Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

A mathematical framework for general classical systems and time irreversibility as its consequence

Holger Stephan[1]

submitted: July 7, 2011

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: stephan@wias-berlin.de

No. 1629
Berlin 2011

2010 *Mathematics Subject Classification.* 82C03, 60J25, 60B05.

Key words and phrases. statistical state, second law of thermodynamics, topological bidual space, Markov operator, Jensen's inequality .

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

It is well known that important models in statistical physics like the Fokker-Planck equation satisfy an H-theorem, i.e., have a decreasing Lyapunov function (or increasing entropy). This illustrates a symmetry break in time and reflects the second law of thermodynamics. In this paper, we show that any physically reasonable classical system has to have this property. For this purpose, we develop an abstract mathematical framework based on the theory of compact topological spaces and convex analysis. Precisely, we show:

1) Any statistical state space can be described as the convex hull of the image of the canonical embedding of the bidual space of its deterministic state space (a compact topological Hausdorff space).

2) The change of any statistical state is effected by the adjoint of a Markov operator acting in the space of observables.

3) Any Markov operator satisfies a wide class of inequalities, generated by arbitrary convex functions. As a corollary, these inequalities imply a time monotone behavior of the solution of the corresponding evolution equations.

Moreover, due to the general abstract setting, the proof of the underlying inequalities is very simple and therefore illustrates, where time symmetry breaks: A model is time reversible for any states if and only if the corresponding Markov operator is a deterministic one with dense range.

In addition, the proposed framework provides information about the structure of microscopic evolution equations, the choice of the best function spaces for their analysis and the derivation of macroscopic evolution equations.

Contents

1 Introduction

For many models, describing the evolution of physical systems, a time decreasing Lyapunov function is known. One example for this is the Brownian motion of a particle, governed by the Fokker-Planck equation in an external field V. Assuming that $u(x,t)$ is the probability density of finding the particle near the point $x \in \mathbb{R}^3$ at time t, the Fokker-Planck equation (in the following denoted by FPE) reads

$$\dot{u} = \nabla^* \bullet \big(\nabla u + u \nabla V\big) \tag{1}$$

Here ∇^* and ∇ are the div and the grad operators, resp. and $\bullet$ denotes the scalar product in the Euclidean space. It is well known (see, e.g., [2]) that the functional

$$H(t) = \Phi[u] = \int u(x,t) \log u(x,t) dx + \int u(x,t) V(x) dx \tag{2}$$

is a Lyapunov function for this equation, i.e., H has the property

$$H(t_1) > H(t_2),\ t_1 < t_2 \ . \tag{3}$$

Moreover, if the potential $V(x)$ is uniformly convex, exponential decay of $u(x,t)$ towards the – suitable normalized – equilibrium $u_\infty(x) = Ce^{-V(x)}$ can be shown [3]. Using this, we can present $H(t)$ from (2) in the typical entropy-form

$$H(t) = F[u] = \int u(x,t) \log \frac{u(x,t)}{u_\infty(x)} dx = \int \frac{u(x,t)}{u_\infty(x)} \log \frac{u(x,t)}{u_\infty(x)} u_\infty(x) dx \tag{4}$$

Sometimes, this Lyapunov function is called free energy or (negative) entropy and the time monotonicity of this function is an expression of the second law of thermodynamics. If $u(x,t_1)$ and $u(x,t_2)$ are two densities at different times $t_2 > t_1$ then a physical change from $u(x,t_1)$ to $u(x,t_2)$ is possible, whereas the opposite process was never observed. From a mathematical point of view there is an operator $\mathbf{T}(t_2,t_1)$ (the solution operator or propagator of equation (1)) with $u(x,t_2) = \mathbf{T}(t_2,t_1)u(x,t_1)$ but there is no operator propagating $u(x,t_2)$ to $u(x,t_1)$.
An old question – firstly stated in the discussion of Boltzmann's equation – is for the origin of this behavior, as a system of classical particles behaves in a reversible manner.
The Fokker-Planck equation can be derived – at least in principle – in the following way: We start with a Hamilton system for $n+1$ particles, one Brownian particle and n particles of the medium. Given initial conditions and assuming adiabatic walls, the evolution of this (nonlinear) system is determined and time reversible.
Because the initial position of the particles cannot be known exactly we pass to the (linear) Liouville equation describing the same problem but the unknown is no longer the trajectory of the particles but a $(n+1)$-particle density function. The Liouville equation is time reversible, too. Now, integrating over the n medium particle, passing to the limit $n \longrightarrow \infty$ and doing some other simplifications, we obtain a more macroscopic equation than the Hamilton system. This is a linear one-particle (v,x)-Fokker-Planck equation for a density function $f(v,x,t)$ in phase space. This density function describes only the state of the one Brownian particle of interest. The influence of the particles of the medium is included via the diffusion part of the equation. This equation is already time irreversible and possesses Lyapunov functions. This time irreversibility occurs seemingly due to "some other simplifications".

Expanding this equation in moments, we obtain equation (1), describing the evolution of the pure spatial density (a rigorous procedure for this step can be found in [10]). As already mentioned, this equation has a time decreasing Lyapunov function and therefore is time irreversible. Both Fokker-Planck equations, for the spatial density and for the density in phase space are one-particle equations. They describe the motion of one particle with no exact initial state and under the influence of stochastic forces.
The question can be posed, if it can happen that by deriving an other macroscopic equation from an other Hamilton system for an other problem we end up with an equation for which the same $H(t)$ is time increasing.
In general, the direct way from a Hamilton system via the Liouville equation to a Fokker-Planck equation is not practicable. On the other hand, we can forget about the derivation and simply assume that we consider a Brownian particle moving in a medium under the influence of a random force. This was how Einstein, Smoluchowski and Langevin derived the diffusion equation (equation (1) without potential $V(x)$). This way can be generalized: Without caring about the underlying interaction, we assume that a particle moves in some random medium, and derive an equation under this hypothesis. The properties of this random force determine the coefficients of this equation. We assume that the coefficients of this equation can be derived in some way, for example by calculation from microscopic considerations or from measurements by solving some inverse problem. We have to make sure that this equation is physically sensible. Here, one important point is positivity, if the solution of this equation is a physical quantity like density, concentration or average number of particles coming from a probability nature of the problem. (In the whole paper the expression "positive" means non negative.)
Again, we can ask, whether $H(t)$ can be time increasing for such an equation. The answer to both questions is: No! $H(t)$ is time decreasing or constant for any such linear equation. Moreover, equation (1) has many other time decreasing Lyapunov functions and no one of them can be time increasing for a physically sensible equation (this notion will be defined more precisely in the following). This behavior is not connected with the underlying evolution equations. Any change of a classical physical state has a direction as soon as random processes are taken into account.
In this paper we will show this in a rigorous way. To inclose essentially arbitrary classical physical systems and to show the above statement precisely a mathematical framework from classical functional analysis will be needed. This framework is the theory of topological spaces and their duals – the spaces of continuous functions and Radon measures on this topological space together with convex analysis. This framework is well known since the first half of the 20^{th} century but it seems to be not common in statistical physics yet.
The aim of this paper is to show
Theorem 1: A statistical state of any closed classical system can be described by an element of the convex hull of a compact topological space. These are probability measures.
Theorem 2: The topological structure of the state space is generated by a suitable chosen set of continuous functions – the observables.
Theorem 3: The change of a statistical state is effected by an adjoint of a Markov operator acting in the space of observables.
Theorem 4: Any Markov operator satisfies a wide class of inequalities, generated by arbitrary convex functions which imply monotonicity for the state change of the system. These inequalities show, what state can be reached from an other state. When dealing with time depending problems the inequalities generate Lyapunov functions which signify time irreversibility and show the logical foundation of the second law of thermodynamics.

The proofs of this theorems result from the explanation in section 2. The theorems are available for any situation whenever statistical states change. The simple proofs of the underlying inequalities illustrate the reason of time irreversibility: A model is time reversible for any states if and only if the corresponding Markov operator is a deterministic one with dense range.
Due to the fact that the main interesting states changes are linear evolution equations we will consider such equations in detail as special cases. To distinguish various evolution equations, we introduce the following notions: We call an equation microscopic, if it is derived from first principles, i.e., there is no need for a deeper description of the problem. We call an equation macroscopic, if it is derived from a microscopic one in a more or less rigorous manner. We call an equation empirical, if it is postulated without strong derivation.
Beside the time irreversibility, the presented framework suggest some more consequences concerning evolution equations:
Proposal 1: The inequalities for Markov operators indicate the optimal function spaces for the densities of the probability measures. In particular, the L_2 space over the stationary measure turns out to be the right Hilbert space for the operator in an evolution equation to be self-adjoint and dissipative.
Proposal 2: No special Lyapunov function stands out to be the entropy of a microscopic equation. Postulating empirical equations for the description of many-particle problems by the modification of one-particle microscopic equations, one has to choose one of the Lyapunov functions for a distribution (Boltzmann, Bose-Einstein, Fermi-Dirac, ...) to describe how the particles behave like an ensemble.
Proposal 3: Choosing suitable observables – a partition of unity – an approximation scheme can be derived that can be understood as a finite physical system with discrete topology.
These natural proposals results in section 4.
The paper is organized as follows: In section 2 we will define states, observables, statistical states and state changes in a soft way necessary for a principle understanding. The whole mathematical framework in a "hard core" version including the main proofs or references to proofs will be explained in an appendix. Sometimes, we refer to the appendix by numbers A.$i.j.k$. In section 3 we will consider evolution equations as a special case of state changes and describe the structure of physically sensitive evolution equations. In section 4 we will collect the main consequences and explain some definitions in more detail.

2 Physical states, observables and state changes

We start to develop a mathematical framework for the description of a general classical physical system. Modeling a physical problem, we have the following typical situation: Either the model is very general, applicable for a large class of problems but is so unspecific that not much information can be extracted from it, or the model is very specific and carries a lot of information but only for a small set of problems. The aim of the present paper is to find an optimum here.
In the following, we will concentrate on classical physical systems. This means, we will assume that the state of the system – all degrees of freedom – is well defined. Moreover, we will assume that the system is observable and that observables commute (this will become clear in the following). In this sense, we will call a system "classical". If the system consists of identical subsystems, say particles, they are distinguishable (numberable). A system of n particles is one system, defined by one parameter (this can be a set of parameters). The real nature of the

system will not be of our interest.

2.1 Deterministic Description

2.1.1 Deterministic States

Let us assume that the state of a classical physical system is completely described by a parameter z of arbitrary nature. We assume that this parameter contains all information in such a way that two different systems can be distinguished by two different states.
The set of all possible states is $\mathcal{Z}$, the state space. To start modeling a system, the state space should be given. We assume that we are able to know the state of the system in principle, although usually, it is not possible to know the state exactly for a given system.
Typical examples for classical physical systems are systems of n classical particles (the state space is a subset of $\mathbb{R}^{6n}$) or discrete systems with a finite or infinite number of degrees of freedom which are often used in biology, economics or chemistry (reactions).

2.1.2 Observables

So far, the state space is an arbitrary set. To deal with it, we have to overlay a suitable structure that has a physical sense and is not too strict to cover all problems but is strict enough to allow interesting analytical results. The objects, generating such a structure in the state space are observables (measurements). The role of observables is crucial for what follows. States and the state space are physical objects that we want to understand whereas the observables are not physical objects. We can choose them as we wish. They are the questions we ask to understand the system.
An observable of a classical system is a real valued function $g : \mathcal{Z} \longrightarrow \mathbb{R}$, defined for all states $z \in \mathcal{Z}$. Being real valued functions, observables carry forward the structure of the reals. The three main structures of the real numbers are

- topological structure (limits, closeness)
- algebraic structure (adding and multiplying measurement results)
- order structure (comparing measurement results)

The second and third structure induce a structure in the set of observables whereas the topological structure in the reals induces a structure in the state space $\mathcal{Z}$. Naturally, we call two states closed to each other if the result of all observables of interest are closed to each other. Thus, given a suitable set of observables, these functions induce – via the pre-image of the open sets in $\mathbb{R}$ – open sets in $\mathcal{Z}$, a topology. Letting g be an observable and B an open set in $\mathbb{R}$, we define a topology in $\mathcal{Z}$ as generated by the sets $g^{-1}(B)$ for all observables and all open sets $B \subset \mathbb{R}$. From the topological definition of continuity (A.1.1.5), all observables used for this construction are continuous functions by definition (A.1.2.6).
A natural assumption about the observables we are interested in is that there are only countable many and that they can distinguish different states (if not, we can identify un-distinguishable states). Thus, $\mathcal{Z}$ becomes a compact topological Hausdorff space.

2.2 Stochastic Description

Probabilities come into the play for two reasons. 1) The state is not determined; and 2) the state change is not determined, i.e., the state after a state change is not determined.

The correct definition of probabilities has to start with the state space and its topology and is important for what follows. The interplay between the state space and its topology, the observables as continuous functions and measures as functions defined on Borel sets induced by the topology is essential for the following results.

2.2.1 Probabilities

If the state is not determined, the system is located in different states for different experiments. Let us assume that we make m experiments and the system is k_i times in the state z_i with $m = k_1 + k_2 + k_3 + ...$ Then, we can say that the system is in a mixed state – in contrast to the pure states z_i – and naturally we call the expression

$$\text{“}z\text{”} = \frac{k_1}{m}\ \text{“}z_1\text{”} + \frac{k_2}{m}\ \text{“}z_2\text{”} + \frac{k_3}{m}\ \text{“}z_3\text{”} + ... \quad \text{with} \quad m = k_1 + k_2 + k_3 + ...$$

the statistical state of the system. Thus, the statistical state is something like a convex combination of pure states. The system is in the state z_i with probability $p_i = \frac{k_i}{m}$.
This seems natural but in a mathematical sense the meaning of a “convex combination” of points of a topological space without any linear structure is not clear, in general. Even if there is a natural linear structure in $\mathcal{Z}$ (for example if $\mathcal{Z}$ is a part of $\mathbb{R}$) the meaning of “the system is in state $\frac{1}{2}z_1 + \frac{1}{2}z_2$” is different from the meaning of “the system is in the states z_1 and z_2 with probability $\frac{1}{2}$”.
What we need is – instead of $\mathcal{Z}$ – a set $\mathcal{Z}'$ with a linear structure and a one-to-one correspondence to $\mathcal{Z}$ to define the convex hull $\mathcal{Z}''$ of $\mathcal{Z}'$ as mixed states in such a way that the extremal elements (the elements that can not be written as a convex combination) of $\mathcal{Z}''$ are $\mathcal{Z}'$.

$$\begin{aligned} \mathcal{Z} \iff\ & \mathcal{Z}' \subset \mathcal{Z}'' \\ & \mathcal{Z}'' = \ \text{conv}\ (\mathcal{Z}') \quad \text{(mixed states)} \\ & \mathcal{Z}' = \ \text{extr}\ (\mathcal{Z}'') \quad \text{(pure states)} \end{aligned}$$

This means: Any element of $\mathcal{Z}'$ can be understood as a pure state, any element of $\mathcal{Z}''$ can be understood as a pure or mixed state and any element of $\mathcal{Z}''$ that is not an element of $\mathcal{Z}'$ can not be understood as a pure state.
Such a construction is well known in functional analysis and is the embedding of a topological space in its bidual. This construction allows us to consider pure and mixed states (and nothing else) in a unique picture together with the observables.

