$ the set of all non-empty subsets of I . For $J \in \alpha(I)$ we define for any set A

$$A^J := \{(x_i)_{i \in J} \mid x_i \in A, i \in J\},$$

and denote by \mathcal{B}_J the smallest σ -algebra that contains Σ_E^J . For $J, H \in \alpha(I)$ with $J \subset H$, we denote with p_J^H the projection

$$\begin{aligned} p_J^H &: E^H \rightarrow E^J, \\ (x_i)_{i \in H} &\mapsto (x_i)_{i \in J}. \end{aligned}$$

A family $(\mathbb{P}_J)_{J \in \alpha(I)}$ of probability measures on (E^J, \mathcal{B}_J) is called projective family iff

$$\mathbb{P}_J(A) = \mathbb{P}_H((p_J^H)^{-1}(A)) \tag{4}$$

holds for all $J \subset H \subseteq I$, $A \in \mathcal{B}_J$. One can show that it is sufficient to demand (4) only for all $A \in \Sigma^J$. The following lemma be needed for the next proposition.

Lemma 1. *[[3] 62.4 Corollary] For any projective family $(P_J)_{J \in \alpha(I)}$ on (E^J, \mathcal{B}_J) there exists a stochastic process $(X_i)_{i \in I}$ with state space E such that for $J = \{i_1, \dots, i_l\} \subseteq I$ and $A_1, \dots, A_l \in \Sigma_E$ we have*

$$\mathbb{P}_J(A_1 \times \dots \times A_l) = \mathbb{P}(X_{i_1} \in A_1, \dots, X_{i_l} \in A_l).$$

Now we are able to announce the following proposition.

Proposition 2. For each pre-kernel p and probability density ν there exists a stochastic process $(X_t)_{t \geq 0}$ which meets (3).

Proof. For arbitrary $t \geq 0$, let us consider the measures

$$\mathbb{P}_{\{1\}}(A) = \nu(A), \quad \mathbb{P}_{\{2\}}(B) = \int_E p(t, x, B) \nu(dx)$$

and

$$\mathbb{P}_{\{1,2\}}(A \times B) = \int_A p(t, x, B) \nu(dx)$$

for all $A, B \in \Sigma$. Then the measures $(\mathbb{P}_i)_{i \in \{\{1\}, \{2\}, \{1,2\}\}}$

$$\mathbb{P}_{\{1\}}(A) = \mathbb{P}_{\{1,2\}}(A \times E) \quad \text{and} \quad \mathbb{P}_{\{2\}}(B) = \mathbb{P}_{\{1,2\}}(E \times B)$$

form a projective family, where $A, B \in \Sigma$. Because of Lemma 1, we obtain for each $t \geq 0$ the existence of two stochastic processes $(\tilde{X}_t^1, \tilde{X}_t^2)$ with $\mathbb{P}_{\{1,2\}}(A \times B) = \mathbb{P}(\tilde{X}_t^1 \in A, \tilde{X}_t^2 \in B)$. Define $X_t := \tilde{X}_t^2$. It remains to show that

$$\mathbb{P}(X_0 \in A, X_t \in B) = \int_A p(t, x, B) \nu(dx),$$

which holds iff $\tilde{X}_t^1 = \tilde{X}_0^2$ is true. This follows from

$$\begin{aligned} \mathbb{P}(\tilde{X}_t^1 \in A) &= \int_A \nu(dx) \\ &= \int_E p(0, x, A) \nu(dx) \\ &= \mathbb{P}(\tilde{X}_0^2 \in A). \end{aligned}$$

□

This motivates the following definition. A measure $\mu: \Sigma \rightarrow [0, 1]$ is called a stationary density of p if

$$\int_A \mu(dx) = \int_E p(t, x, A) \mu(dx)$$

holds for all $A \in \Sigma$.

Note, if $(X_t^\mu)_{t \geq 0}$ is the process that evolves according to p with stationary density μ as the starting density, then

$$\mathbb{P}(X_0^\mu \in A) = \int_A \mu(dx) = \int_E p(t, x, A) \mu(dx) = \mathbb{P}(X_t^\mu \in A, X_0^\mu \in E) = \mathbb{P}(X_t^\mu \in A)$$

holds, which legitimizes to name it stationary density.

The following definition is crucial for the introduction of the transfer operator below.

Definition 3 (μ -compatible). A pre-kernel p is called μ -compatible if and only if for each $t \geq 0$ and $A \in \Sigma_E$ with $\mu(A) = 0$ there exists a set $B \in \Sigma_E$ with $\mu(B) = 0$ and

$$p(t, x, A) = 0 \text{ for all } x \in E \setminus B.$$

One may notice that this condition is a weaker demand than absolutely continuity. To see that, consider the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ and a pre-kernel for time $t = 0$

$$p(0, x, A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{else.} \end{cases}$$

Then, for each $x \in E$ the measure $p(0, x, \cdot)$ is a Dirac delta measure, and, therefore, not absolutely continuous. However, p is λ -compatible.