2.2.2 Statistical states

Let $\mathcal{Z}$ be a compact topological Hausdorff space and $\mathcal{O}(\mathcal{Z})$ the set of open sets in $\mathcal{Z}$ defining the topology. The dual $\mathcal{Z}^*$ of a topological space is the space of functionals on $\mathcal{Z}$, compatible with the topology – the space $\mathcal{C}(\mathcal{Z})$ of continuous real-valued functions on $\mathcal{Z}$. This is the space of observables.
The bidual $\mathcal{Z}^{**}$ of $\mathcal{Z}$ – the dual of $\mathcal{C}(\mathcal{Z})^*$ – is the space of linear functionals on $\mathcal{C}(\mathcal{Z})$ – the space of Radon measures on $\mathcal{Z}$. These measures are real-valued functions on the Borel sets $\mathcal{B}(\mathcal{Z})$ of $\mathcal{Z}$ generated by the open sets $\mathcal{O}(\mathcal{Z})$ in $\mathcal{Z}$.

$$\begin{aligned} \mathcal{Z} \quad & \overset{\text{dual}}{\Longrightarrow} \quad \mathcal{Z}^* = \mathcal{C}(\mathcal{Z}) \quad \text{(continuous functions)} \\ \mathcal{Z}^* \quad & \overset{\text{dual}}{\Longrightarrow} \quad \mathcal{Z}^{**} = \mathcal{C}^*(\mathcal{Z}) \quad \text{(Radon measures)} \end{aligned}$$

We point out one important continuous function $\mathbb{1}$ that maps every point $z \in \mathcal{Z}$ to 1, i.e., $\mathbb{1}(z) = 1$. This function is continuous in any topological space.
A continuous function is said to be positive, written $g \geq 0$, if $g(z) \geq 0$ for all $z \in \mathcal{Z}$. The pointwise multiplication of two functions f and g we denote by $f \cdot g$. Non trivial theorems show that the norm in $\mathcal{C}$, the order and the pointwise multiplication are consistent with each other and make $\mathcal{C}$ to a Banach space, Banach lattice and Banach algebra.
The dual product or dual pairing of a continuous function $g \in \mathcal{C}(\mathcal{Z})$ and a Radon measure $p \in \mathcal{C}(\mathcal{Z})^*$, denoted by

$$\overline{g} = \langle g, p \rangle = \int_{\mathcal{Z}} g(z) p(dz) \ ,$$

is the natural observable of a statistical state – the mean value.
It is well known that a topological space is a subset of its bidual. The map of $\mathcal{Z}$ in its bidual $\mathcal{Z}^{**} = \mathcal{C}(\mathcal{Z})^*$ is the canonical embedding $\iota : \mathcal{Z} \longrightarrow \mathcal{Z}^{**}$. The image of a point $z \in \mathcal{Z}$ of the canonical embedding is the functional $\iota(z) \in \mathcal{C}(\mathcal{Z})^*$ that acts on any element $g \in \mathcal{C}(\mathcal{Z})$ in the same way as g acts as a functional on z, i.e.,

$$g(z) = \langle g, \iota(z) \rangle, \ g \in \mathcal{C}(\mathcal{Z})$$

There is only one element $p \in \mathcal{C}(\mathcal{Z})^*$ satisfying this equation – the point measure (or Dirac measure or "δ-function") $\iota(z) = \delta_z$. We denote the set of point measures by

$$\mathcal{S}_e^*(\mathcal{Z}) = \{\delta_z \in \mathcal{C}^*(\mathcal{Z}) \mid z \in \mathcal{Z}\} \ .$$

The canonical embedding is a one-to-one map between $\mathcal{Z}$ and $\mathcal{S}_e^*(\mathcal{Z})$ (A.1.5.6). These are the pure states. As a subset of a linear vector space, we can calculate convex combinations of elements of $\mathcal{S}_e^*(\mathcal{Z})$.

$$\mathcal{S}^*(\mathcal{Z}) = \overline{\mathrm{conv}\{\delta_z \ \in \mathcal{C}^* \mid z \in \mathcal{Z}\}}^{\,*}$$

where the over-lining means the weak* closure (A.1.4.5). It turns out that the measures in $\mathcal{S}^*(\mathcal{Z})$ are precisely the positive and normalized measures in $\mathcal{C}(\mathcal{Z})^*$

$$\mathcal{S}^*(\mathcal{Z}) = \big\{p \in \mathcal{C}^*(\mathcal{Z}) \mid p \geq 0, \ \|p\| = \langle \mathbb{1}, p \rangle = p(\mathcal{Z}) = 1\big\}$$

where the positivity $p \geq 0$ is the order structure on measures

$$p \geq 0 \overset{\text{def}}{\Longleftrightarrow} p(B) \geq 0, \ B \in \mathcal{B}(\mathcal{Z}) \ ,$$

i.e., p is positive if and only if the value on each Borel set is nonnegative. Such and only such measures can be understood as probabilities: $p(B) = \int_B p(dz)$ is the probability finding the system in a state contained in B. Note, that p is a function of sets and not a function of points like probability densities.
$\mathcal{S}^*(\mathcal{Z})$ is the statistical state space. From the definition it is clear that $\mathcal{S}^* = \mathrm{conv}(\mathcal{S}_e^*)$. Moreover, $\mathcal{S}_e^* = \mathrm{extr}(\mathcal{S}^*)$. Thus, $\mathcal{S}_e^*$ and $\mathcal{S}^*$ are precisely the desired objects $\mathcal{Z}'$ and $\mathcal{Z}''$.

$$\begin{aligned} \mathcal{Z} \iff \ & \mathcal{S}_e^* \subset \mathcal{S}^* \subset \mathcal{C}^*(\mathcal{Z}) \\ & \mathcal{S}^* = \mathrm{conv}(\mathcal{S}_e^*), \quad \text{mixed states} \\ & \mathcal{S}_e^* = \mathrm{extr}(\mathcal{S}^*), \quad \text{pure states} \end{aligned}$$

Here, the star of $\mathcal{S}^*$ denotes that $\mathcal{S}^*$ belongs to dual objects and not that $\mathcal{S}^*$ is the dual object of something.
If there is no danger of misunderstanding, we will use the notion state instead of statistical state, and omit $\mathcal{Z}$ in $\mathcal{C}(\mathcal{Z})$, $\mathcal{C}^*(\mathcal{Z})$, $\mathcal{S}^*(\mathcal{Z})$, ...

2.2.3 Observables and generalized observables

We started from a set of real-valued functions – observables – to define a suitable topology in the state space. Now, having a topology in $\mathcal{Z}$ and looking for its bidual we construct its dual $\mathcal{C}(\mathcal{Z})$ that can be understood as the space of observables in a natural way. Of course, any of the starting observables are continuous by definition of the topology and therefore they are elements of $\mathcal{C}(\mathcal{Z})$. It is important – for instance for the definition of evolution equations – that this space $\mathcal{C}(\mathcal{Z})$ is separable due to the fact that initially we took a countable set of observations to define the topology.
The characteristic function $\mathbb{1}_B$ of a given Borel set $B \in \mathcal{B}$ with $\mathbb{1}_B(z) = 1$ for $z \in B$ and $\mathbb{1}_B(z) = 0$ otherwise, is not a continuous function, in general. Nevertheless, this function is a useful observable, as it observes whether the system is in a state in B or not. Such functions are elements of $\mathcal{C}^{**}(\mathcal{Z})$ the bidual of $\mathcal{C}(\mathcal{Z})$ and the tri-dual of $\mathcal{Z}$. An exact description of $\mathcal{C}^{**}(\mathcal{Z})$ as the space of functionals of Radon measures is very difficult. For us it is only important that any measurable function on $\mathcal{Z}$ is an element of $\mathcal{C}^{**}(\mathcal{Z})$. Measurable functions can be understood as continuous functions on $\mathcal{Z}$ with the topology generated by the Borel sets $\mathcal{B}$. We denote it by $\mathcal{C}_{\mathcal{B}}(\mathcal{Z})$ and call these functions generalized observables.
The dual pairing makes sense for any observables including generalized ones

$$\overline{\xi} = \langle \xi, p \rangle = \int_{\mathcal{Z}} \xi(z) p(dz) \;, \xi \in \mathcal{C}^{**}, \; p \in \mathcal{C}^{*}$$

In probability theory generalized observables are called random variables.

2.2.4 The duality between states and observables

The dual pairing $\langle g, p \rangle = \int_{\mathcal{Z}} g(z) p(dz)$ between a continuous function g and a probability measure p is the general form of an integral, combining an intensive value g, i.e., a function of points and an extensive value p, i.e., a positive additive function of sets.
Roughly speaking, the dual pairing is a sum of terms that are products of observables for given state and the probability to find the system in this state and can be understood as a limit of the bilinear form

$$\langle g, p \rangle = \lim_{n,m\to\infty} \sum_{i,j=1}^{n,m} g(z_i) p(B_j) \Phi(z_i, B_j)$$

where $(B_j)_{j=1}^m$ is a disjunct decomposition of $\mathcal{Z}$ and $\Phi(z, B)$ is the function, counting whether a point is in a set, i.e., $\Phi(z, B) = 1$ for $z \in B$ and 0 otherwise. The functions with one argument fixed $\Phi(z, \cdot) = \delta_z$ and $\Phi(\cdot, B) = \mathbb{1}_B$ are the point measures and the characteristic functions, resp.
The goal of the investigation is to get information about the state of the system. For this purpose we use suitable observables to obtain the desired information. This is the only way to get information about the system. From a mathematical point of view the observables are test functions. This duality of states and observables is important for the derivation of approximations and of macroscopic equations from microscopic ones.

2.3 Deterministic state changes

The state of a system may change. This state change is a function $\gamma : \mathcal{Z} \to \mathcal{Z}$, mapping a state into an other state. We call γ a deterministic state change. Since we have a topology in $\mathcal{Z}$ we

consider only functions, conserving the topology – continuous functions and denote the set of such functions by $\mathrm{End}(\mathcal{Z})$ – the semigroup of endomorphisms on the topological space $\mathcal{Z}$. This is – in contrast to continuous functions with real values – not a vector space because the values of γ are points in $\mathcal{Z}$, a set without a linear structure.
A typical example of such a function is a semiflow of a dynamical system at given time (see sub-subsection 3.5.1).
Irreversibility means that there is a functional, estimating the state before the change with the state afterwards. In general, we can not expect that this is possible for arbitrary deterministic state changes. This situation becomes completely different if we consider statistical state changes. The reason is that not any change is physically sensible but only such a change for which the result can be understood as a probability measure, i.e., an element of $\mathcal{S}^*$.

2.3.1 Deterministic Markov operators

Due to the canonical embedding of $\mathcal{Z}$ in $\mathcal{S}^*$, any function $\gamma \in \mathrm{End}(\mathcal{Z})$ induces some map Γ in such a way that $z' = \gamma(z)$ implies $\delta_{z'} = \Gamma(\delta_z)$.

$$\begin{array}{cccccc} \mathcal{Z} & \ni & z & \overset{\text{emb.}}{\Longleftrightarrow} & \delta_z \in \mathcal{S}_e^* \subset \mathcal{S}^* \\ & & \gamma\big\downarrow & & \big\downarrow \Gamma = ? \\ \mathcal{Z} & \ni & \gamma(z) & \overset{\text{emb.}}{\Longleftrightarrow} & \delta_{\gamma(z)} \in \mathcal{S}_e^* \subset \mathcal{S}^* \end{array}$$

The nature of this map Γ, the image of γ, is well known. It is a linear operator, the adjoint of a deterministic Markov operator.
A deterministic Markov operator is a linear operator that depends on a continuous function $\gamma \in \mathrm{End}(\mathcal{Z})$ and acts on continuous functions $g \in \mathcal{C}(\mathcal{Z})$ via the rule

$$\big(\mathbf{M}_\gamma g\big)(z) = (g \circ \gamma)(z) = g\big(\gamma(z)\big)$$

where $\circ$ denotes the composition. We denote by

$$\mathcal{M}_e = \big\{\mathbf{M}_\gamma \in \mathcal{L}(\mathcal{C}) \;\big|\; \gamma \in \mathrm{End}(\mathcal{Z})\big\}$$

the set of deterministic Markov operators. Other terms are composition operator, Frobenius-Perron operator or Ruelle transfer operator.
Deterministic Markov operators have some obvious properties

$$\mathbf{M}_\gamma \mathbb{1} = \mathbb{1},\; \mathbf{M}_\gamma \geq 0$$

i.e., they conserve positivity and constants. Here, $\mathbf{L} \geq 0$ denotes the cone-positivity of an operator in $\mathcal{C}$, this is, the result of the action of $\mathbf{L}$ on a positive function is a positive function:

$$\mathbf{L} \geq 0 \;\overset{\text{def}}{\Longleftrightarrow}\; \mathbf{L}g \geq 0,\; g \geq 0,\; g \in \mathcal{C}(\mathcal{Z})$$

The notion of cone-positivity is in contrast to the notion of form-positivity of operators or bilinear forms in Hilbert spaces. Our main order in this paper is cone-positivity. We will call it simply positivity.
Deterministic Markov operators are bounded with $\|\mathbf{M}_\gamma\| = 1$.
The operator $\mathbf{M}_\gamma^*$, the adjoint of $\mathbf{M}_\gamma$ is a bounded and positivity conserving operator in $\mathcal{C}^*(\mathcal{Z})$ (A.2.3.5). It acts according to the rule (here $B \in \mathcal{B}$ is a Borel set)

$$\big(\mathbf{M}_\gamma^* p\big)(B) = \big\langle \mathbb{1}_B, \mathbf{M}_\gamma^* p\big\rangle = \big\langle \mathbf{M}_\gamma \mathbb{1}_B, p\big\rangle = \big\langle \mathbb{1}_B \circ \gamma, p\big\rangle = \big\langle \mathbb{1}_{\gamma^{-1}(B)}, p\big\rangle = p\big(\gamma^{-1}(B)\big)$$

Because of

$$\langle g, \mathbf{M}_\gamma^* \delta_z \rangle = \langle \mathbf{M}_\gamma g, \delta_z \rangle = g\big(\gamma(z)\big) = \langle g, \delta_{\gamma(z)} \rangle$$

we have $\mathbf{M}_\gamma^* \delta_z = \delta_{\gamma(z)}$. Thus, $\mathbf{M}_\gamma^*$ is the wanted object Γ. This is, the linear operator $\mathbf{M}_\gamma^*$ acts in $\mathcal{S}_e^*$ in the same way as γ acts in $\mathcal{Z}$.
Moreover, from

$$\|\mathbf{M}_\gamma^* p\| = \langle \mathbb{1}, \mathbf{M}_\gamma^* p \rangle = \langle \mathbf{M}_\gamma \mathbb{1}, p \rangle = \langle \mathbb{1}, p \rangle = \|p\|$$

and the positivity it follows $\mathbf{M}_\gamma^* \mathcal{S}^* \subset \mathcal{S}^*$, i.e., $\mathbf{M}_\gamma^*$ maps not only pure states to pure states but also statistical states to statistical states. Thus, $\mathbf{M}_\gamma^*$ is an example for a statistical state change. The linearity of $\mathbf{M}_\gamma^*$ underlines the property of probabilities describing superpositions of experiments and not superpositions of interactions (probabilities do not interact with each other). We want to underline that the connection between linear operators $\mathbf{M}_\gamma^*$ and "nonlinear" objects γ is really an equivalence and has nothing to do with the usual linearization of nonlinearities. Such a connection is well known from the equivalence of a dynamical system and its corresponding linear Liouville equation.

2.4 Statistical state change

So far, we have defined statistical states and the image of deterministic state changes in the statistical state space. Now, we have to define general state changes. This should be general operators $\mathbf{L} \in \mathcal{L}(\mathcal{C}^*)$ mapping statistical states into statistical states: $\mathbf{L}\mathcal{S}^* \subset \mathcal{S}^*$. Only such operators are physically sensible. We cannot understand the meaning of an operator, mapping a statistical state out of $\mathcal{S}^*$. The disadvantage of such a definition is that such a operator must not have a pre-dual one. This means, in general, there is no operator $\mathbf{N} \in \mathcal{L}(\mathcal{C})$ with $\mathbf{N}^* = \mathbf{L}$. This is caused by the non reflexivity of the space $\mathcal{C}$ (A.2.3.7).
The approach analogous to the definition of mixed states as the closure of convex combinations of pure states is not practicable either. Similarly, we cannot define a general statistical state change as the closure of convex combinations of adjoints of deterministic Markov operators. Convex combinations of adjoints of deterministic Markov operators map statistical states into statistical states, but it is not clear what closure has to be taken for this set.
Consequently, in a first step we define suitable operators in the space of observables before taking their adjoints. This makes the observables more original than statistical states.