2.2 Connection between operator and pre-kernel

Consider for some $t \geq 0$ the following operator

$$\begin{aligned} \mathcal{U}_t: L^\infty(\mu) &\rightarrow L^\infty(\mu) \\ f &\mapsto \left(x \mapsto \int_E f(y) p(t, x, dy) \right). \end{aligned} \quad (5)$$

Let us show that this operator is not necessarily well-defined. For example, if we fix some $x_0 \in E$, and consider the following pre-kernel

$$p(t, x, A) := \begin{cases} 1 & \text{if } x_0 \in A, \\ 0 & \text{else,} \end{cases}$$

for all $x \in E$, $A \in \Sigma$, $t \geq 0$. Note that p depends only on the set A for $t \geq 0$, and that it is independent of x . Then $(\mathcal{U}_t f)(x) = f(x_0)$ holds, but $f \in L^\infty(\mu)$ is only well defined up to a null set, so $f(x_0)$ could be arbitrary. In fact, in this example \mathcal{U}_t is only well-defined if and only if $\mu(\{x_0\}) > 0$ holds.

Fortunately, we are able to show that the operator is well defined as long as p is μ -compatible. To see that, consider $f, g \in L^\infty(\mu)$ such that

$$f(x) = g(x) \text{ for all } x \in E \setminus A$$

for a null set A . From the μ -compatibility, we obtain a null set B with

$$p(t, x, A) = 0 \text{ for all } x \in E \setminus B.$$

For $x \in E \setminus B$, we obtain

$$\begin{aligned} \int_E g(y) p(t, x, dy) &= \int_{E \setminus A} g(y) p(t, x, dy) \\ &= \int_{E \setminus A} f(y) p(t, x, dy) \\ &= \int_E f(y) p(t, x, dy). \end{aligned} \quad (6)$$

It only remains to show that $\mathcal{U}_t f \in L^\infty(\mu)$. From $f \in L^\infty(\mu)$, we find a constant $M > 0$ and a null-set $A \in \Sigma_E$ such that $|f(x)| \leq M$ for all $x \in E \setminus A$. If p is μ -compatible, we have a null-set B such that $p(t, x, A) = 0$ holds for all $x \in E \setminus B$. Therefore, we have for all $x \in E \setminus B$ that

$$\begin{aligned} \int_E f(y) p(t, x, dy) &= \int_{E \setminus A} f(y) p(t, x, dy) \\ &\leq \int_{E \setminus A} M p(t, x, dy) \\ &= M. \end{aligned}$$

Thus $\mathcal{U}_t f \in L^\infty(\mu)$. Analogous one can see that $\|\mathcal{U}_t f\|_{L^\infty(\mu)} \leq \|f\|_{L^\infty(\mu)}$ holds. We are now in a position to sum up the results to obtain

Proposition 4. The operator \mathcal{U}_t from (5) is well-defined and continuous if p is μ -compatible.

Due to Hölder's inequality, the term

$$\langle f, g \rangle_\mu = \int_\Omega f(x) \overline{g(x)} \mu(dx)$$

is well-defined for $f \in L^p(\mu)$ and $g \in L^q(\mu)$ with $\frac{1}{q} + \frac{1}{p} = 1$, for $1 \leq p, q \leq \infty$, where $\frac{1}{\infty} := 0$. We define the transfer operator $\mathcal{T}_t: L^1(\mu) \rightarrow L^1(\mu)$ as the unique operator that satisfies

$$\langle \mathcal{U}_t f, g \rangle_\mu = \langle f, \mathcal{T}_t g \rangle_\mu, \quad (7)$$

for all $f \in L^\infty$, $g \in L^1(\mu)$. In general, for an arbitrary operator $R: L^\infty(\mu) \rightarrow L^\infty(\mu)$ an adjoint operator in the sense of (7) does not exist. However, we can show that for an operator of the special form as in (5) it does exist. Consider a sequence of measurable functions $(\mathbb{1}_{A_n})_{n \in \mathbb{N}}$ which converges point-wise to some measurable function $\mathbb{1}_A$. Because of $\mathbb{1}_E \in L^1(p)$, it follows from Lebesgue's dominated convergence theorem that $\mathcal{U}_t \mathbb{1}_{A_n}$ converges point wise to $\mathcal{U}_t \mathbb{1}_A$. In the following theorem this is the key to show the existence of the transfer operator.

Theorem 5. Let (E, Σ_E, μ) be a σ -finite measurable space and

$$R: L^\infty(\mu) \rightarrow L^\infty(\mu)$$

a continuous Operator such that for any sequence of measurable functions of the form $(\mathbb{1}_{A_n})_{n \in \mathbb{N}}$ which converges point wise to some measurable function $\mathbb{1}_A$ we have that $R\mathbb{1}_{A_n}$ converges point wise to $R\mathbb{1}_A$ and that $(R\mathbb{1}_{A_n})_{n \in \mathbb{N}}$ is bounded. Furthermore, assume that for $f \geq 0$ we have $Rf \geq 0$, then it exists a unique adjoint operator R^* , i.e.

$$\langle Rf, g \rangle_\mu = \langle f, R^*g \rangle_\mu$$

for all $f \in L^\infty$, $g \in L^1(\mu)$.

Proof. To see uniqueness, assume there exists two adjoint operators R_1^*, R_2^* . Then we have

$$\int_A ((R_1^* - R_2^*)g)(x) \mu(dx) = 0 \text{ for all } A \in \Sigma_E, g \in L^1(\mu).$$

Therefore, $R_1^* = R_2^*$. To see existence, let us first restrict to the case $\mathbb{K} = \mathbb{R}$. For $f \in L^1(\mu)$, consider the functional

$$\begin{aligned} y' : L^\infty(\mu) &\rightarrow \mathbb{R} \\ h &\mapsto \langle Rh, f \rangle_\mu. \end{aligned}$$

We show now the existence of a function $g \in L^1(\mu)$ with the property $y'(h) = \int_\Omega h(x)g(x)\mu(dx)$. If we define the operator S as $Sf := g$, then

$$\langle Rh, f \rangle_\mu = \langle h, Sf \rangle_\mu \quad \forall h \in L^\infty(\mu), f \in L^1(\mu),$$

and it remains to show that S is linear and continuous.

We start with showing the existence of a function g . Define $\nu : \Sigma \rightarrow \mathbb{R}$ by $\nu(A) := y'(\mathbb{1}_A)$ for $A \in \Sigma$. We show now that ν is a signed measure, i.e.

- $\nu(\emptyset) = 0$,
- $\nu(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \nu(A_i)$ for each sequence $(A_i)_{i \in \mathbb{N}}$ of disjoint, measurable sets.

First, we have $\nu(\emptyset) = 0$ since y' is linear. To see the second property, let us given a sequence $(A_i)_{i \in \mathbb{N}}$ of disjoint, measurable sets, and denote with $B := \bigcup_{i \in \mathbb{N}} A_i$ and $B_n := \bigcup_{i=1}^n A_i$ the corresponding union of sets. Since $\mathbb{1}_{B_n}$ converges point wise towards $\mathbb{1}_B$, we obtain that $R\mathbb{1}_{B_n}$ converges point wise towards $R\mathbb{1}_B$. Because of $|R\mathbb{1}_{B_n}(x)f(x)| \leq M|f(x)|$ for some $M \in \mathbb{R}^+$, we obtain from the Lebesgue's dominated convergence theorem that

$$y'(\mathbb{1}_{B_n}) = \langle R\mathbb{1}_{B_n}, f \rangle_\mu \rightarrow \langle R\mathbb{1}_B, f \rangle_\mu = y'(\mathbb{1}_B) \quad \text{for } n \rightarrow \infty$$

holds. Therefore, we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^n \nu(A_i) &= \lim_{n \rightarrow \infty} \sum_{i=1}^n y'(\mathbb{1}_{A_i}) \\ &= \lim_{n \rightarrow \infty} y'(\mathbb{1}_{B_n}) \\ &= y'(\mathbb{1}_B) \\ &= \nu(\bigcup_{i \in \mathbb{N}} A_i), \end{aligned}$$

which shows that ν is indeed a measure. For $A \in \Sigma$ with $\mu(A) = 0$, we have $\mathbb{1}_A = 0$ μ -almost surely and, therefore, $\nu(A) = y'(\mathbb{1}_A) = 0$. Due to the Radon-Nikodym theorem, we obtain a function $g \in L^1(\mu)$ with

$$y'(h) = \int_\Omega h(x)g(x)\mu(dx).$$

If we define $S : L^1(\mu) \rightarrow L^1(\mu)$ as $Sf := g$, it only remains to show that S is linear and continuous. To see linearity, consider some $f_1, f_2 \in L^1(\mu)$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. From

$$\langle Rh, f_i \rangle_\mu = \langle h, Sf_i \rangle_\mu \quad \forall h \in L^\infty(\mu), i = 1, 2,$$

we obtain

$$\langle Rh, \alpha_1 f_1 + \alpha_2 f_2 \rangle_\mu = \langle h, \alpha_1 S f_1 + \alpha_2 S f_2 \rangle_\mu \quad \forall h \in L^\infty(\mu).$$

Because of the uniqueness from the Radon-Nikodym theorem, we have

$$\alpha_1 S f_1 + \alpha_2 S f_2 = S(\alpha_1 f_1 + \alpha_2 f_2).$$

To see continuity, consider for some $f \in L^1(\mu)$ the function \tilde{S}_f given by

$$\tilde{S}_f(x) = \begin{cases} \frac{|(Sf)(x)|}{(Sf)(x)} & \text{if } (Sf)(x) \neq 0, \\ 0 & \text{else.} \end{cases}$$

Because of $|\tilde{S}_f(x)| \leq 1$ for all $x \in E$, we have $\tilde{S}_f \in L^\infty(\mu)$. Therefore, we have

$$\begin{aligned} \|Sf\|_{L^1(\mu)} &= \int_E |(Sf)(x)| \mu(dx) \\ &= \int_E (Sf)(x) \tilde{S}_f(x) \mu(dx) \\ &= \int_E f(x) (R\tilde{S}_f)(x) \mu(dx) \\ &\stackrel{(*)}{\leq} \|f\|_{L^1(\mu)} \|R\|_{L^\infty(\mu)} \|\tilde{S}_f\|_{L^\infty(\mu)} \\ &\leq \|f\| \|R\|, \end{aligned}$$

where we have used the Hölder's inequality in (*).