2.4.1 Markov operators

By

$$\mathcal{M} = \big\{ \mathbf{M} \in \mathcal{L}(\mathcal{C}) \;\big|\; \mathbf{M} \geq 0,\ \mathbf{M}\mathbb{1} = \mathbb{1} \big\}$$

we define a set of linear operators in $\mathcal{C}$, the so-called Markov operators. Their adjoints, the operators $\mathcal{M}^* = \big\{ \mathbf{M}^* \in L(\mathcal{C}^*) \;\big|\; \mathbf{M} \in \mathcal{M} \big\}$ map statistical states into statistical states.

$$\mathbf{M} \geq 0,\ \mathbf{M}\mathbb{1} = \mathbb{1} \quad\Longrightarrow\quad \mathbf{M}^* \mathcal{S}^* \subset \mathcal{S}^*$$

This easily follows from

$$\|\mathbf{M}^* p\| = \langle \mathbb{1}, \mathbf{M}^* p \rangle = \langle \mathbf{M}\mathbb{1}, p \rangle = \langle \mathbb{1}, p \rangle = \|p\|$$

and the positivity of the adjoint of a positive operator. $\mathcal{M}$ and $\mathcal{M}^*$ are convex sets.
There is a simple description of Markov operators and their adjoints:
For any Markov operator $\mathbf{M}$ there can be found a function P of points $z \in \mathcal{Z}$ and Borel sets $B \in \mathcal{B}$, i.e., $P = P(z, B)$ such that $P(\cdot, B) \in \mathcal{C}$ and $P(z, \cdot) \in \mathcal{S}^*$ and

$$(\mathbf{M}g)(z) = \int_{\mathcal{Z}} g(z')P(z, dz') \tag{5}$$

Thus, a Markov operator is an integral operator with a kernel $P(z, B)$ that is a continuous function for fixed set B and a probability measure for fixed state z. The last property shows that the observable after a statistical state change is a convex combination of the observable before.
Analogously, the adjoint $\mathbf{M}^*$ acts on measures as

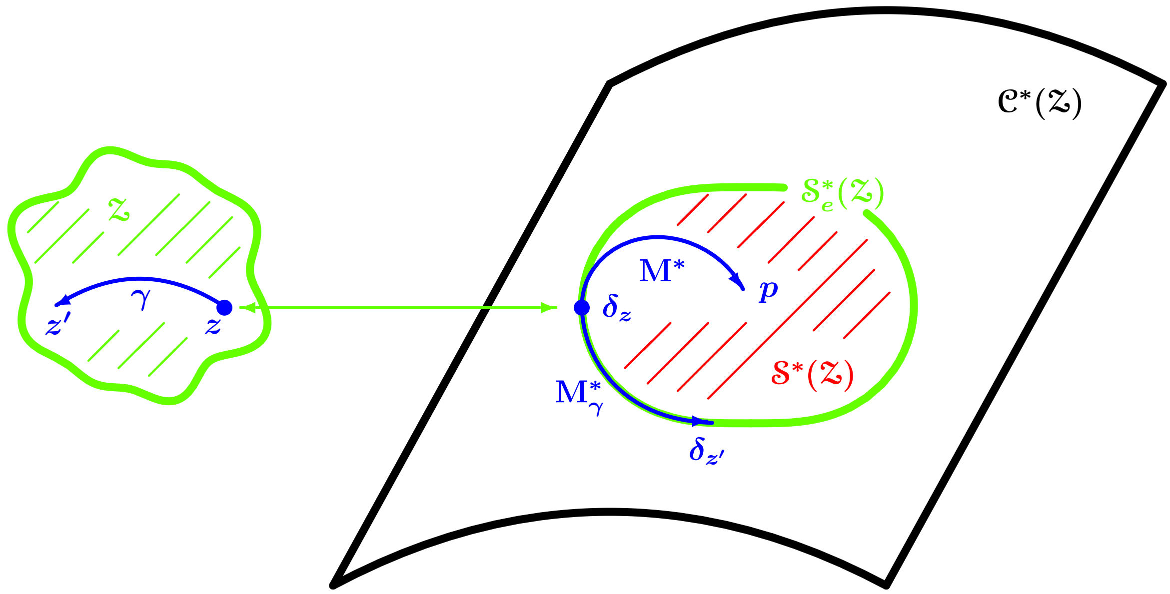
$$(\mathbf{M}^*p)(B) = \int_{\mathcal{Z}} P(z, B)p(dz) \tag{6}$$

It is an integral operator with the same kernel as $\mathcal{M}$.

2.4.2 Connection between the deterministic and stochastic worlds

The following picture is a formal illustration of the connection between the deterministic and stochastic worlds.

$$\begin{array}{ccccccccccc}
\mathcal{Z} & \ni & z & \Longleftrightarrow & \delta_z & \in & \mathcal{S}^*_e & = & \mathrm{extr}(\mathcal{S}^*) & \subset & \mathcal{S}^* \subset \mathcal{C}^* \\
\mathrm{End}(\mathcal{Z}) & \ni & \gamma & \Longleftrightarrow & \mathbf{M}_\gamma & \in & \mathcal{M}_e & = & ???(\mathcal{M}) & \subset & \mathcal{M}
\end{array}$$



Note, that $\mathrm{extr}(\mathcal{S}^*)$ – the extremal elements of $\mathcal{S}^*$ – is not something like the boundary of $\mathcal{S}^*$. The elements of $\mathcal{M}_e$ are not the extremal elements of $\mathcal{M}$. Moreover, it seems to be an open mathematical problem to find a topology in such a way that $\mathcal{M}$ is the closure of the convex hull of $\mathcal{M}_e$.

2.5 Irreversibility

2.5.1 Stationary states

The famous Perron-Frobenius-Krein-Rutman Theorem claims: For a compact $\mathcal{Z}$, any adjoint of a Markov operator $\mathbf{M}^*$ acting on $\mathcal{S}^*(\mathcal{Z})$ possess a fix point $\mu \in \mathcal{S}^*(\mathcal{Z})$, satisfying $\mathbf{M}^*\mu = \mu$. This is a stationary state of the state change $\mathbf{M}^*$. It can be not unique. Sometimes such a state is called equilibrium state. We believe this notation is reserved for a stationary state that is the time limit of an evolution process.
Note that from this theorem it follows that in particular any adjoint of a deterministic Markov operator $\mathbf{M}^*_\gamma$ has a stationary state in $\mathcal{S}^*$. This is in contrast to the situation for γ in $\mathcal{Z}$. As an example we can take a circle for $\mathcal{Z}$ and γ is a rotation of this circle around an arbitrary angle. This rotation has no fix point (unless the angle is 2π), whereas the corresponding $\mathbf{M}^*_\gamma$ has a stationary state in $\mathcal{S}^*(\mathcal{Z})$ – the homogeneous measure on the circle.

2.5.2 Jensen's inequality for Markov operators

For an arbitrary Markov operator $\mathbf{M}$, and an arbitrary convex function $F : \mathbb{R} \longrightarrow \{\mathbb{R} \cup +\infty\}$ Jensen's inequality

$$F(\mathbf{M}g) \leq \mathbf{M}F(g)\ , \mathbf{M} \in \mathcal{M},\ g \in \mathcal{C}(\mathcal{Z}) \tag{7}$$

holds. The inequality has to be understood as an inequality between continuous functions. It means that for any $z \in \mathcal{Z}$, $F((\mathbf{M}g)(z)) \leq (\mathbf{M}F(g))(z)$. This is a simple consequence of Jensen's inequality for sequences. Equality holds for strictly convex functions if and only if $\mathbf{M}$ is a deterministic Markov operator.
Pairing this inequality with a stationary measure μ, we get

$$\langle F(\mathbf{M}g), \mu\rangle \leq \langle \mathbf{M}F(g), \mu\rangle = \langle F(g), \mathbf{M}^*\mu\rangle = \langle F(g), \mu\rangle \tag{8}$$

Thus, $\langle F(g), \mu\rangle$ is a value that cannot increase after a state change. This already shows an asymmetry of statistical state changes.

2.5.3 Jensen's inequality for adjoints of Markov operators

Now, we carry over inequality (7) from the space of observables $\mathcal{C}$ to the convex set of probability measures $\mathcal{S}^*$.
For a suitable normalization (boundedness from below) we consider convex functions $F : \mathbb{R} \rightarrow \{\mathbb{R} \cup +\infty\}$ with $F(1) = 0$. Any convex function can be normalized in such a way by adding a suitable linear function.
We define a functional H on two arbitrary probability measures p and q as a generalized Legendre transform of $\langle F(g), q\rangle$

$$H[p, q] = \sup_{g\in\mathcal{C}} \big(\langle g, p\rangle - \langle F(g), q\rangle\big),\ p, q \in \mathcal{S}^* \tag{9}$$

Then, for an arbitrary Markov operator $\mathbf{M}$ inequality

$$0 \leq H[\mathbf{M}^*p, \mathbf{M}^*q] \quad \leq \quad H[p, q],\ p, q \in \mathcal{S}^* \tag{10}$$

holds. For a strictly convex F equality holds for any states if and only if $\mathbf{M}$ is a deterministic Markov operator with dense range (A.3.3.3).

As a special case, for q we can choose the stationary state μ of $\mathbf{M}^*$. Then, we have $\mathbf{M}^*\mu = \mu$ and (10) becomes

$$0 \leq H[\mathbf{M}^* p, \mu] \quad \leq \quad H[p, \mu], \;\; p \in \mathcal{S}^* \tag{11}$$

Inequalities (10) and (11) mean that there is a functional H for which the distance between two arbitrary states or between one arbitrary state and the equilibrium state before and after a state change can not increase. Moreover, any convex function F generates such a functional and the inequality is valid for any Markov operator.
The meaning of the functional (9) is not straightforward. To illustrate it, we consider the case that p has a density h with respect to q. Then,

$$H[p, q] = \sup_{g \in \mathcal{C}} \big(\langle g, p\rangle - \langle F(g), q\rangle\big) = \int_{\mathcal{Z}} F^*\big(h(z)\big) q(dz) \tag{12}$$

where F^* is the convex conjugate of F. If p and q have a densities with respect to the Lebesgue measure, say $p(dz) = h_p(z)dz$ and $q(dz) = h_q(z)dz$, the expression of H simplifies to

$$H[h_p, h_q] = \int_{\mathcal{Z}} F^*\big(h(z)\big) q(dz) = \int_{\mathcal{Z}} F^* \left(\frac{h_p(z)}{h_q(z)}\right) h_q(z) dz \; . \tag{13}$$

This is the often used expression for the entropy (see, e.g., [2] with the restriction that q has to be the stationary measure). Note, that the general expression for this functional, valid in any situation and for non stationary measures, is (9).
The inequalities (7) and (10) are the main inequalities, appropriable in any situations. The inequalities (8) and (11) require the use of the stationary measure. In degenerate situations this measure can be concentrated in a small part of the state space. Then, these inequalities carry only sparse information (see the remark in subsection 3.4).
As an example we can take the convex function $F(x) = \frac{1}{r}|x|^r$ with $r > 1$ and its convex conjugate $F^*(x) = \frac{1}{r'}|x|^{r'}$ with $\frac{1}{r} + \frac{1}{r'} = 1$ and conclude from (8) and (13)

$$\int_{\mathcal{Z}} |(\mathbf{M} g)(z)|^r \mu(dz) \quad \leq \quad \int_{\mathcal{Z}} |g(z)|^r \mu(dz) \tag{14}$$

$$\int_{\mathcal{Z}} |(\mathbf{M}^* p)_\mu(z)|^{r'} \mu(dz) \quad \leq \quad \int_{\mathcal{Z}} |p_\mu(z)|^{r'} \mu(dz) \tag{15}$$

here p_μ denotes the density of p with respect to μ.

2.5.4 A summary of the main postulates

The construction of the mathematical description is based on the following postulates. Whenever we deal with a system and these postulates are fulfilled, all results in this paper can be applied to this system.
Postulate 1: The states z and the state space $\mathcal{Z}$ are well defined. $\mathcal{Z}$ does not change in time and consists of all possible states. (completeness)
Postulate 2: $\mathcal{Z}$ is compact, this is: There is a "countable number of observables of interest" (see the remark in sub-subsection 4.3.1). (compactness and separability)
Postulate 3: Observables are real valued functions and provide a commutative algebra (classical systems).
Postulate 4: Statistical state change do not depend on states (linearity).

Postulate 5: State changes in the statistical spaces have an analogon in the space of observables (existence of a pre-dual operator).

For any such systems there is a large family of functionals, generated by any convex function that decreases for a state change except for the pure deterministic case. Moreover, this family of functionals is independent on the system, on the states or on the changes. As a special case, these functionals reduce the distance to any stationary state after a change.

For simple situations, for example if the state space is the Euclidean space, the evolution equation of interest is the FPE with smooth potential and the preferred measure is the Lebesgue measure, the proposed framework does not generate new results. The gain of the construction lies in its generality. For example, when dealing with state spaces with interfaces, various dimensions, disconnected or discrete parts the right set of probability measures is not obvious. For example, it may happen that we have two different point measures in one point due to the fact that a function with a jump at an interface is included in the set of continuous functions. Once the topology in $\mathcal{Z}$ is given, there is nothing to define more. Any probability measure lies in the corresponding convex hull in the bidual of $\mathcal{Z}$ and there is nothing else.

3 Evolution problems

In this paper we investigate irreversibility via the study of state changes and not via evolution equations because of greater generality. Often, the origin of the investigation of irreversibility is an evolution equation or a class of such equations as $\dot{g}(t) = \mathbf{A}g(t)$ where the dot denotes the time derivative. This equation is used to show inequalities like (3) of the form $\frac{d}{dt}H(t) \le 0$. Formally, such inequalities are easy to derive. For example, inequality (30) follows straightforward from (29) . But a rigorous derivation of such an inequality requires regularity conditions (because usually $\mathbf{A}$ is a unbounded operator) that restrict the field of application of the method.
Otherwise, the solution operators of these equations are state changes for fixed time and therefore bounded operators that can be handled in an easier way. However, irreversibility has to be shown via an inequality of the form (3) without $\frac{d}{dt}H(t)$ which is formally more difficult.
Moreover, not every state change can be realized by an evolution equation. For example, consider a system in two states, say a ball that can be white or black. At one time it is white, at another time it is black. For this change, an evolution equation can never be provided.
Thus, in the following we will consider evolution equations as special cases of state changes.

3.1 Examples of states and state spaces

Typical examples for state spaces are

- $z = (v, x)$ Brownian particle
- $z = (x)$ Idealized Brownian particle with given velocity distribution
- m classical particles with generalized coordinates q_i and momentums p_i. The state $z = (p_1, ..., p_m, q_1, ..., q_m) \in \mathcal{Z} \subset \mathbb{R}_{6m}$ is a point in the Euclidean space.
- m chemically reacting stirred species $\mathbf{X}_1, ..., \mathbf{X}_m$. k_i is the number of particles (molecules) of $\mathbf{X}_i$ and the state is $z = (k_1, ..., k_m) \in \mathcal{Z} = \mathbb{N}_0^m$. We know all about the system, if we know each number of all possible molecules.

  Examining only stirred particles implies that we do not have to take into consideration the space distribution of the particles as they are homogeneously distributed.
- The state space $\mathcal{Z} = \{z_1, z_2, ..., z_n\}$ is finite, if the system can be only in a finite number of states. This situation is typical for discrete models like cellular automates in economy, biology, society or reactions with a conserved number of particles.

  Moreover, systems with a finite number of states arise in the approximation of continuous systems like diffusion equations or solid body dynamics. In these cases the finite system can be understood, as a system of pots connected by tubes with asymmetric valves where the content of the pots changes by flows through the tubes.

The structure of evolution equations of the type $\dot{g}(t) = \mathbf{A}g(t)$, i.e., the set of possible operators $\mathbf{A}$ in this equation is a result of the interplay between the topology in $\mathcal{Z}$ and the topology of the reals for time. If the topology for the time is discrete, i.e., time is numbered by integers, in the equation $g(n+1) = \mathbf{A}g(n)$, there is nothing to investigate – $\mathbf{A}$ can be any Markov operator.
If time is a real number, it depends on the topology in $\mathcal{Z}$ what operators $\mathbf{A}$ are possible. There is a rule of thumb: The more discrete the topology in $\mathcal{Z}$, the purer the set of operators $\mathbf{A}$.
As examples, we restrict ourself to two abstract and often used state spaces:

- $\mathcal{Z} = \{z_1, z_2, ..., z_n\}$ (finite state space with discrete topology)

  Due to the discrete topology, any real function of n arguments is a continuous one. Thus, $\mathcal{C}(\mathcal{Z}) = \mathbb{R}^n$ with the sup-norm and $\mathcal{C}^*(\mathcal{Z}) = \mathbb{R}^n$ with the L_1-norm.

  $\mathcal{S}^*$ is the $(n-1)$-dim Simplex in $\mathbb{R}^n$ and $\mathcal{S}^*_e$ are its vertices, i.e., the ordinary base vectors.

  Operators in $\mathcal{C}(\mathcal{Z})$ and $\mathcal{C}^*(\mathcal{Z})$ are matrices. Markov operators are stochastic matrices $(m_{ij})_{i,j=1}^n$, i.e., matrices with positive entries $m_{ij} \geq 1$ and row sum $\sum_{j=1} m_{ij} = 1$. Thus, $m_{i1}, m_{i2}, ..., m_{in}$ are the coefficients of a convex combination.

  Adjoints of Markov operators are their transposed, i.e., matrices with positive entries and column sum 1.

- $\mathcal{Z} \subset \mathbb{R}^m$ (a compact or suitable compactified domain in the Euclidean space)

  $\mathcal{C}(\mathcal{Z})$ and $\mathcal{C}^*(\mathcal{Z})$ are the well known Banach spaces of continuous functions and Radon measures on $\mathcal{Z}$.

3.2 Semi-groups of Markov operators and evolution equations

In this subsection we analyze the special case, in which the state change is caused by an evolution equation. This is, we consider trajectories of statistical states $p(t)$, where $t \geq 0$ is a real time parameter. If $p(t_2)$ and $p(t_1)$ are two states for different times $t_2 > t_1 \geq 0$, then an operator $\mathbf{T}^*(t_2, t_1)$ acting as $p(t_2) = \mathbf{T}^*(t_2, t_1)p(t_1)$ should be the adjoint of a Markov operator $\mathbf{T}(t_2, t_1)$.
The corresponding evolution equation would be an equation like $\dot{p}(t) = \mathbf{A}^*(t)p(t)$ with an initial state, say $p(0) = p_0$. Here, we restrict ourselves to equations $\dot{p}(t) = \mathbf{A}^*p(t)$ with operators, not depending on time, explicitely (for the general situation see the remark in sub-subsection 4.3.3).
This implies that $\mathbf{T}^*(t_2, t_1)$ is a function of only the difference $t_2 - t_1$. Thus, the operator family $\mathbf{T}^*(t)$ is a semigroup with the semigroup property $\mathbf{T}^*(t_2 + t_1) = \mathbf{T}^*(t_2)\mathbf{T}^*(t_1) = \mathbf{T}^*(t_1)\mathbf{T}^*(t_2)$ and $\mathbf{T}^*(0) = \mathbf{I}^*$ the identity in $\mathcal{C}^*$. Formal, $p(t) = \mathbf{T}^*(t)p_0$ is the solution to some equation $\dot{p}(t) = \mathbf{A}^*p(t)$. This connection is right in a rigorous way, if $\mathbf{T}^*(t)$ is a strong continuous semigroup in time. Then, $\mathbf{A}^*$ is its generator (A.4.1.4).
Due to the very "coarse" topology in $\mathcal{C}^*(\mathcal{Z})$, there are no strong continuous (even no weak continuous) semigroups, except those with bounded generators. Thus, there are no interesting evolution equations in $\mathcal{C}^*(\mathcal{Z})$. This is the second reason (the first was the non existence of a pre-dual of a general bounded operator in a non reflexive function space) to study the problem at first in the space of observables. In this space we have a necessary and sufficient condition for a generator of a continuous semigroup of Markov operators.
Following [1], we will say that an operator $\mathbf{A}$ with dense domain $D(\mathbf{A}) \subset \mathcal{C}(\mathcal{Z})$ satisfies the positive minimum principle, if

$$(\mathbf{A}g)(z_+) \leq 0, \ g \in D(\mathbf{A}) \ , \tag{16}$$

where z_+ is the point where g contains its maximum: $g(z_+) = \sup_{z\in\mathcal{Z}} g(z)$ (A.1.2.12). A generator of a continuous semigroup in $\mathcal{C}(\mathcal{Z})$ is a generator of a positive continuous semigroup if and only if it satisfies (16). If additionally $\mathbf{A}\,\mathbb{1} = 0$, then the solution to the equation

$$\dot{g}(t) = \mathbf{A}g(t), \ g(0) = g_0 \tag{17}$$

is $g(t) = \mathbf{T}(t)g_0$, where $\mathbf{T}(t)$ is a semigroup of Markov operators. Formally $\mathbf{A} = \mathbf{T}'(0)$ (see A.4.1.4).

This is a surprising statement, because the test of (16) for a given generator $\mathbf{A}$ is very simple and often can be done straightforward. The main analytical problem is to show that a given operator is a generator at all.
Moreover, the solution $g(t)$ satisfy (17) is a classical sense.
Once, $\mathbf{T}(t)$ is a semigroup, $\mathbf{T}^*(t)$ is a semigroup, too. But due to the missing strong continuity, an evolution equation for $p(t) = \mathbf{T}^*(t)p_0$ holds only in a weak* sense, i.e.,

$$\frac{d}{dt}\big\langle g, p(t)\big\rangle = \big\langle \mathbf{A}g, p(t)\big\rangle,\ g \in D(\mathbf{A}) \ . \tag{18}$$

$\mathbf{A}^*$ exists since $D(\mathbf{A})$ is dense in $\mathcal{C}(\mathcal{Z})$, so we can write $\big\langle \mathbf{A}g, p(t)\big\rangle = \big\langle g, \mathbf{A}^* p(t)\big\rangle$, i.e., $p(t)$ is the solution with $p(0) = p_0$ of some equation in weak* sense. We will write this in the following way

$$\dot{p}(t) \overset{*}{=} \mathbf{A}^* p(t),\ p(0) = p_0 \ . \tag{19}$$

In general, it is very difficult and no need to write down $\mathbf{A}^*$ explicitly. From a practical point of view, it is better to solve equation (17) and calculate the adjoint $\mathbf{T}^*(t)$ than to solve (19).
Nevertheless, under some regularity conditions and if $p(t)$ has a density $W(t)$ with respect to some given measure, a formal equation similar to $\dot{W}(t) = \mathbf{A}^* W(t)$ can be derived. One example is the FPE (1). The real meaning of such equation and its range of application are discussed in sub-subsection 4.2.3.
In the literature on probability theory equation (17) is called Chapman-Kolmogorov backward equation and equation (19) forward equation. This notion becomes clear if one considers evolution equations, depending on time explicitely (see, e.g., [6]).
As analogon to quantum mechanics, we can call equation (17) – describing the evolution of an observable – *Heisenberg representation* or *Heisenberg picture*. Equation (19) is the corresponding *Schrödinger picture*. In contrast to quantum mechanics – dealing with a Hilbert space – both pictures are not equivalent because $\mathcal{C}(\mathcal{Z})$ is a non reflexive space. This non-reflexivity is of principial nature, because the sense of the crossover from $\mathcal{Z}$ to $\mathcal{S}^*$ was to enlarge the set of states by mixed states.
Due to this asymmetry of the pictures the *Heisenberg pictures* is the right one to start with.
Solving a given problem for a given state space $\mathcal{Z}$ and a given initial probability measure p_0, we have to do the following steps:

1) Derive (or guess) the operator $\mathbf{A}$ describing the desired evolution.
2) Check the positive minimum principle (16).
3) Show that $\mathbf{A}$ is a generator (A.4.2.7).
4) Calculate the solution operator $\mathbf{T}(t)$. This is an integral operator like (5) with a kernel $G(z, B, t)$ (this is Green's function) given by a family of probability measures $G(z, \cdot, t)$.
5) Calculate $p(t) = \mathbf{T}^*(t)p_0$, where $\mathbf{T}^*(t)$ is also an integral operator like (6) with the same kernel as $\mathbf{T}(t)$.

3.3 Stationary states

In sub-subsection 2.5.1 we mentioned that any adjoint of a Markov operator possess a fix point $\mu \in \mathcal{S}^*(\mathcal{Z})$, satisfying $\mathbf{M}^*\mu = \mu$. As $\mathbf{T}(t)$ is a Markov operator for fixed t, there exists a fix point $\mu \in \mathcal{S}^*(\mathcal{Z})$, satisfying $\mathbf{T}^*(t)\mu = \mu$. Actually, much more is true as the Markov-Kakutani

Theorem claims: For any semigroup $\mathbf{T}(t)$ of Markov operators there exists a not necessarily unique probability measure $\mu \in \mathcal{S}^*(\mathcal{Z})$ such that $\mathbf{T}^*(t)\mu = \mu$ for any t. This μ solves $\mathbf{A}^*\mu = 0$. The converse is also true: If μ is a solution to $\mathbf{A}^*\mu = 0$, then $\mathbf{T}^*(t)\mu = \mu$ for any t.
Note, that this does not imply that μ is an equilibrium state, i.e., $\mathbf{T}^*(t)p_0$ need not to converges towards μ for any or even some $p_0 \in \mathcal{S}^*(\mathcal{Z})$.

3.4 Inequalities and Lyapunov functions

Given a general inequality for Markov operators like $G(\mathbf{M}g) \leq G(g)$ or $G^*(\mathbf{M}^*p) \leq G^*(p)$ Lyapunov functions follow, immediately: If $\mathbf{T}(t)$ is a semigroup of Markov operators, we set with $t_2 > t_1$ $g(t_1) = \mathbf{T}(t_1)g_0$, $g(t_2) = \mathbf{T}(t_2)g_0$, $p(t_1) = \mathbf{T}^*(t_1)p_0$, $p(t_2) = \mathbf{T}^*(t_2)p_0$. Setting $\mathbf{M} = \mathbf{T}(t_2-t_1)$, from the semigroup property we conclude $g(t_2) = \mathbf{T}(t_2)g_0 = \mathbf{T}(t_2-t_1)\mathbf{T}(t_1)g_0 = \mathbf{M}^*g(t_1)$ and analogously $p(t_2) = \mathbf{M}^*p(t_1)$. Thus, from the above inequalities follows $G(g(t_2)) \leq G(g(t_1))$ or $G^*(p(t_2)) \leq G^*(p(t_1))$.
Using this abstract result, from the inequalities (8), (10) and (11) the following Lyapunov functions for evolution equations can be derived (F is an arbitrary convex function).
Let $g(t)$ be the solution of some equation $\dot{g}(t) = \mathbf{A}g(t)$ with initial data $g(0) = g_0$ and μ is some solution to $\mathbf{A}^*\mu = 0$, then

$$\langle F(g(t_2)), \mu\rangle \leq \langle F(g(t_1)), \mu\rangle, \ t_2 > t_1 \tag{20}$$

Let $p(t)$ be the solution of some equation $\dot{p}(t) \stackrel{*}{=} \mathbf{A}^*p(t)$ with $p(0) = p_0$ and $q(t)$ be another solution with an other initial data $q(0) = q_0$, then for the functional (9)

$$H[p(t_2), q(t_2)] \leq H[p(t_1), q(t_1)], \ t_2 > t_1 \tag{21}$$

holds. Analogously, we can conclude from (11)

$$H[p(t_2), \mu] \leq H[p(t_1), \mu], \ t_2 > t_1 \tag{22}$$

for the stationary measure μ.
In degenerate situations, e.g., if the coefficients of equation (29) vanish in some part of the state space, the stationary measure μ is concentrated in this part. Then, the inequalities (20) and (22) carry information only about this part of the state space. In this case it is practical to use inequality (21) in the following way: Set $q(t) = p(t + t_0)$, i.e., $q(t)$ is the same solution as $p(t)$ shifted in time. Then, inequality

$$H[p(t_2), p(t_0 + t_2)] \leq H[p(t_1), p(t_0 + t_1)], \ t_2 > t_1$$

holds and make sense even for degenerate situations before the trajectory $p(t)$ reaches the degenerate part of the state space.
Note, that all inequalities are rigorous results in contrast to the inequality $\frac{d}{dt}H(t) \leq 0$ often derived in a formal way.

3.5 Examples for evolution problems

The natural next step is to describe the set of all physical sensible evolution equations for a given state space $\mathcal{Z}$. Having a necessary and sufficient criterion for generators generating trajectories of statistical states – the positive maximum principle (16) – we have to describe all operators with this property. It is easy to check this criterion for a given operator, but unfortunately, the description of all such operators is a very difficult problem and even for simple state spaces $\mathcal{Z}$ unsolved, yet. Supposably, the general way to do this is via the definition of the generator:

1) Give $\mathcal{Z}$ and its topology.
2) Describe the structure of a Markov operator in $\mathcal{C}(\mathcal{Z})$, i.e., the specific expression of P in (5).
3) Describe the analytic properties of semigroups of Markov operators $\mathbf{T}(t)$.
4) Calculate the limit $\mathbf{A}g = \lim_{t\to 0} \frac{1}{t}(\mathbf{T}(t) - \mathbf{I})g$.

In the following we will describe the result of this procedure for three important cases.

3.5.1 Dynamical systems and their Liouville equations

The first case is the one of deterministic trajectories, i.e. dynamical systems.
For $t \geq 0$ let $\varphi_t : \mathcal{Z} \to \mathcal{Z}$ be a time-family of continuous maps of the state space into itself with the properties $\varphi_{t_1+t_2} = \varphi_{t_2} \circ \varphi_{t_1}$ (semiflow property) and $\varphi_0 = id$ (identity). Then, if certain regularity properties are fulfilled the trajectory $z_t = \varphi_t(z_0)$ is the solution of an ordinary differential equation, say $\dot{z}_t = a(z_t)$ with initial data z_0.
The image of the semiflow φ_t in $\mathcal{Z}$ via the canonical embedding is a semigroup of a deterministic Markov operator.

$$\varphi_t \iff \mathbf{T}(t) = \mathbf{M}_{\varphi_t}$$

The semigroup property easily follows from the semiflow property. For arbitrary $g \in \mathcal{C}$ it holds

$$\begin{aligned} \mathbf{T}(t_1+t_2)g &= \mathbf{M}_{\varphi_{t_1+t_2}}g = \mathbf{M}_{\varphi_{t_2}\circ\varphi_{t_1}}g = g \circ (\varphi_{t_2} \circ \varphi_{t_1}) = (g \circ \varphi_{t_2}) \circ \varphi_{t_1} = \mathbf{M}_{\varphi_{t_1}}\mathbf{M}_{\varphi_{t_2}}g = \\ &= \mathbf{T}(t_1)\mathbf{T}(t_2)g \end{aligned}$$

Typical examples for semiflows are shifts and the corresponding generators are differential operators of first order. As usual, the definition of differential operators requires a metric structure in $\mathcal{Z}$, something that is not given in a general topological space. Surprisingly, there is a complete algebraic description of generators of semigroups of deterministic Markov operators. Such a generator is an abstract derivation:
A generator $\mathbf{A}$ is a generator of a semigroup of deterministic Markov operator if and only if the domain $D(\mathbf{A})$ is a sub-algebra of $\mathcal{C}(\mathcal{Z})$, $\mathbf{A}\mathbb{1} = 0$ and

$$\mathbf{A}(f \cdot g) = f \cdot \mathbf{A}g + g \cdot \mathbf{A}f$$

A derivation is always a unbounded operator (A.4.4.3). This is one reason why in finite dimensional spaces there are no semigroups of deterministic Markov operators.
If, for example, $\mathcal{Z} \subset \mathbb{R}^m$ is a compact domain and

$$\begin{aligned} \dot{z}_1 &= a_1(z_1, ..., z_m) \\ \cdots &\;.\; \quad \cdots \\ \dot{z}_m &= a_m(z_1, ..., z_m) \end{aligned} \tag{23}$$

is the dynamical system, i.e., the system of differential equations corresponding to the semiflow φ_t, then the corresponding equation to $g(z,t) = (\mathbf{T}(t)g_0)(z)$ is

$$\frac{\partial}{\partial t} g(z,t) = \sum_{i=1}^{m} a_i(z) \frac{\partial g(z,t)}{\partial z_i} \tag{24}$$

The right hand side of this equation is the general form of an abstract derivation in a domain of the Euclidean space.
The equation, corresponding to $p(t,B) = (\mathbf{T}^*(t)p_0)(B)$ is (assuming that $p(t,B)$ is regular enough to have a density $W(z,t)$ with respect to the Lebesgue measure)

$$\frac{\partial}{\partial t}W(z,t) = -\sum_{i=1}^{m}\frac{\partial}{\partial z_i}\big(a_i(z)W(z,t)\big) . \tag{25}$$

– the Liouville equation of the dynamical system (23). Conversely, (23) is the equation of the characteristics of (24) or (25).
For any two solutions W_1 and W_2 to equation (25) any functional H in the form (13)

$$H[W_1(t),W_2(t)] = \int_Z W_2(z,t)F^*\left(\frac{W_1(z,t)}{W_2(z,t)}\right)\,dz = H[W_1(0),W_2(0)]$$

is constant in time as expected since the right hand side of (24) is an abstract derivation.
Actually, this statement with $W_2(z,t)$ being the density of the stationary measure in phase space, is a general form of Liouville's famous theorem on phase space volume conservation.

3.5.2 The finite state space

In this case a complete description of Markov generators is possible.
A Markov generator is a matrix and has to have the form

$$\mathbf{A} = \begin{pmatrix} -d_1 & a_{12} & \cdots & a_{1n} \\ a_{21} & -d_2 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & -d_n \end{pmatrix}$$

where $a_{ij} \geq 0$ and $d_i = a_{i1} + ... + a_{i,i-1} + a_{i,i+1} + ... + a_{in}$ is the row sum. The corresponding evolution equations in $\mathcal{C}$ and $\mathcal{C}^*$ are

$$\dot{g}_i = (\mathbf{A}g)_i = \sum_{j\neq i}^{n} a_{ij}(g_j - g_i) \tag{26}$$

$$\dot{p}_i = (\mathbf{A}^*p)_i = \sum_{j\neq i}^{n}(a_{ji}p_j - a_{ij}p_i) \tag{27}$$

and can be understood in the strong sense. The functional H defined by (12) reads as

$$H[p(t),q(t)] = \sum_{i=1}^{n} q_i(t)F^*\left(\frac{p_i(t)}{q_i(t)}\right)$$

The monotonicity in time follows from

$$\frac{d}{dt}H(t) = -\sum_{i,j=1}^{n} a_{ji}q_j(t)\left[F^*\left(\frac{p_j}{q_j}\right) - F^*\left(\frac{p_i}{q_i}\right) - \left(\frac{p_j}{q_j} - \frac{p_i}{q_i}\right)F^{*\prime}\left(\frac{p_i}{q_i}\right)\right] \leq 0 .$$

due to the convexity of F^*.