Let us now consider the case $\mathbb{K} = \mathbb{C}$. Since $Rf \geq 0$ for $f \geq 0$, we know that R maps real-valued functions onto real-valued functions. Therefore, we already have the existence of S acting on all real-valued functions. We continue S on $L^\infty(\mu)$ over the field \mathbb{C} in the following way: For a function $f: \Omega \rightarrow \mathbb{C}$ set

$$Sf := S(\operatorname{Re}(f)) + iS(\operatorname{Im}(f)),$$

then S inherits continuity, linearity, by standard results from analysis. Moreover it is the adjoint of R . \square

It was the essential part of the above proof to find for a special functional $y': L^\infty(\mu) \rightarrow \mathbb{R}$ a function $g \in L^1(\mu)$ with

$$y'(h) = \int_E h(x) g(x) \mu(dx). \quad (8)$$

In general, the existence of a function g as described in (8) is only granted for functionals $y': L^p(\mu) \rightarrow \mathbb{R}$ where $1 \leq p < \infty$ (p.15 Theorem II.2.4). We can now define the transfer operator as follows:

Definition 6. Given a μ -compatible pre-kernel p and the operator \mathcal{U}_t by (5), then we define the (generalized) transfer operator \mathcal{T}_t by

$$\mathcal{T}_t := \mathcal{U}_t^*,$$

where \mathcal{U}_t^* is the adjoint of \mathcal{U}_t .

We show now that the interpretation as explained in the introduction coincides with the transfer operator defined here. To see this, consider an arbitrary starting density f , i.e. $\int_E f(x) \mu(dx) = 1$ and $f \geq 0$ μ -almost surely. From Proposition 2, we know the existence of a process $(X_t^f)_{t \geq 0}$ such that

$$\mathbb{P}(X_0^f \in A) = \int_A f(x) \mu(dx)$$

and

$$\mathbb{P}(X_t^f \in A) = \int_E p(t, x, A) f(x) \mu(dx) \text{ for all } A \in \Sigma, t \geq 0. \quad (9)$$

We can then state the following

Theorem 7. *Let $(X_t^f)_{t \geq 0}$ be given as stated above, then*

$$\mathbb{P}(X_t^f \in A) = \int_A \mathcal{T}_t f(x) \mu(dx) \text{ for all } A \in \Sigma, t \geq 0,$$

where \mathcal{T}_t denotes the corresponding transfer operator given by Definition 6.

Proof.

$$\begin{aligned} \mathbb{P}(X_t^f \in A) &= \int_E f(x) p(t, x, A) \mu(dx) \\ &= \int_E f(x) \left(\int_A p(t, x, dy) \right) \mu(dx) \\ &= \int_E f(x) \left(\int_E \mathbf{1}_A(y) p(t, x, dy) \right) \mu(dx) \\ &= \int_E f(x) (\mathcal{U}_t \mathbf{1}_A)(x) \mu(dx) \\ &= \int_E (\mathcal{T}_t f)(x) \mathbf{1}_A(x) \mu(dx) = \int_A (\mathcal{T}_t f)(x) \mu(dx). \end{aligned}$$

□

2.3 Preparation for projection theory

Since our transfer operator propagates probability densities it seems natural to use $L^1(\mu)$ as the underlying space. However, for a projection onto finite dimensional space for numerical reasons the space $L^2(\mu)$ is favorable. In general an embedding is not possible, since neither $L^2(\mu) \subseteq L^1(\mu)$ nor $L^1(\mu) \subseteq L^2(\mu)$ holds. But if we confine our considerations to the case of a stationary measure μ , we have

$$L^\infty(\mu) \subseteq L^2(\mu) \subseteq L^1(\mu).$$

When now restricting the transfer operator \mathcal{T}_t to $L^2(\mu)$, we have to deal with the question, whether

$$\mathcal{T}_t(L^2(\mu)) \subseteq L^2(\mu) \quad (10)$$

holds, which would be obligatory to make use of the projection machinery in $L^2(\mu)$. For stationary measure μ we now show (10) by extending \mathcal{U}_t onto $L^2(\mu)$.

The corresponding unique adjoint is then equal to the transfer operator on $L^2(\mu)$ and, therefore, we get (10).

Assume that we have a stationary density μ of a given pre-kernel p . Consider then the extended operator

$$\begin{aligned} \tilde{\mathcal{U}}_t: L^2(\mu) &\rightarrow L^2(\mu), \\ f &\mapsto \left(x \mapsto \int_E f(y) p(t, x, dy) \right). \end{aligned} \quad (11)$$

We show:

Lemma 2. *The operator $\tilde{\mathcal{U}}_t$ from (11) is well-defined if μ is a stationary density of p .*

Proof. Let $[f] \in L^2(\mu)$ and consider two representative functions \tilde{f} and \tilde{g} of $[f]$. We prove the following properties:

1. It holds that $\int_E |\tilde{f}(y)| p(t, x, dy) < \infty$ for all $t \geq 0$, $x \in E$ μ -almost surely. Therefore, \tilde{f} is $p(t, x, \cdot)$ -integrable.
2. It holds that $\int_E \tilde{f}(y) p(t, x, dy) = \int_E \tilde{g}(y) p(t, x, dy)$ μ -almost surely.
3. It holds that $\tilde{\mathcal{U}}_t f \in L^2(\mu)$ and $\tilde{\mathcal{U}}_t$ is continuous.

First, one may notice that

$$\begin{aligned} \int_E \mathbb{1}_A(x) \mu(dx) &= \int_E p(t, x, A) \mu(dx) \\ &= \int_E \int_E \mathbb{1}_A(y) p(t, x, dy) \mu(dx). \end{aligned} \quad (12)$$

Proof of 1. One may consider a non-negative monotonic increasing sequence of simple² functions $(f_n)_{n \in \mathbb{N}}$ that converges point-wise towards $|\tilde{f}|$. Due to the lemma of Beppo Levi, we get

$$\lim_{n \rightarrow \infty} \int_E f_n(y) p(t, x, dy) = \int_E |\tilde{f}(y)| p(t, x, dy).$$

Using the lemma of Beppo Levi again yields

$$\lim_{n \rightarrow \infty} \int_E \int_E f_n(y) p(t, x, dy) \mu(dx) = \int_E \int_E |\tilde{f}(y)| p(t, x, dy) \mu(dx).$$

Finally, applying (12) to the simple functions leads us to

$$\int_E |\tilde{f}(x)| \mu(dx) = \int_E \int_E |\tilde{f}(y)| p(t, x, dy) \mu(dx).$$

Since \tilde{f} is μ -integrable, we have $\int_E \int_E |\tilde{f}(y)| p(t, x, dy) \mu(dx) < \infty$ and, therefore,

$$\int_E |\tilde{f}(y)| p(t, x, dy) < \infty$$

μ -almost surely.

²A simple function is a measurable function that can be written as a sum of indicator functions.

Proof of 2. Since μ is a stationary density of p , we have

$$0 = \int_A \mu(dx) = \int_E p(t, x, A) \mu(dx).$$

Since p is by definition non-negative, we obtain a set $B \in \Sigma_E$ with $\mu(B) = 0$ and

$$p(t, x, A) = 0 \text{ for all } t \in I, x \in E \setminus B.$$

This shows that p is μ compatible, and the property follows then from (6).

Proof of 3. Baxter and Rosenthal have shown this property for Markov kernels already in [4]. The proof does not change for the pre-kernel and is given by:

$$\begin{aligned} \|\tilde{\mathcal{U}}_t f\|_{L^2(\mu)}^2 &= \int_E |\tilde{\mathcal{U}}_t f(x)|^2 \mu(dx) \\ &= \int_E \left(\int |f(y)| p(t, x, dy) \right)^2 \mu(dx) \\ &\stackrel{(*)}{\leq} \int_E \int |f(y)|^2 p(t, x, dy) \mu(dx) \\ &\stackrel{(12)}{=} \int_E |f(x)|^2 \mu(dx) \\ &= \|f\|_{L^2(\mu)}^2, \end{aligned}$$

where we have used Jensen's inequality in step (*).

□

Consider now the weighted scalar product

$$\langle f, g \rangle_\mu = \int_E f(x) \overline{g(x)} \mu(dx),$$

then $(L^2(\mu), \langle \cdot, \cdot \rangle_\mu)$ is a Hilbert space, and, therefore, the adjoint $\tilde{\mathcal{U}}_t^*$ of $\tilde{\mathcal{U}}_t$ exists, i.e.

$$\langle \tilde{\mathcal{U}}_t f, g \rangle_\mu = \langle f, \tilde{\mathcal{U}}_t^* g \rangle_\mu$$

for all $f, g \in L^2(\mu)$, and is well-defined.

We denote the adjoint of $\tilde{\mathcal{U}}_t$ by $\tilde{\mathcal{T}}_t = \tilde{\mathcal{U}}_t^*$. It follows for all $f \in L^\infty(\mu)$ and $g \in L^2(\mu)$ that

$$\langle \tilde{\mathcal{U}}_t f, g \rangle_\mu = \langle f, \tilde{\mathcal{T}}_t g \rangle_\mu = \langle f, \mathcal{T}_t g \rangle_\mu.$$

In particular, we have

$$\int_A ((\tilde{\mathcal{T}}_t - \mathcal{T}_t)g)(x) \mu(dx) = 0 \text{ for all } A \in \Sigma_E, g \in L^2(\mu).$$

Thus, $\tilde{\mathcal{T}}_t g = \mathcal{T}_t g$ for all $g \in L^2(\mu)$ which implies (10).