3.5.3 The domain in $\mathbb{R}^m$

For the case of a domain in $\mathbb{R}^m$ the structure of a Markov generator can be found in [6, 13]. Moreover, in [13] a general H-theorem on a formal level is proved.
In this case, the structure of the generator of Markov semigroups can be described for inner points of $\mathcal{Z}$

$$(\mathbf{A}g)(z) = \sum_{i,j=1}^{m} b_{ij}(z)\frac{\partial^2 g}{\partial z_i \partial z_j} + \sum_{i=1}^{m} a_i(z)\frac{\partial g}{\partial z_i} + \fint_{\mathcal{Z}} \big(g(z') - g(z)\big)Q(z, dz') \tag{28}$$

where $\big(b_{ij}(z)\big)_{i,j=1}^{m}$ in a positive (in the sense of bilinear forms in Hilbert spaces) semidefinite matrix and $Q(z,B) \geq 0$ is a nonnegative function defined on all points $z \in \mathcal{Z}$ and Borel sets $B \subset \mathcal{Z}$. $Q(z,B)$ can be unbounded for $z \in B$. In this case, the integral operator has to be understood as a principal value integral.
Note, that expression (28) is a formal one and valid only for inner points of $\mathcal{Z}$. Neither the structure of the operator at boundary points, nor the regularity properties of the coefficients a_i, b_{ij}, Q are known for general situations.
An equation for probability densities cannot be derived in a strong sense, in general. Assuming, the probability measure $p(t)$ has a density $W(t)$ with respect to the Lebesgue measure, and assuming $Q(z, dz') = Q(z, z')dz'$, the evolution equation for $W(t)$ is

$$\begin{aligned}\frac{\partial}{\partial t}W(z,t) &= -\sum_{i=1}^{m}\frac{\partial}{\partial z_i}\big(a_i(z)W(z,t)\big) + \sum_{i,j=1}^{m}\frac{\partial^2}{\partial z_i \partial z_j}\big(b_{ij}(z)W(z,t)\big) + \\ &+ \fint_{\mathcal{Z}}\big(Q(z',z)W(z',t) - Q(z,z')W(z,t)\big)dz' \; .\end{aligned} \tag{29}$$

Normally, this is the form in which the Chapman-Kolmogorov forward equation is used (see [6]).
The functional H for two densities W_1 and W_2 of two solutions in the form (13) reads as

$$H[W_1(t), W_2(t)] = \int_{\mathcal{Z}} W_2(z,t)F^*\left(\frac{W_1(z,t)}{W_2(z,t)}\right) dz$$

the monotonicity follows from

$$\begin{aligned}\frac{d}{dt}H(t) &= -\int_{\mathcal{Z}}\sum_{i,j=1}^{n} F^{*\prime\prime}\left(\frac{W_1(z,t)}{W_2(z,t)}\right)\left(\frac{W_2}{W_1}\frac{\partial}{\partial z_i}\frac{W_1}{W_2}\right)\left(\frac{W_2}{W_1}\frac{\partial}{\partial z_j}\frac{W_1}{W_2}\right)b_{ij}(z)W_2(z,t)dz - \\ &- \int_{\mathcal{Z}}\int_{\mathcal{Z}}\Big(F^*(\theta) - F^*(\theta') - F^{*\prime}(\theta')(\theta - \theta')\Big)W_2(z)Q(z,z')dz\,dz' \leq 0 \; ,\end{aligned} \tag{30}$$

with $\theta = \frac{W_1(z)}{W_2(z)}$, $\theta' = \frac{W_1(z')}{W_2(z')}$.
Note, that the coefficients a_i correspond to the Liouville equation and therefore disappear after derivating with respect to time.
Note, that inequality (30) is a formal one. A rigorous derivation of (30) requires regularity results like: a prove of the existence of a density solving (29); a prove that $W_2(z,t) > 0$, to compose $\frac{W_1}{W_2}$; differentiability of $H[W_1(t), W_2(t)]$ with respect to time. Moreover (30) requires smooth functions F^*. In general, the rigorous inequality (22) holds as soon as $\mathbf{A}$ satisfy (16).

4 Consequences and discussion

We will concentrate on consequences of the mathematical setting especially on the duality of states as physical values and observables as test values and the irreversibility inequalities (7), (8), (10), (11). We will distinguish the consequences in physical ones (irreversibility and derivation of macroscopic and empirical equations) and mathematical ones (tools for analytical investigations of the equations and general approximation methods).

4.1 Physical consequences

4.1.1 H-theorems and the second law of thermodynamics

The proof of inequality (10) (A.3.2.3) shows that irreversibility occurs as soon as the force, acting on the system contains a random part. This random part comes into the play for several reasons, on the one hand in a direct way involving noise and on the other in an indirect way. One example is the use of a mollifier function to smear the trajectories of particles. Another example is the assumption of homogeneous distributed scattering angles in the derivation of the Boltzmann equation. Boltzmann's "Stoßzahlansatz" is a further implicit randomization of the problem. This is well discussed in [7] using a variant of Ehrenfest's urn model. Here is assumed that the number of interacting particles is proportional to the number of all particles, an assumption about homogeneity of the underlying space. Both assumptions are the intrinsic reason of the irreversibility in the Boltzmann equation.
Inequality (7) as an origin of the inequalities (8), (10) and (11) show that irreversibility is connected with a leak of information. Convex combinations do not commute with state changes. It turns out that the second law of thermodynamics is a mathematical consequences of the properties of probabilities (positivity and normalization). Provocatively, one can say: May be somewhere exists a world without a second law of thermodynamics, but we cannot describe such a world with probabilities.

4.1.2 Microscopic equations and entropy

As seen from the inequalities (7), (8), (10), (11) any convex function provides such inequalities and there are valid for any state changes. Thus, there is no natural procedure to pick out one of them like $F(x) = x \log x$ or $F(x) = \log x$ to define an entropy.
It is generally accepted that entropy is not a value of interest when investigating a deterministic system described for example by a Hamilton system. We believe, the same is true for any non-deterministic microscopic equation like (29) or in the general form (19). The difference between a Hamilton system and this equations is only the nature of the acting force, deterministic or random.
Beside (2), any of the functions

$$H(t) = \Phi[u] = \int F\big(u(x,t)e^{V(x)}\big)e^{-V(x)}dx$$

is a Lyapunov function for the FPE (1) with any convex F. It is not clear, why we have to single out (2) as the entropy for a Brownian particle.
Completely different is the situation if we derive a macroscopic equation from a microscopic one or postulate an empirical equation. In this case the entropy comes out during the derivation or has to be postulated in advance.

4.1.3 The derivation of the BBGKY-hierarchy and the Boltzmann equation

The derivation of the BBGKY-hierarchy and the Boltzmann equation as a consequence of the BBGKY is a method to simplify a problem for n identical classical particles. Starting with an equation for a n-particle density $W(z_1, ..., z_n)$ and using symmetry properties we obtain an equation for a function $f(z)$, depending only on one parameter.
This can be done, assuming permutation symmetry for $W(z_1, ..., z_n)$. We believe, this assumption is not correct and yields a false understanding of $f(z)$ as a density function. It is easy to see, that the assumption $W(z_1, z_2) = W(z_2, z_1)$ (in the case of two particles) is never fulfilled for a deterministic state: Assume for example, particle 1 is in state θ_1 and particle 2 is in state θ_2 with certainty. Thus, $W_1(z_1, z_2) = \delta(z_1 - \theta_1)\delta(z_2 - \theta_2)$. This function has no symmetry and is not the same as $W_2(z_1, z_2) = \delta(z_2 - \theta_1)\delta(z_1 - \theta_2)$. Likewise, the symmetrical density $(W_1(z_1, z_2) + W_2(z_1, z_2))/2$ describes an other situation, namely a case when the states of the particles are not determined. The symmetrization of $W(z_1, ..., z_n)$ is an approach to improve the actually non physical procedure of numbering un-distinguishable particles. The real meaning of identical and un-distinguishable particles is that there is no observables that can distinguish the particles, i.e., we have to allow only observables with permutation symmetry. Using such observables (we follow the derivation from [4]), from a n-particle density $W(z_1, ..., z_n)$ we obtain a function, depending on only one parameter $f(z)$ that is not a probability density but a moment – the average number of particles. This is still a many-particle function, although it depends on only one state parameter z.
The BBGKY-hierarchy has the same set of Lyapunov functions as the original n-particle equation. Passing from the first equation of the BBGKY-hierarchy to the Boltzmann equation using the factorization $f(z_1, z_2) = f(z_1)f(z_2)$ (i.e., assuming propagation of chaos for infinite many particles) from all Lyapunov functions this procedure selects one or some special ones, satisfying some algebraic properties connected with the ansatz $f(z_1, z_2) = f(z_1)f(z_2)$.

4.1.4 Empirical equations

Equations like the FPE (1) are one-particle equations. In this sense, it is a microscopic (or first principle) equation. It describes the probability density of one particle. This density has to be understood as a sum of independent experiments and the equation is therefore linear. It is clear that the same density function describes the concentration of many non-interacting particles. In this sense, the FPE is a macroscopic many-particle equation. The difference of this equations is in their interpretation.
As a one-particle equation it predicts the probability of finding the particle somewhere at time t. If this time is reached, the particle is in a determined point with certainty. The prediction and the reality do not coincide.
As a many-particle equation it predicts the concentration, i.e., the whereabout of all particles at time t. If this time is reached, the prediction and the reality coincide.
If we want to describe a n-particle system with interaction, the correct but usually impracticable way is the description of these n-particles as one system and a following simplification. This is the silver bullet, feasible in academic situations.
The other, empirical, way for the derivation of an equation for a n-particle problem is (see [11] and the references there) to start with a suitable one-particle equation, to interpret it as a model for n non-interacting particles and to try to consider the interactions afterwards. We illustrate this procedure for the diffusion equation with constant diffusion coefficient D.

Step 1: We assume that the equation $\dot{f} = D\Delta f$ describe the evolution of the probability density of one particle of interest.
Step 2: We assume that the equation $\dot{u} = D\Delta u$ describe the evolution of the concentration of many of these particles without interaction.
Step 3: We want to include interactions in an empirical way, introducing, for example, a dependence of the diffusion coefficient on some quantities like $D \to \mathbf{D}(x, u, \nabla u, ...)$.
Remark: There are many ways to do this. To do the right choice, we have to understand equation $\dot{u} = D\Delta u$ as a many particle equation, i.e., to point out what is the driving force for the particles to diffuse. We can write the equation in the form

$$\dot{u} = D\Delta u = \nabla^* \bullet (D\nabla u) \tag{31}$$

and understand the gradient of the concentration ∇u as a driving force (for the question, where the diffusion coefficient has to be placed see sub-subsection 4.2.3).
The introduction of a driving force implies that the motion of the particles is a cooperative effect. But by looking at (31) as a microscopic one-particle-equation, no cooperative effect can be found as there is only one particle. From the point of view of a one-particle-equation this makes the introduction of a driving force a purely arbitrary step. Indeed, for any smooth convex function F any decomposition

$$\dot{u} = D\Delta u = \nabla^* \bullet \Big(D\big[\partial^2 F(u)\big]^{-1} \nabla \partial F(u)\Big) \tag{32}$$

with $\big[\partial^2 F(u)\big]^{-1}$ being the inverse of the Hessian $\partial^2 F(u)$, defines a driving force $\nabla\lambda = \nabla\partial F(u)$, the gradient of some potential. The connection between the driving force and the concentration, namely $u = (\partial F)^{-1}(\lambda)$ describes the statistics of the considered ensemble of particles. λ is called chemical potential and the driving force is its gradient. The statistic function $(\partial F)^{-1}$ is the inverse of the derivative of the free energy.
Step 4: Depended on the nature of the particles define a driving force $\nabla\lambda = \nabla\partial F(u)$ as the gradient of the derivative of some free energy $F(u)$ or the inverse of some statistics $u = (\partial F)^{-1}(\lambda)$. Now, we write down equation (32).
Step 5: Generalize this equation as you like introducing a general diffusion coefficient $\mathbf{D}$

$$\begin{aligned} \dot{u} &= \nabla \cdot \mathbf{D}\nabla\lambda \\ u &= (\partial F)^{-1}(\lambda) \end{aligned}$$

This system possesses only one Lyapunov function, the free energy $H(t) = \int F\big(u(x,t)\big)dx$ due to

$$\frac{d}{dt}H(t) = \frac{d}{dt}\int F(u)dx = \int \partial F(u)\dot{u}dx = -\int \Big(\mathbf{D}\big[\partial^2 F(u)\big]^{-1}\lambda \bullet \lambda\Big)\, dx \leq 0$$

if the product of two matrices $\mathbf{D}\big[\partial^2 F(u)\big]^{-1}$ is positive definite. Note, that equation (31) as a one-particle-equation possesses the whole family of Lyapunov functions.
This method of derivation is very successful especially in the case of drift-diffusion equations for classical charged particles, where the driving force is $\nabla(\lambda + \psi)$ with the chemical potential λ and the electro-static potential ψ (see, e.g., [11]).
In the special case of classical particles the free energy is $F(u) = u(\log u - 1)$, the connection between the chemical potential and the concentration is $u = (\partial F)^{-1}(\lambda) = e^{\lambda}$, the well known Boltzmann distribution and the corresponding evolution equation is $\dot{u} = \nabla \cdot \big(\mathbf{D}u\nabla \log u\big)$.

The free energy (sometimes called entropy) $F(u) = u(\log u - 1)$ is the limit of the simplest case of n completely distinguishable particles. Defining the free energy as the logarithm of the number of possibilities to arrange the particles, by Stirling's formula we have

$$F(u) \approx \lim_{n\to\infty} \log n! \approx \log \left(\frac{n}{e}\right)^n \approx u(\log u - 1)$$

The definition of the entropy as $F(u) = u(\log u - 1)$ seems to come from this property and is the correct one only in the special case of completely distinguishable particles.

4.2 Mathematical consequences

4.2.1 The natural function spaces

The construction – initial observables $\longrightarrow$ topology $\longrightarrow$ observables $\longrightarrow$ probability measures – defines the function spaces $\mathcal{C}$ and $\mathcal{C}^*$ for the investigation of the problem in a natural way. Once the state space $\mathcal{Z}$ and its topology are defined (this is the physical part of the investigation) the following procedure is preset. Regardless of geometric difficulties of the state space (inhomogeneities, degeneracy, varying dimensions), continuous functions and probability measures on $\mathcal{Z}$ are well defined.

Considering evolution problems, if the operator $\mathbf{A}$ is a Markov generator, equations (17) and (19) are solvable for any time $t \geq 0$ (including 0). Equation (17) is solvable in a classical sense. Even for non homogeneous problems or other exotic situations, the introduction of a weak sense of the solution is not required. Equation (19) is solvable in $\mathcal{S}^*$, globally in time . Its solution is not worse then a measure. More irregular objects like derivatives of point measures, other distributions or Young measures are not required. However, the solution is not better than a measure, in general. It cannot be expected that the measure has a density with respect to any fixed measure.

The solvability of (19) for any time $t \geq 0$ includes the solvability for any initial point measure δ_z. Thus, Green's function (a measure, in general) is well defined by $G(t, z, B) = \mathbf{T}^*(t)\delta_z(B)$ and the widely used formulae

$$\begin{aligned} g(t,z) &= \big(\mathbf{T}(t)g_0\big)(z) = \int_z g_0(z')G(t,z,dz') \\ p(t,B) &= \big(\mathbf{T}^*(t)p_0\big)(B) = \int_z G(t,z',B)p(dz') \end{aligned}$$

are correct in a rigorous way.

4.2.2 Equations for densities. L_2-theory. Detailed balance

As mentioned, an evolution problem is globally solvable in $\mathcal{S}^*$. Nevertheless, it may be useful, to consider densities in place of measures. If such a density is considered as an element of $L_r(\mu)$, for example, we have the possibility to choose an optimal measure μ for the problem. This choice is not without consequences, because it impose restrictions to the state space. Jensen's inequalities in the forms (14) and (15) give a hint for the right measure. If p_0 has a density with respect to the stationary measure μ inequality (15) that the $p(t)$ has a density with respect μ for all times. Moreover, the norm cannot increase, i.e., $\mathbf{T}(t)$ and $\mathbf{T}^*(t)$ are contraction semigroups in $L_r(\mu)$ and $L_{r'}(\mu)$, resp. If $L_r(\mu)$ was chosen with another measure than the stationary one, blow up effects can occur.

To find out an equation that satisfies $h(t)$, we consider the case $r = 2$, i.e., the Hilbert space $L_2(\mu)$. We define $L_2(\mu)$ as the closure of $\mathcal{C}$ in the norm

$$\|f\|^2_{L_2(\mu)} = \int_{\mathcal{Z}} |f(z)|^2 \mu(dz)$$

The scalar product $(\cdot,\cdot)_\mu$ in $L_2(\mu)$ on the dense subset $\mathcal{C}$ is defined as

$$(f,g)_\mu = \langle g \cdot f, \mu\rangle,\ \ g, f \in \mathcal{C}$$

We define an operator $\mathbf{B}$ in $L_2(\mu)$ as the extension $\mathbf{B} = \mathbf{A}\big|_{\mathcal{C}}$ of $\mathbf{A}$:

$$(\mathbf{B}f, g)_\mu = \langle g \cdot \mathbf{A}f, \mu\rangle,\ \ f, g \in D(\mathbf{A})$$

Because of the inequality $2g \cdot \mathbf{A}g \leq \mathbf{A}g^2$ (A.5.3.3) we have

$$\frac{d}{dt}\|g\|^2_{L_2(\mu)} = \frac{d}{dt}\langle g^2, \mu\rangle = 2\langle g \cdot \dot{g}, \mu\rangle = 2\langle g \cdot \mathbf{A}g, \mu\rangle \leq 2\langle \mathbf{A}g^2, \mu\rangle = 2\langle g^2, \mathbf{A}^*\mu\rangle = 0$$

Thus, $\mathbf{B}$ is dissipative. The equation for the density $h(t)$ follows from

$$\begin{aligned}\frac{d}{dt}(g, h(t))_\mu &= \frac{d}{dt}\langle g \cdot h(t), \mu\rangle = \frac{d}{dt}\langle g, p(t)\rangle = \langle \mathbf{A}g, p(t)\rangle = \langle \mathbf{A}g \cdot h(t), \mu\rangle = (\mathbf{B}g, h(t))_\mu = \\ &= (g, \mathbf{B}^* h(t))_\mu,\ \ g \in D(\mathbf{A})\end{aligned}$$

Thus, the evolution equation for $h(t)$ is

$$\dot{h}(t) = \mathbf{B}^* h(t)$$

and valid only in a weak sense (since L_2 is reflexive, there is no weak* topology) because of the definition of $\mathbf{B}$ as a closure of a bilinear form. Moreover, only initial data with densities with respect to μ are allowed.
For matrices, $\mathbf{A}^*$ is the transposed matrix to $\mathbf{A}$, but $\mathbf{B}^*$ is not the transposed of $\mathbf{B} = \mathbf{A}$, but the "transposed of $\mathbf{B}$ with respect to μ." In the infinite dimensional case there is no notion of transposition. This is obvious as $\mathbf{A}$ and $\mathbf{A}^*$ are acting in different spaces. For $\mathbf{B}$ and $\mathbf{B}^*$ such a connection holds even in the infinite dimensional case. We can say that $\mathbf{B}^*$ is the "transposed" operator of $\mathbf{B} = \mathbf{A}$, defined via the scalar product $(\cdot,\cdot)_\mu$

$$\langle g \cdot \mathbf{A}f, \mu\rangle = (\mathbf{B}f, g)_\mu = (g, \mathbf{B}^* f)_\mu \tag{33}$$

Thus, the adjoint of $\mathbf{B}$ is defined via $\mathbf{A}$ and not via $\mathbf{A}^*$. $\mathbf{B}^*$ is also the extension of some Markov generator and not of an adjoint.
From (33) an answer to the question – Whether a Markov generator is self-adjoint? – can be given. It is clear that an equation like $\mathbf{A} = \mathbf{A}^*$ makes no sense for operators acting in different spaces. Even for matrices such an equation makes no sense. Although both $\mathbf{A}$ and $\mathbf{A}^*$ act in $\mathbb{R}_n$, the norms in this spaces are different. This question of self-adjointness can be asked in $L_2(\mu)$. An operator $\mathbf{B}$ is self-adjoint in $L_2(\mu)$ if $\mathbf{B} = \mathbf{B}^*$ holds or, equivalently, in $\mathcal{C}$

$$\langle g \cdot \mathbf{A}f, \mu\rangle = \langle f \cdot \mathbf{A}g, \mu\rangle,\ \ f, g \in D(\mathbf{A}) \tag{34}$$

Assuming self-adjointness, setting $f = \mathbb{1}$ from (34) for all $g \in D(\mathbf{A})$ it follows

$$\langle g, \mathbf{A}^*\mu\rangle = \langle \mathbb{1} \cdot \mathbf{A}g, \mu\rangle = \langle g \cdot \mathbf{A}\mathbb{1}, \mu\rangle = 0$$

Since this holds for all $g \in D(\mathbf{A})$, we have $\mathbf{A}^*\mu = 0$ and therefore μ is the stationary measure. Thus, if the extension of a Markov generator in a Hilbert space defined on some measure is self-adjoint, this measure has to be the stationary measure of its adjoint.
Defining a multiplication operator $\mathbf{Q}_\mu : \mathcal{C} \longrightarrow \mathcal{C}^*$ via $\mathbf{Q}_\mu g = g\mu$ (A.3.1.3) , equality (34) is equivalent to the operator equality

$$\mathbf{A}^*\mathbf{Q}_\mu = \mathbf{Q}_\mu \mathbf{A} \tag{35}$$

In a finite state space, if $\mathbf{A}$ and $\mathbf{A}^*$ are matrices, this is precisely the definition of detailed balance. Thus, a natural extension of the notion of detailed balance is (35) together with $\mathbf{A}^*\mu = 0$.
Or, in other words: A Markov generator $\mathbf{A}$ is a detailed balance operator, if its extension in the L_2-Hilbert space w.r.t. the stationary measure of its adjoint is self-adjoint. As the extension $\mathbf{B}$ of a Markov operator $\mathbf{A}$ into $L_2(\mu)$ is always dissipative, $-\mathbf{B}$ is a positive semi-definite operator in the sense of bilinear forms. Thus, cone-positivity imply form-positivity. The converse is not true. This is easy to see, considering, e.g., forth order differential operators.
The value $\|g\|_{\mathbf{B}} = \sqrt{-(\mathbf{B}g, g)_\mu}$ is the so-called energy norm. If $\mathbf{A}$ is the Laplacian, this norm is the starting point for the definition of the Sobolev spaces.
The notion of detailed balance for matrices is important for the investigation of discrete Markov processes. Being positive definite matrices in the right space, the analysis of such problems is much simpler than the analysis of general ones. The extension of this notion to general Markov generators makes it possible to use the methods form detailed balance theory as soon as we have established (34) (see, e.g., [3]). Indeed, this is a wide class of operators, as we do not consider self-adjoint operators in a given Hilbert space but chose the best Hilbert space for a given operator. For example, any pure differential operator (with $Q \equiv 0$) of type (28) in one dimension or any Fokker-Planck operator with suitable boundary conditions is self-adjoint in the right Hilbert space, regardless of the first order derivative terms. This can be easily established (for the Fokker-Planck operator this is done in sub-subsection 4.2.3), using the explicite expression for the stationary measure.

4.2.3 The optimal setting of the Fokker-Planck equation

We rewrite the Fokker-Planck equation (1) as

$$\dot{u} = \Delta u + \nabla^* \bullet (u \nabla V) \tag{36}$$

and ask for the best setting for this equation. Again, ∇^* and ∇ are the div and the grad operators, resp. and $\bullet$ denotes the scalar product in the Euclidean space. We assume $x \in \mathbb{R}^3$ and suitable boundary conditions, e.g., homogeneous Neumann conditions. For smooth situations the solution of this equation is smooth enough and can be considered in any function space, e.g., $L_2(dx)$ (like that we denote the L_2 space with respect to the Lebesgue measure). It is clear that the operator on the right hand side of equation (36) is not symmetric in the $L_2(dx)$ due to the first order term.
This is not astonishing. Taking $L_2(dx)$ we choose implicitly a measure for the problem – the homogeneous measure in $\mathbb{R}^3$. This means, we expect that the particle feel in every point the same space. But the particle moves in a potential $V(x)$. Indeed, this potential signifies a inhomogeneity of the space. Thus, the chose of $L_2(dx)$ seems not to be the optimal space for the problem.

Moreover, for non smooth potentials $V(x)$ is not clear whether this equation has a global solution in $L_2(dx)$ at all.
Using the presented formalism, at first, we have to wright down the equation for the observables in $\mathcal{C}(\mathcal{Z})$ (we assume a suitable compactification for $\mathbb{R}^3$, e.g., the 3-dimensional ball). This is the so-called Fokker-Planck backwards equation

$$\dot{g} = \mathbf{A}g = \Delta g - \nabla g \cdot \nabla V \tag{37}$$

This equation has a classical solution in $\mathcal{C}$ even for non smooth potentials. In this case we have to improve the domain of $\mathbf{A}$ in such a way that ∇g vanishes where ∇V is not continuous.
The adjoint operator $\mathbf{A}^*$ is the right hand side of (36) if the solution of the corresponding equation (19) has a smooth density.
In general, we are done. The adjoint of the solution operator $\mathbf{T}(t)$ for this equation provides the solution in $\mathcal{S}^*$, i.e., $p(t) = \mathbf{T}^*(t)p_0$.
Assuming that $p(t)$ has a density, (36) is not the best equation for the determination of this density. We have to find the stationary solution μ as a solution to the equation

$$0 = \langle \mathbf{A}g, \mu \rangle = \int_{\mathbb{R}^3} \big(\Delta g - \nabla g \cdot \nabla V\big)\mu(dx) \tag{38}$$

It is easy to check that $\mu(dx) = e^{-V(x)}dx$ is the solution to (38) (we omit the normalization constant for simplicity).
Now, we are looking for an equation for the density $h(t)$ with $p(t) = h(t)\mu$. For this purpose, we have to extend $\mathbf{A}$ to an operator $\mathbf{B}$ in the Hilbert space $L_2(\mu)$ to find $\mathbf{B}^*$ in $L_2(\mu)$ so that we can write down $\dot{h} = \mathbf{B}^*h$. Thus, we have to calculate the bilinear form $(f, \mathbf{B}g)_\mu$.

$$\begin{aligned}(f, \mathbf{B}g)_\mu &= \langle f \cdot \mathbf{A}g, \mu\rangle = \int_{\mathbb{R}^3} f\big(\Delta g - \nabla g \cdot \nabla V\big)\mu(dx) = \\ &= \int_{\mathbb{R}^3} \big(f\Delta g - f(\nabla g \cdot \nabla V)\big)e^{-V(x)}dx = -\int_{\mathbb{R}^3} \nabla f \cdot \nabla g\, e^{-V(x)}dx\end{aligned}$$

(here we integrated by parts

$$\int_{\mathbb{R}^3} f\Delta g e^{-V(x)}dx = -\int_{\mathbb{R}^3} \nabla g \cdot \nabla(f e^{-V(x)})dx = -\int_{\mathbb{R}^3} \big((\nabla g \cdot \nabla f)e^{-V(x)} - f(\nabla g \cdot \nabla V)e^{-V(x)}\big)dx$$

using the Neumann boundary conditions). Thus, the operator $\mathbf{B}$ is symmetric (and therefore self-adjoint because it is a generator (A.4.5.2)) and the right equation for $h(t)$ is

$$\dot{h} = \mathbf{B}^*h = \mathbf{B}h = \mathbf{A}h = \Delta h - \nabla h \cdot \nabla V \tag{39}$$

Taking the solution to this equation we have $p(t) = h(t)\mu$.
The connection with equation (36) is $u(x, t) = h(x, t)e^{-V(x)}$ and can be established by simple calculations. The discrepancy between the equations (36) and (39) is that (39) is available in more general situations because $h(t)$ is a smoother function, tending to $\mathbb{1}$ for large times, whereas $u(x, t)$ tends to $e^{-V(x)}$ what may not be a function at all.
In general, the Fokker-Planck equation contains a diffusion coefficient $D(x)$. There is a long discussion, where this diffusion coefficient has to be placed with respect to the spatial derivatives. In one dimension the question is, whether $\dot{u} = D(x)\partial_{xx} + ...$ or $\dot{u} = \partial_{xx}D(x) + ...$ or $\dot{u} = \partial_x D(x)\partial_x + ...$ is correct. The answer is, that all of them are right depending on the setting. The first is the right form in $\mathcal{C}$ and in a classical sense, the second is the right form in $\mathcal{C}^*$ in a weak* sense and the third is the right form in a weak sense in $L_2(\mu)$. These equations describe the evolution of an observation, of the probability measure and of the density of this probability measure with respect to the stationary measure, resp.

4.2.4 The integration by parts rule

The extension of the integration by parts rule to more general situations is a frequently investigated problem. Actually this is a problem of extending the notion of a derivative.
We propose the following extension – based on the algebraic definition of an abstract derivation. Given an abstract derivation $\mathbf{A}$, pairing with some $\mu \in \mathcal{S}^*$, we have

$$\langle f \cdot \mathbf{A}g, \mu\rangle + \langle g \cdot \mathbf{A}f, \mu\rangle = \langle \mathbf{A}(f \cdot g), \mu\rangle = \langle f \cdot g, \mathbf{A}^* \mu\rangle$$

If μ is the stationary measure for $\mathbf{A}^*$, we have $\mathbf{A}^* \mu = 0$. Hence,

$$\langle f \cdot \mathbf{A}g, \mu\rangle = -\langle g \cdot \mathbf{A}f, \mu\rangle$$

This is equivalent to

$$(f, \mathbf{B}g)_\mu = -(g, \mathbf{B}f)_\mu$$

and is the most general form of the integration by parts rule. This is a pure algebraic formula and does not require any metric structure in $\mathcal{Z}$ for the definition of a differential quotient.

4.2.5 Approximations

Now, we are looking for a natural approximation scheme for the weak* equation (18) for the probability measure $p(t)$ with initial date $p(0) = p_0$

$$\frac{d}{dt}\big\langle g, p(t)\big\rangle = \big\langle \mathbf{A}g, p(t)\big\rangle,\ g \in D(\mathbf{A})$$

Following the main procedure, we have to chose suitable observables, say linear independent $g_1, ..., g_n$ to extract information about $p(t)$. We assume that $g_i \in D(\mathbf{A})$ and $\mathcal{C}_n = \operatorname{span}(g_1, ..., g_n)$ and get the system of equations

$$\frac{d}{dt}\big\langle g_i, p(t)\big\rangle = \big\langle \mathbf{A}g_i, p(t)\big\rangle,\ i = 1, ..., n \tag{40}$$

This is a moment expansion for the generalized moments $m_i(t) = \big\langle g_i, p(t)\big\rangle$. This system can be solved if it is closed. This is the case if and only if $\mathbf{A}g_i \in \mathcal{C}_n$ what is not the case, in general. We can close the system replacing $\mathbf{A}$ by $\mathbf{A}_n = \mathbf{P}_n\mathbf{A}$, where $\mathbf{P}_n$ is a projector into $\mathcal{C}_n$. Such a projector has to be defined via a base in $\mathbf{C}^*$, the dual of $\mathbf{C}$. As we are interested in probability measures, we take a base $q_1, ..., q_n \in \mathcal{S}^*$. Then, the projector acts on $g \in \mathcal{C}$ as

$$\mathbf{P}_n g = \sum_{j=1}^{n} \langle g, q_j\rangle g_j$$

It is natural to assume that the projection is the identity in $\mathcal{C}_n$. This is $\mathbf{P}_n g_i = g_i$. Thus, the bases $(g_i)_{i=1}^n$ and $(q_j)_{j=1}^n$ are bi-orthonormal, i.e., they are connected by

$$\langle g_i, q_j\rangle = \delta_{ij} \tag{41}$$

With this definition of the projector we have

$$\mathbf{A}_n g_i = \mathbf{P}_n \mathbf{A} g_i = \sum_{j=1}^{n} \langle \mathbf{A}g_i, q_j\rangle g_j = \sum_{j=1}^{n} a_{ij} g_j$$

with $a_{ij} = \langle \mathbf{A} g_i, q_j \rangle$. From (40) we conclude

$$\frac{d}{dt}\big\langle g_i, p(t)\big\rangle = \sum_{j=1}^{n} a_{ij}\big\langle g_i, p(t)\big\rangle,\ i = 1, ..., n$$

and with the definition of the moments

$$\dot{m}_i(t) = \sum_{j=1}^{n} a_{ij} m_j(t),\ i = 1, ..., n \tag{42}$$

The desired approximate solution is $p(t) = \sum_{j=1}^{n} m_j(t) q_j$. One can make sure by acting with g_i on this equation and using (41). An approximation scheme with a projector generated by an bi-orthonormal base is called Galerkin-Petrov scheme.