3 Basic properties and examples

In this section, we show some basic properties of the transfer operator. First, we show that the transfer operator of a Markov kernel³ has a certain property, which is also known as the Chapman-Kolmogorov property.

Proposition 8. Let $(\mathcal{T}_t)_{t \geq 0}$ be the family of a transfer operator for a given μ -compatible Markov kernel p . Then, it holds

$$\mathcal{T}_s \mathcal{T}_t = \mathcal{T}_{s+t} \quad \text{for all } s, t \geq 0.$$

Proof. Since p is a Markov kernel, we have that for all $A \in \Sigma_E$

$$p(t+s, x, A) = \int_E p(s, y, A) p(t, x, dy) \quad \text{for all } s, t \geq 0.$$

Therefore, it follows that for $f \in L^\infty(\mu)$

$$\int_A f(y) p(t+s, x, dy) = \int_E \int_A f(y') p(s, y, dy') p(t, x, dy) \quad \text{for } A \in \Sigma_E. \quad (13)$$

To prove the claim, it is only necessary to show $\langle \mathcal{T}_s \mathcal{T}_t g, f \rangle_\mu = \langle g, \mathcal{U}_{s+t} f \rangle_\mu$ for all $g \in L^1(\mu)$, $f \in L^\infty(\mu)$ because the adjoint is unique.

$$\begin{aligned} \langle \mathcal{T}_s \mathcal{T}_t g, f \rangle_\mu &= \langle \mathcal{T}_t g, \mathcal{U}_s f \rangle_\mu \\ &= \langle g, \mathcal{U}_t (\mathcal{U}_s f) \rangle_\mu \\ &= \int_E g(x) \overline{(\mathcal{U}_t (\mathcal{U}_s f))(x)} \mu(dx) \\ &= \int_E g(x) \overline{\left(\int_E (\mathcal{U}_s f)(y) p(t, x, dy) \right)} \mu(dx) \\ &= \int_E g(x) \overline{\left(\int_E \int_E f(y') p(s, y, dy') p(t, x, dy) \right)} \mu(dx) \\ &\stackrel{(13)}{=} \int_E g(x) \overline{\left(\int_E f(y) p(t+s, x, dy) \right)} \mu(dx) \\ &= \langle g, \mathcal{U}_{s+t} f \rangle_\mu. \end{aligned}$$

□

The next statement shows how to extract essential probability information from the transfer operator.

Proposition 9. Let μ be a probability density. Let us given a μ -compatible pre-kernel p and the according transfer operator \mathcal{T}_t . Let denote by $(X_t)_{t \geq 0}$ the according process from Proposition 2 with initial probability density μ . For any sets $A, B \in \Sigma_E$ with $\mu(A) \neq 0$, we have, then, that

$$\frac{\langle \mathcal{T}_t \mathbb{1}_A, \mathbb{1}_B \rangle_\mu}{\langle \mathbb{1}_A, \mathbb{1}_A \rangle_\mu} = \mathbb{P}(X_t \in B \mid X_0 \in A).$$

³See Section 1.1

Proof. We have

$$\begin{aligned}
\langle \mathcal{T}_t \mathbb{1}_A, \mathbb{1}_B \rangle_\mu &= \int_E \mathcal{T}_t \mathbb{1}_A(x) \mathbb{1}_B(x) \mu(dx) \\
&= \int_E \mathbb{1}_A(x) \mathcal{U}_t \mathbb{1}_B(x) \mu(dx) \\
&= \int_E \left(\mathbb{1}_A(x) \int_E \mathbb{1}_B(y) p(t, x, dy) \right) \mu(dx) \\
&= \int_A p(t, x, B) \mu(dx) \\
&= \mathbb{P}(X_t \in B, X_0 \in A),
\end{aligned}$$

and

$$\begin{aligned}
\langle \mathbb{1}_A, \mathbb{1}_A \rangle_\mu &= \int_E \mathbb{1}_A(x) \mathbb{1}_A(x) \mu(dx) \\
&= \int_A \mu(dx) \\
&= \mathbb{P}(X_0 \in A).
\end{aligned}$$

□

We now conclude this paper with some examples for application.

3.1 Deterministic process

Let $S : E \rightarrow E$ measurable transformation on the phase space E which is non-singular according to μ . Define then

$$p(t, x, A) = \begin{cases} 1 & S(x) \in A, \\ 0 & S(x) \notin A, \end{cases} \quad \text{for } t > 0, x \in E, \text{ and } A \in \Sigma_E.$$

Note that p is μ -compatible if and only if S is non-singular according to μ . For a given initial density f according to a random variable X , we get

$$\begin{aligned}
\mathbb{P}(S(X) \in A) &= \mathbb{P}(X \in S^{-1}(A)) \\
&= \int_{S^{-1}(A)} f(x) \mu(dx) \\
&= \int_E p(t, x, A) f(x) \mu(dx).
\end{aligned}$$

Therefore, the constructed pre-kernel meets (9). This implies that the transfer operator \mathcal{T}_t defined by μ and p propagates the densities according to the transformation S .