Dealing with Markov generators, it is natural to demand the projector to conserve positivity. This yields $g_i \geq 0$. Moreover, $\mathbb{1}$ should be in any finite dimensional space $\mathcal{C}_n$. This yields $\mathbf{P}_n \mathbb{1} = \mathbb{1}$ and therefore $\sum_{i=1}^{n} g_i = \mathbb{1}$. Hence, $(g_i)_{i=1}^{n}$ is a decomposition of the unity.

From $\sum_{i=1}^{n} \langle \mathbf{A} g_i, q_j \rangle = \langle \mathbf{A} \sum_{i=1}^{n} g_i, q_j \rangle = \langle \mathbf{A}\mathbb{1}, q_j \rangle = 0$ for $j = 1, ..., n$, we conclude $a_{jj} = -\sum_{j \neq i=1}^{n} a_{ij}$. Thus, equation (42) is of type (27):

$$\dot{m}_i(t) = \sum_{i \neq j=1}^{n} \big(a_{ij} m_j(t) - a_{ji} m_i(t)\big),\ i = 1, ..., n \tag{43}$$

From $g_i \geq 0$, $\sum_{i=1}^{n} g_i = \mathbb{1}$ and $m_i(t) = \big\langle g_i, p(t)\big\rangle$ we have $m_i \geq 0$ and $\sum_{i=1}^{n} m_i = 1$. Thus, the moments can be understood as probability measures in some statistical state space $\mathcal{S}_n^*$. It lefts to analyze the properties of this space.

Because $(g_i)_{i=1}^{n}$ is a decomposition of the unity, it is naturally to define the following closed (A_i and B_i) and open (U_i) sets

$$\begin{aligned} A_i &= \{z | g_i(z) = 1\} = g_i^{-1}\big(\{1\}\big) \\ B_i &= \{z | g_i(z) = 0\} = g_i^{-1}\big(\{0\}\big) \\ U_i &= \{z | g_i(z) > 0\} = g_i^{-1}\big((0, \infty)\big) = \mathcal{Z} \setminus B_i \end{aligned}$$

From (41) and $q_j \in \mathcal{S}^*$ follows $q_j(A_j) = 1$ and $q_j(A_i) = 1$ for $i \neq j$. From the point of view of the observables g_j, states in the sets A_j are un-distinguishable. Thus, A_j can be understood as one state and q_j is the corresponding point measure. Finally, the approximation scheme can be understood as the following physical problem: We have n pots, located in the sets A_j that can be assumed to be points. The measures q_j are concentrated in this pots and therefore describe the probability for the system to be in a state in A_j. a_{ij} is the rate, carrying probability from A_j to A_i. If $\mathcal{Z}$ is a domain in the Euclidean space and the sets A_i are given points there, the described method is a variant of the finite volume method, where the solution is a linear combination of point measures. The non-zero rates a_{ij} describe the edges of a graph of the geometrical connection of the points A_i.

The sets $(U_j)_{j=1}^{n}$ establish an open cover of the compact set $\mathcal{Z}$. They describe the uncertainty for a state to belong to the closed sets A_i. If $z_i \in U_i \setminus U_j$, $z_j \in U_j \setminus U_i$ and $z_{ij} \in U_i \cap U_j$ than we have the situation similar to the three ball from sub-subsection 4.3.2: $z_i \sim z_{ij}$, $z_j \sim z_{ij}$ but $z_i \not\sim z_j$ where $\sim$ denotes the distinguishability of two states by observables.

The evolution equation (43) itself, can be solved explicitely for all time knowing the spectrum of $\mathbf{A}_n$. If $\lambda_1 = 0, -\lambda_2, ..., -\lambda_n$ are the eigenvalues of $\mathbf{A}_n$, the resolvent $\mathbf{R}_n(\lambda)$ of $\mathbf{A}_n$ can be written as

$$\mathbf{R}_n(\lambda) = (\lambda - \mathbf{A}_n)^{-1} = \mathbf{C}_1\frac{1}{\lambda} + \mathbf{C}_2\frac{1}{\lambda+\lambda_2} + ... + \mathbf{C}_n\frac{1}{\lambda+\lambda_n}$$

with matrices $\mathbf{C}_1$, ..., $\mathbf{C}_n$ not depending on λ. Then, the solution $\mathbf{T}(t)$ can be written in an explicit way for all times $t \geq 0$ as

$$\mathbf{T}(t) = \mathbf{C}_1 + \mathbf{C}_2 e^{-\lambda_2 t} + ... + \mathbf{C}_n e^{-\lambda_n t}$$

$\mathbf{C}_1^*$ is the projector of the stationary space (the linear hull of all stationary states).

4.3 Some Remarks

Here, we collect some further comments on the mathematical framework.

4.3.1 The compactness of the state space

We require compactness of the state space $\mathcal{Z}$. From a mathematical point of view this is a very important assumption. It causes the following statements – essential for the whole explanation

- A continuous function attains its maximum and minimum on $\mathcal{Z}$.
- $\mathcal{Z}$ is compact if and only if $\mathcal{C}(\mathcal{Z})$ is separable
- For operators with $\mathbf{M} \geq 0$ the properties $\mathbf{M}\mathbb{1} = \mathbb{1}$ and $\|\mathbf{M}\| \leq 1$ are equivalent.
- There exists a stationary probability measure $\mu \in \mathcal{S}^*(\mathcal{Z})$ with $\mathbf{A}^*\mu = 0$.

It may seem that the requirement of compactness of the state space constrains the class of considerable problems in a non acceptable way. But this is not the case. A topological space is compact if and only if the set of continuous real-valued functions is separable, i.e., contains a countable basis (A.1.2.6). Thus, using a non compact state space we have to motivate why we need a non-countable set of observables. Often, the topology is canonically given from the nature of the state space. For example, let us consider a diffusing particle with the position $x \in \mathbb{R}$ as states. Of course, we use the usual topology in $\mathbb{R}$ that makes $\mathbb{R}$ non-compact. Observables – continuous functions on $\mathbb{R}$ – distinguish any points in $\mathbb{R}$ in the same way. But this is not natural. It is more difficult to distinguish very far points than near ones. This means, an observable should differ decreasingly for far points, i.e., a observable should have limits for $x \to \pm\infty$. Taking as observables only the set of continuous functions with such limits we obtain a different topology in $\mathbb{R}$ which is compact, the so-called two point compactification of $\mathbb{R}$. This means actually, that we add two points to the state space, namely $+\infty$ and $-\infty$. The result is that the state space becomes larger but compact, the space of observables becomes smaller and the space of probability measures becomes larger: The point measures $\delta_{+\infty}$ and $\delta_{-\infty}$ are now contained in $\mathcal{S}^*$, because the limits of the observables for $x \longrightarrow \pm\infty$ are well defined real numbers and therefore functionals.
This solves a very dissatisfying paradox for the diffusion equation $u_t = u_{xx}$ (and many others). For non-compact $\mathbb{R}$, for $t \to \infty$ the solution $u(x,t)$ goes to zero for any x, whereas the whole mass is constant $\int u(x,t)dx = \int u(x,0)dx$. The mass is somewhere, but the solution does not feel it.

For the compactified $\mathbb{R}$, $u(x,t)$ tends to $\frac{1}{2}\delta_{+\infty} + \frac{1}{2}\delta_{-\infty}$, showing that the mass is collected in the boundary point of the state space but not outside. The right compactification makes the state space time-complete.
Any non-compact topological space can be compactified by adding one or some more points. The way in which this has to be done depends on the specific situation. As a result the space of observables becomes smaller and the space of probability measures becomes larger. This shows that a non-compact state space in an indication that we forgot to include some states.

4.3.2 Observables and generalized observables

As already mentioned, the distinction of observables and generalized observables is important for the existence of continuous semigroups in the case of continuous time.
In addition, there is a fundamental difference of both types of observables in the case of a non-discrete topology: Being a continuous function, in contrast to a generalized observable, an observable generally cannot observe a single point but only a point with its neighborhood. This reflects the problem that the result of measurements is indeed a real number but does not behave like a real number. As an example, we can consider a real measurement of the masses of three balls. We measure the masses and get the three results m_1, m_2 and m_3 with the property $m_1 = m_2$, $m_2 = m_3$ but $m_3 > m_1$, due to the accuracy. Real numbers do not have this property. But, this is typical for open covers of the state space (see sub-subsection 4.2.5). Thus, observables are practical observables, whereas generalized observables are theoretical ones.

4.3.3 Generalizations

Completely analogously we can consider state changes between two different state spaces $\gamma : \mathcal{Z}_1 \to \mathcal{Z}_2$. This leads to Markov operators $\mathbf{M} : \mathcal{C}(\mathcal{Z}_1) \to \mathcal{C}(\mathcal{Z}_2)$ with the properties $\mathbf{M} \geq 0$ and $\mathbf{M}\mathbb{1}_1 = \mathbb{1}_2$. In this situation, there are no stationary states or semigroups.
Problems, sometimes called non-Markovian, depending on time explicitely, i.e. equations of time $\dot{g} = \mathbf{A}(t)g$ are covered by this construction, too. It is much more difficult to show the existence of a solution of such an equation . But if a solution exists, it can be described by $g(t_2) = \mathbf{T}(t_2, t_1)g(t_1)$ with a family $\mathbf{T}(t_2, t_1)$ of Markov operators. Thus, for such systems time irreversibility holds as well.

4.3.4 Quantum Mechanics

When trying to apply the presented methods for quantum problems, at first we have to construct the compact state space. The state of a quantum system is a vector in the unit sphere in a Hilbert space $\mathcal{H}$. This sphere has a natural topology in which it is not compact. Theoretically, this sphere can be compactified but there is no natural way to do this.
The problem becomes clearer by using a different approach. We take some self-adjoint operator $\mathbf{H}$ – for example the Hamiltonian – with a discrete set of eigenvectors $f_1, f_2, ...$ as a base in $\mathcal{H}$. This set can be compactified considering converging sequences as continuous functions. This set is isomorphic to the set of bounded self-adjoint operators in $\mathcal{H}$ diagonalizable under the same base $f_1, f_2, ...$, i.e., the self-adjoint operators commuting with $\mathbf{H}$.
This is a well know fact: Dealing with quantum systems we have to consider an other concept of observables than the commutative algebra of continuous functions under the pointwise multiplication.

Appendix. The mathematical framework

In this section we collect the main implicitly used mathematical properties of the introduced objects. All proofs not given here can be found in the following literature: The theory of the spaces $\mathcal{C}$ and $\mathcal{C}^*$ in [9, 5]. The topological properties and the bidual formalism in [8]. The theory of positive semigroups in $\mathcal{C}$ in [1] and the proofs of the properties of the Lyapunov functionals and Jensen's inequality in [12].

A.1 The main function spaces and their properties

A.1.1 Topology

Definition 1: $\mathcal{Z}$ is a compact topological Hausdorff space.
Definition 2: $\mathcal{O}(\mathcal{Z})$ is the set of open sets of $\mathcal{Z}$.
Definition 3: $\mathcal{B}(\mathcal{Z})$ is the set of Borel sets of $\mathcal{Z}$, generated by $\mathcal{O}(\mathcal{Z})$.
Definition 4: Any open cover of $\mathcal{Z}$ contains a finite cover (Definition of compactness).
Definition 5: A function between topological spaces is called continuous if the pre-image of every open set is open. In the Euclidean space, this definition of continuity coincides with the usual (ε, δ) definition.

A.1.2 The Banach space of continuous functions (observables)

Definition 1: $\mathcal{C}(\mathcal{Z})$ is the set of all continuous functions from $\mathcal{Z}$ in the reals $\mathbb{R}$.
Notation 2: $f, g \in \mathcal{C}(\mathcal{Z})$.
Theorem 3: $\mathcal{C}(\mathcal{Z})$ is the topological dual of $\mathcal{Z}$.
Theorem 4: $\mathcal{C}(\mathcal{Z})$ is a vector space
Theorem 5: $\mathcal{C}(\mathcal{Z})$ is a Banach space with the norm $\|g\| = \sup_{z\in\mathcal{Z}} |g(z)|$
Theorem 6: $\mathcal{C}(\mathcal{Z})$ is a separable Banach space (contains a countable base) if and only if $\mathcal{Z}$ is compact.
Theorem 7: $\mathcal{C}(\mathcal{Z})$ is a Banach algebra under the pointwise multiplication $(f \cdot g)(z) = f(z)g(z)$
Theorem 8: $\mathcal{C}(\mathcal{Z})$ has a natural order $g \geq 0 \iff g(z) \geq 0,\ z \in \mathcal{Z}$
Theorem 9: $\mathcal{C}(\mathcal{Z})$ is a Banach lattice under this order.
Definition 10: g_+ and g_- are the positive and negative parts of g, $|g| = g_+ + g_-$ is the modulus.
Theorem 11: $g_+, g_-, |g| \in \mathcal{C}(\mathcal{Z})$
Theorem 12: $\exists\ z_+, z_- \in \mathcal{Z}$ with $g(z_+) = \sup_{z\in\mathcal{Z}} g(z)$, $g(z_-) = \inf_{z\in\mathcal{Z}} g(z)$.

A.1.3 The Banach space of Radon measures

Definition 1: $\mathcal{C}^*(\mathcal{Z})$ is the set of all Radon measures defined on $\mathcal{B}(\mathcal{Z})$.
Theorem 2: $\mathcal{C}^*(\mathcal{Z})$ is the topological bidual of $\mathcal{Z}$.
Theorem 3: $\mathcal{C}^*(\mathcal{Z})$ is the dual as a linear space of $\mathcal{C}(\mathcal{Z})$ with the dual pairing $\langle g, p\rangle = \int_{\mathcal{Z}} g(z)p(dz)$ (Riesz's Theorem)
Theorem 4: $\mathcal{C}^*(\mathcal{Z})$ is a Banach space with the norm $\|p\| = \sup_{\|g\|=1}\langle g, p\rangle$.
Theorem 5: $\mathcal{C}^*(\mathcal{Z})$ has a natural order (denote by $\geq_1$) as a space of measures $p \geq_1 0 \iff p(B) \geq 0,\ B \in \mathcal{B}(\mathcal{Z})$
Theorem 6: $\mathcal{C}^*(\mathcal{Z})$ has a natural order (denote by $\geq_2$) as a dual of a Banach lattice $p \geq_2 0 \iff \langle g, p\rangle \geq 0,\ g \geq 0,\ g \in \mathcal{C}(\mathcal{Z})$
Theorem 7: The orders $\geq_1$ and $\geq_2$ are equivalent (denote in the following by $\geq$)

Theorem 8: The weak* topology in $\mathcal{C}^*(\mathcal{Z})$ coincides with the weak topology in $\mathcal{C}(\mathcal{Z})$ and the topology in $\mathcal{Z}$, i.e., linear combinations of point measures are dense.

A.1.4 The spaces and their topologies

Definition 1: $\mathcal{C}_{\mathcal{B}}(\mathcal{Z})$ is the set of continuous functions on $\mathcal{Z}$ with the topology generated by the Borel sets $\mathcal{B}$ as open sets (measurable functions).
Theorem 2: $\mathcal{C}_{\mathcal{B}}(\mathcal{Z}) \subset \mathcal{C}^{**}(\mathcal{Z}) = \mathcal{Z}^{***}$.
Definition 3: Weak topology in $\mathcal{C}$: g_n tends to g, if $\langle g_n, p\rangle \to \langle g, p\rangle$ for all $p \in \mathcal{C}^*$.
Definition 4: Weak topology in $\mathcal{C}^*$: p_n tends to p, if $\langle \xi, p_n\rangle \to \langle \xi, p\rangle$ for all $\xi \in \mathcal{C}^{**}$.
Definition 5: Weak* topology in $\mathcal{C}^*$: p_n tends to p, if $\langle g, p_n\rangle \to \langle g, p\rangle$ for all $g \in \mathcal{C}$.
Definition 6: ι is the canonical embedding $\iota : \mathcal{Z} \to \mathcal{Z}^{**}$.
Theorem 7: The weak* topology in $\mathcal{C}^*(\mathcal{Z})$ coincides with the topology in $\mathcal{Z}$, i.e., point measures are dense, i.e., the pointwise convergency of continuous functions is the weak* convergency.
Theorem 8: In a non reflexive Banach space the weak* topology is weaker then the weak topology and the weak topology is weaker then the strong topology.