To see why the transfer operator coincides with the Frobenius-Perron Oper-

ator, note that for $A \in \Sigma_E$ and $f \in L^1(\mu)$ we have:

$$\begin{aligned}
\int_A (\mathcal{T}_t f)(x) \mu(dx) &= \int_E (\mathcal{T}_t f)(x) \mathbb{1}_A(x) \mu(dx) \\
&= \int_E f(x) (\mathcal{U}_t \mathbb{1}_A)(x) \mu(dx) \\
&= \int_E f(x) \int_E \mathbb{1}_A(y) p(t, x, dy) \mu(dx) \\
&= \int_E f(x) p(t, x, A) \mu(dx) \\
&= \int_E f(x) \mathbb{1}_{S^{-1}(A)}(x) \mu(dx) \\
&= \int_{S^{-1}(A)} f(x) \mu(dx).
\end{aligned}$$

which agrees (1). Therefore, the Frobenius-Perron operator is a special case of our transfer operator defined in (6).

3.2 Markov process

Since any Markov process can be described by a Markov kernel⁴, and since each Markov kernel is a pre-kernel, our scheme also covers Markov processes. One is often interested in a stochastic differential equation of the form

$$dX_t = V(X_t, t)dt + \epsilon dW_t, \quad (14)$$

on the interval $I = [t_0, T]$, where ϵ can be interpreted as temperature. Further, assume that a unique stochastic process exists for each $a \in \mathbb{R}^n$, denoted by $(X_t^a)_{t \in I}$, that solves

$$X_t^a = a + \int_s^t f(u, X_u) du + \int_s^t \epsilon dW_u, \quad t_0 \leq s \leq t \leq T.$$

It follows from [2], Theorem 9.2.3, that each solution of (14) is a Markov process, and its Markov kernel, in particular also its pre-kernel, is given by

$$p(t, x, A) := \mathbb{P}(X_t^x \in A).$$

3.3 Classical Mechanics

Let us consider a molecule with N atoms, then in classical molecular dynamics its Hamiltonian is given by

$$H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q),$$

where $q \in \mathbb{R}^{3N}$ denotes the conformational state and $p \in \mathbb{R}^{3N}$ denotes the momentum state, $M \in \mathbb{R}^{3N \times 3N}$ the mass matrix, and V a differentiable potential.

⁴See Section 1.1

If the initial values q_0 and p_0 are known, we can solve the differential equation

$$\begin{aligned} \dot{q}(t) &= \frac{\partial H}{\partial p(t)} = M^{-1}p(t) & \dot{p}(t) &= \frac{\partial H}{\partial(t)} = -\nabla V(q(t)) \\ q(0) &= q_0 & p(0) &= p_0. \end{aligned} \quad (15)$$

We denote the new state (q_t, p_t) by the Hamiltonian flow $\phi^t(q_0, p_0) = (q_t, p_t)$. If $\Omega_c = \mathbb{R}^{3N}$ is the conformational space and $\Omega_m = \mathbb{R}^{3N}$ the momentum space, then the phase space Γ is given as $\Gamma = \Omega_c \times \Omega_m$. Let us denote by $\pi_1: \Gamma \rightarrow \Omega_c$ the projection onto state space $(q, p) \mapsto q$. In molecular dynamics one is often interested on metastable sets according to the state space, i.e. one is interested in the process $\pi_1 \phi^t$. Let $f_0: \Gamma \rightarrow [0, \infty)$ be a probability density on the phase space, which is absolute continuous according to the Lebesgue measure, and consider $F(q) = \int_{\Omega_m} f_0(q, p) dp$, then F is a probability density on the configuration space. In particular, consider the probability space $(\Gamma, \mathcal{B}(\Gamma), f_0)$ and $(\Omega_c, \mathcal{B}(\Omega_c), F)$, where $\mathcal{B}(\Gamma), \mathcal{B}(\Omega_c)$ denotes the corresponding Borel- σ -algebra. If we assume that F is always positive, then

$$\begin{aligned} X_t &: \Gamma \rightarrow \Omega_c, \\ x &\mapsto \pi_1 \phi^t x \end{aligned}$$

is a stochastic process according to the pre-kernel

$$p(t, q, A) := \frac{1}{F(q)} \int_{\Omega_m} p_1(t, (q, p), A) f_0(q, p) dp, \quad (16)$$

with

$$p_1(t, x, A) = \begin{cases} 1 & \text{if } \pi_1 \phi^t x \in A \\ 0 & \text{else.} \end{cases}$$

This is due to the fact, that

$$\begin{aligned} f_0(X_0 \in A) &= \int_{\Gamma} \mathbb{1}_{X_0 \in A}(x) f_0(x) dx \\ &= \int_{\Omega_c} \int_{\Omega_m} \mathbb{1}_A(q) f_0(q, p) dp dq \\ &= \int_A F(q) dq \end{aligned}$$

and

$$\begin{aligned} f_0(X_t \in B, X_0 \in A) &= \int_{\Gamma} \mathbb{1}_{X_0 \in A}(x) \mathbb{1}_{X_t \in B}(x) f_0(x) dx \\ &= \int_A \left(\int_{\Omega_m} \mathbb{1}_B(\pi_1 \phi^t(q, p)) f_0(q, p) dp \right) dq \\ &= \int_A p(t, q, B) F(q) dq. \end{aligned}$$

From Theorem 7 we know that the transfer operator according to the pre-kernel p from (16) propagates probability densities in the conformational space. In

particular, if f_0 is stationary according to ϕ^t , i.e. $\int_A f_0(x) dx = \int_{\phi^{-t}(A)} f_0(x) dx$, then the transfer operator is given as

$$\mathcal{T}_t: L^1(\Omega_c) \rightarrow L^1(\Omega_c),$$

$$f \mapsto \left(q \mapsto \frac{1}{F(q)} \int_{\Omega_m} f(\pi_1 \phi^{-t}(q, p)) f_0(q, p) dp \right),$$

which is the operator that has been introduced and analyzed by Schütte [16]. This can be shown by verifying

$$\langle \mathcal{U}_t f, g \rangle_F = \langle f, \mathcal{T}_t g \rangle_F$$

for any $f \in L^\infty(F)$, $g \in L^1(F)$, which follows from Lemma 3.10 in [16].

Conclusion

In this article, we derived a generalized transfer operator for an arbitrary dynamical system. We have proven that the here defined transfer operator exists, is well-defined, posses the propagation property and can be restricted to the subspace $L^2(\mu)$ if μ is stationary.

In particular, this article provides the justification to consider transfer operators on Markov kernels as it has been done in [14, 15, 17]. Furthermore, this transfer operator includes the transfer Operator considered by Schütte [16], which does not provide the Chapman Kolmogorov property, and, therefore, does not derive from a Markov kernel.

All in all, the main achievement is that we have extended the Frobenius-Perron Operator from deterministic systems to stochastic systems.

Acknowledgment We gratefully thank the collaborative research center SFB 765 of the German Science Foundation (DFG) for financial support of Adam Nielsen and we thank the DFG research center Matheon for financial support of Konstantin Fackeldey.

References

- [1] M. C. Mackey A. Lasota, Chaos, fractals, and noise - stochastic aspects of dynamics., Springer, New York, 1994.
- [2] L. Arnold, Stochastic differential equations: Theory and applications, R. Oldenburg Verlag München Wien, 1973.
- [3] H. Bauer, Measure and integration theory., De Gruyter, Berlin, New York, 1978.
- [4] J.R. Baxter and J.S. Rosenthal, Rates of convergence for everywhere-positive markov chains, *Statistics & Probability Letters* **22** (1995), 333–338.
- [5] P.Deuffhard Ch. Schütte, W. Huisinga, Transfer operator approach to conformational dynamics in biomolecular systems, Konrad-Zuse-Zentrum für Informationstechnik Berlin (1999).

- [6] Predrag Cvitanović, Carl Philip Dettmann, Ronnie Mainieri, and Gabor Vattay, Trace formulas for stochastic evolution operators: Weak noise perturbation theory, *Journal of statistical physics* **93** (1998), no. 3-4, 981–999.
- [7] Michael Dellnitz, Gary Froyland, and Oliver Junge, The algorithms behind gaio - set oriented numerical methods for dynamical systems, In *Ergodic theory, analysis, and efficient simulation of dynamical systems*, Springer, 2000, pp. 145–174.
- [8] Michael Dellnitz, Gary Froyland, and Stefan Sertl, On the isolated spectrum of the perron-frobenius operator, *Nonlinearity* **13** (2000), no. 4, 1171.
- [9] Michael Dellnitz, Andreas Hohmann, Oliver Junge, and Martin Rumpf, Exploring invariant sets and invariant measures, *Chaos* **7** (1997), 221.
- [10] Michael Dellnitz and Oliver Junge, On the approximation of complicated dynamical behavior, *SIAM Journal on Numerical Analysis* **36** (1998), 491–515.
- [11] Jiu Ding and Aihui Zhou, Finite approximations of frobenius-perron operators. a solution of ulam’s conjecture to multi-dimensional transformations, *Physica D: Nonlinear Phenomena* **92** (1996), no. 1, 61–68.
- [12] S.N. Ethier and T.G. Kurtz, Markov processes, characterization and convergence., *Wiley Series in probability and mathematical statistics*, 1986.
- [13] Gary Froyland, Extracting dynamical behaviour via markov models, *PROCEEDINGS, NEWTON INSTITUTE*, Birkhauser, pp. 283–324.
- [14] W. Huisinga, Metastability of markovian systems: A transfer operator based approach in application to molecular dynamics, Ph.D. thesis, Freie Universität Berlin, 2001.
- [15] M. Sarich, Projected transfer operators - discretization of markov processes in high-dimensional state space., Ph.D. thesis, Freie Universität Berlin, 2011.
- [16] C. Schütte, Conformational dynamics: Modelling, theory, algorithm, and application to biomolecules., *Habilitation*, Freie Universität Berlin, 1999.
- [17] M. Weber, A subspace approach to molecular markov state models via a new infinitesimal generator., *Habilitation*, Freie Universität Berlin, 2011.