A.1.5 The convex set of probability measures (statistical states)

Definition 1: $\mathcal{S}^*$ is the set of positive and normalized measures in $\mathcal{C}^*(\mathcal{Z})$ (probability measures, statistical states, mixed states)

$$\mathcal{S}^*(\mathcal{Z}) = \left\{p \in \mathcal{C}^*(\mathcal{Z}) \;\middle|\; p \geq 0,\ \|p\| = \langle \mathbb{1}, p\rangle = p(\mathcal{Z}) = 1\right\}$$

Definition 2: $\mathcal{S}^*_e(\mathcal{Z})$ is the set of point measures in $\mathcal{C}^*(\mathcal{Z})$ (Dirac measures, δ-functions, pure states)

$$\mathcal{S}^*_e(\mathcal{Z}) = \{\delta_z \in \mathcal{C}^* \mid z \in \mathcal{Z}\}$$

Theorem 3: $\mathcal{S}^*(\mathcal{Z})$ is convex and closed in the weak* topology
Theorem 4: $\mathcal{S}^*_e(\mathcal{Z}) = \operatorname{extr}(\mathcal{S}^*)$
Theorem 5: $\mathcal{S}^*(\mathcal{Z}) = \overline{\operatorname{conv} \mathcal{S}^*_e(\mathcal{Z})}^{\,*}$
Theorem 6: $\mathcal{S}^*_e(\mathcal{Z}) = \iota\mathcal{Z}$

A.2 Markov operators and their adjoints

A.2.1 General operators

Definition 1: $\mathcal{L}(\mathcal{X})$ is the set of bounded linear operators in a Banach space $\mathcal{X}$.
Definition 2: In a Banach lattice $\mathcal{X}$ an operator $\mathbf{L} \in \mathcal{L}(\mathcal{X})$ is called (cone-)positive, if $g \geq 0$ imply $\mathbf{L}g \geq 0$.
Definition 3: $\mathbf{L}^* \in \mathcal{L}(\mathcal{C}^*)$ is the adjoint to $\mathbf{L} \in \mathcal{L}(\mathcal{C})$ $\langle \mathbf{L}g, p\rangle = \langle g, \mathbf{L}^*p\rangle$ for all $g \in \mathcal{C}$, $p \in \mathcal{C}^*$.
Definition 4: The adjoint can be similarly defined for densely defined unbounded operators.

A.2.2 Markov operators

Definition 1: A Markov operator $\mathbf{M}$ is a linear bounded operator in $\mathcal{C}(\mathcal{Z})$ with the properties $\mathbf{M}\mathbb{1} = \mathbb{1}$ and $\mathbf{M} \geq 0$.
Notation 2: The set of Markov operators is denoted by $\mathcal{M}$.
Theorem 3: $\mathcal{M}$ is convex.
Theorem 4: Markov operators are bounded $\|\mathbf{M}\| \leq 1$.

A.2.3 Adjoints of Markov operators

Notation 1: $\mathbf{M}^*$ is the adjoint of $\mathbf{M}$ if $\langle \mathbf{M}g, p\rangle = \langle g, \mathbf{M}^* p\rangle$ for all $g \in \mathcal{C}$ and $p \in \mathcal{C}^*$.
Notation 2: The set of adjoints Markov operators is denoted by $\mathcal{M}^*$.
Theorem 3: $\mathcal{M}^*$ is convex.
Theorem 4: $\|\mathbf{M}^*\| \le 1$.
Theorem 5: $\mathbf{M}^*\mathcal{S}^* \subset \mathcal{S}^*$.
Theorem 6: There exists a $\mu \in \mathcal{S}^*$ (fix point, stationary state) with $\mathbf{M}^*\mu = \mu$ (Perron-Frobenius-Krein-Rutman Theorem).
Counter example 7: Due to the non-reflexivity of $\mathcal{C}$, a bounded operator in $\mathcal{C}^*(\mathcal{Z})$ must not have a pre-dual. Proof by example:
Let $\xi_1, ..., \xi_n \in \mathcal{C}^{**}$ be n generalized observables not contained in $\mathcal{C}$ with the properties $\xi_i \ge 0$ and $\xi_1 + ... + \xi_n = \mathbb{1}$. Let $q_1, ..., q_n \in \mathcal{S}^*$ be n probability measures. Then, the projector $\mathbf{N}$ on the n-dimensional subspace $\mathrm{span}(q_1, ..., q_n)$ defined by

$$\mathbf{N}p = \sum\nolimits_{i=i}^{n} \langle \xi_i, p\rangle q_i$$

is bounded in $\mathcal{C}^*$ as every projector. The dual of this operator is

$$\mathbf{N}^*\xi = \sum\nolimits_{i=i}^{n} \langle \xi, q_i\rangle \xi_i$$

a projector in $\mathcal{C}^{**}$ but the restriction of $\mathbf{N}^*$ to $\mathcal{C}$ is unbounded and even not densely defined in $\mathcal{C}$. This example illustrates the problems of defining a statistical state change not via Markov operators.

A.2.4 Deterministic Markov operators

Definition 1: A deterministic Markov operator is a composition with a function $\gamma \in \mathrm{End}(\mathcal{Z})$, acting as

$$\big(\mathbf{M}_\gamma g\big)(z) = (g \circ \gamma)(z) = g\big(\gamma(z)\big) \ .$$

Notation 2: We denote the set of deterministic Markov operators by $\mathcal{M}_e$.
Theorem 3: $\mathcal{M}_e \subset \mathcal{M}$.
Theorem 4: $\mathbf{M} \in \mathcal{M}_e$ if and only if $\mathbf{M}(f \cdot g) = \mathbf{M}f \cdot \mathbf{M}g$, $f, g \in \mathcal{C}$ (algebra homomorphism)
Theorem 5: $\mathbf{M} \in \mathcal{M}_e$ if and only if $\mathbf{M}|g| = |\mathbf{M}g|$ (lattice homomorphism)
Theorem 6: $\mathbf{M} \in \mathcal{M}_e$ if and only if $\mathbf{M}F(g) = F(\mathbf{M}g)$, $g \in \mathcal{C}$, F strictly convex
Theorem 7: $\mathbf{M} \in \mathcal{M}_e$. Then, $\mathbf{M}^*\mathcal{S}^*_e \subset \mathcal{S}^*_e$
Theorem 8: $\mathbf{N}$ in $\mathcal{C}^*$ satisfies $\mathbf{N}\mathcal{S}^*_e \subset \mathcal{S}^*_e$ and has a pre-dual, then this pre-dual is in $\mathcal{M}_e$.
Theorem 9: $\mathbf{M} \in \mathcal{M}_e$ is an abstract pull-back operator, i.e., $\mathbf{M}_{\psi\circ\varphi} = \mathbf{M}_\varphi \mathbf{M}_\psi$ holds for $\psi, \varphi \in \mathrm{End}(\mathcal{Z})$
Theorem 10: $\mathbf{M} \in \mathcal{M}_e$, then $\mathbf{M}^*$ is an abstract push-forward operator, i.e., $\mathbf{M}^*_{\psi\circ\varphi} = \mathbf{M}^*_\psi \mathbf{M}^*_\varphi$ holds for $\psi, \varphi \in \mathrm{End}(\mathcal{Z})$

A.3 Inequalities for Markov operators

A.3.1 Multiplication operators

Theorem 1: Given $f \in \mathcal{C}$, the multiplication operator $\mathbf{P}_f$ defined by $\mathbf{P}_f g = f \cdot g$ is bounded in $\mathcal{C}$ with $\|\mathbf{P}_f\| = \|f\|$.

Theorem 2: The adjoint $\mathbf{P}_f^*$ of $\mathbf{P}_f$ is a bounded operator in $\mathcal{C}^*$ and acts via $(\mathbf{P}_f^* p)(B) = \int_B f(z)p(dz) = \langle \mathbb{1}_B \cdot f, p\rangle$.
Theorem 3: The operator $\mathbf{Q}_p : \mathcal{C} \to \mathcal{C}^*$ defined by $\mathbf{Q}_p f = \mathbf{P}_f^* p$ is a bounded operator.
Theorem 4: The inverse operator $\mathbf{Q}_p^{-1}$ can be defined on some domain in $\mathcal{C}^*$ (Radon-Nikodym Theorem). $\mathbf{Q}_p^{-1} q = f$ is the Radon-Nikodym derivative of q with respect to p.

A.3.2 Maximum principle

Theorem 1: Let $g_- = \min_{z\in\mathcal{Z}} g(z)$ and $g_+ = \max_{z\in\mathcal{Z}} g(z)$. Then, the following maximum principle for Markov operators $\mathbf{M}$ holds: $g_-\mathbb{1} \le \mathbf{M}g \le g_+\mathbb{1}$.
This follows directly from the definition of Markov operators and implies A.4.2.6.

A.3.3 Jensen's inequality

Theorem 1: For an arbitrary Markov operator $\mathbf{M}$, and an arbitrary convex function $F : \mathbb{R} \to \{\mathbb{R} \cup +\infty\}$ Jensen's inequality

$$F(\mathbf{M}g) \le \mathbf{M}F(g) \;, \mathbf{M} \in \mathcal{M},\; g \in \mathcal{C}(\mathcal{Z}) \tag{44}$$

holds. Equality holds for strictly convex functions if and only if $\mathbf{M}$ is a deterministic Markov operator.
Definition 2: We define a functional H on two arbitrary probability measures p and q as a generalized Legendre transform of $\langle F(g), q\rangle$

$$H[p,q] = \sup_{g\in\mathcal{C}} \big(\langle g,p\rangle - \langle F(g),q\rangle\big),\; p,q \in \mathcal{S}^* \tag{45}$$

Theorem 3: Then, for an arbitrary Markov operator $\mathbf{M}$ it holds

$$H[\mathbf{M}^*p, \mathbf{M}^*q] \le H[p,q] \;,\quad p,q \in \mathcal{S}^* \;,\quad \mathbf{M}^* \in \mathcal{M}^* \;. \tag{46}$$

Here, we recall the proof published in [12], because of its simplicity.
Denoting the range of $\mathbf{M}$ by $R(\mathbf{M})$ and using Jensen's inequality (44), we get

$$\begin{aligned} H[\mathbf{M}^*p, \mathbf{M}^*q] &= \sup_{g\in\mathcal{C}} \big(\langle g, \mathbf{M}^*p\rangle - \big\langle F(g), \mathbf{M}^*q\big\rangle\big) = \sup_{g\in\mathcal{C}} \big(\langle \mathbf{M}g, p\rangle - \big\langle \mathbf{M}F(g), q\big\rangle\big) \le \\ &\le \sup_{g\in\mathcal{C}} \big(\langle \mathbf{M}g, p\rangle - \big\langle F(\mathbf{M}g), q\big\rangle\big) = \sup_{h\in R(\mathbf{M})} \big(\langle h, p\rangle - \big\langle F(h), q\big\rangle\big) \le \\ &\le \sup_{h\in\mathcal{C}} \big(\langle h, p\rangle - \big\langle F(h), q\big\rangle\big) = H[p,q] \;. \end{aligned}$$

Remark 4: There are two places, involving the sign $\le$. The first comes from Jensen's inequality (44) and the second from the enlargement of the set where the supremum was taken from $h \in R(\mathbf{M})$ to $h \in \mathcal{C}$. Form this, it is clear that equality holds if $\mathbf{M}$ is a deterministic Markov operator with weakly dense range.
Corollary 5: As a special case, for q we can choose the equilibrium measure μ of $\mathbf{M}^*$. Then, we have $\mathbf{M}^*\mu = \mu$ and (46) becomes

$$H[\mathbf{M}^*p, \mu] \;\;\le\;\; H[p,\mu],\; p \in \mathcal{S}^* \tag{47}$$

A.4 Semigroups of Markov operators

A.4.1 Strong continuous semigroups and their generators

Definition 1: In $\mathcal{C}$, a operator family $\mathbf{T}(t)$ with $t \geq 0$ is called semigroup, if$\mathbf{T}(t_1 + t_2) = \mathbf{T}(t_1)\mathbf{T}(t_2) = \mathbf{T}(t_2)\mathbf{T}(t_1)$ and $\mathbf{T}(0) = \mathbf{I}$ holds.
Definition 2: A semigroup is strong continuous if $\lim_{t\to 0} \|\mathbf{T}(t)g - g\| = 0$ holds for all $g \in \mathcal{C}$.
Definition 3: The set

$$D(\mathbf{A}) = \{g \in \mathcal{C} | \lim_{t\to 0} \frac{1}{t}(\mathbf{T}(t)g - g) \;\text{ exists}\}$$

is called domain of the generator of the semigroup $\mathbf{T}(t)$.
Definition 4: The operator

$$\mathbf{A}g = \lim_{t\to 0} \frac{1}{t}(\mathbf{T}(t)g - g),\; g \in D(\mathbf{A})$$

is called the generator of the semigroup $\mathbf{T}(t)$.
Theorem 5: $D(\mathbf{A})$ is dense in $\mathcal{C}$.
Theorem 6: There exists a adjoint of $\mathbf{A}$, denoted by $\mathbf{A}^*$

A.4.2 Markov semigroups and Markov generators

Definition 1: A strong continuous semigroup $\mathbf{T}(t)$ is called a Markov semigroup, if $\mathbf{T}(t) \in \mathcal{M}$ for all $t \geq 0$.
Definition 2: The generator of a Markov semigroup is called Markov generator.
Theorem 3: $\mathbb{1} \in D(\mathbf{A})$, $\mathbf{A}\mathbb{1} = 0$.
Theorem 4: $\mathbf{T}^*(t)$ is a weak* continuous semigroup.
Theorem 5: There exists a $\mu \in \mathcal{S}^*$ (fix point, stationary state) with $\mathbf{T}^*(t)\mu = \mu$ for any t (Markov-Kakutani Theorem). This holds for any commuting set of Markov operators.
Theorem 6: A Markov generator satisfy the positive minimum principle. This is,

$$(\mathbf{A}g)(z_+) \leq 0,\; g \in D(\mathbf{A})\;,$$

where z_+ is the point with $g(z_+) = \sup_{z\in\mathcal{Z}} g(z)$.
Theorem 7: If a operator $\mathbf{A}$ is densely defined in $\mathcal{C}(\mathcal{Z})$, satisfy the positive minimum principle, $\mathbf{A}\mathbb{1} = 0$ and the equation $(\lambda - \mathbf{A})g = f$ is solvable for some $\lambda > 0$ and any $f \in \mathcal{C}(\mathcal{Z})$), then, $\mathbf{A}$ is a Markov generator.
Theorem 8: $\mu \in \mathcal{S}^*$ is a solution to $\mathbf{A}^*\mu = 0$ if and only if $\mathbf{T}^*(t)\mu = \mu$ for any t.

A.4.3 Inequalities for generators

Theorem 1: Let $\mathbf{M} \in \mathcal{M}$ satisfy $F(\mathbf{M}g) \leq \mathbf{M}F(g)$. Then, $F(\mathbf{T}(t)g) \leq \mathbf{T}(t)F(g)$ and for $g \in D(\mathbf{A})$ holds $\mathbf{A}G(g) \leq G'(g)\mathbf{A}g$.
Proof: Since $\mathbf{T}(t)$ is a semigroup of Markov operators, from $F(\mathbf{M}g) \leq \mathbf{M}F(g)$ follows $F(\mathbf{T}(t)g) \leq \mathbf{T}(t)F(g)$. Take $g \in D(\mathbf{A})$ we have

$$\begin{aligned} 0 \;&\geq\; \lim_{t\to 0} \frac{1}{t}\big(F(\mathbf{T}(t)g) - \mathbf{T}(t)F(g)\big) = \lim_{t\to 0} \frac{1}{t}\big((F(\mathbf{T}(t)g) - F(g)) - (\mathbf{T}(t)F(g) - F(g))\big) = \\ &=\; F'(g)\mathbf{A}g - \mathbf{A}F(g) \end{aligned}$$

Corollary 2: For $g \in D(\mathbf{A})$ inequality $\mathbf{A}g \cdot \mathrm{sign}(g) \leq \mathbf{A}|g|$ holds (Kato's inequality).

Proof: Take A.4.3.1 with the convex function $F(g) = |g|$.
Corollary 3: For $g \in D(\mathbf{A})$ inequality $2g\mathbf{A}g \leq \mathbf{A}g^2$ holds.
Proof: Take A.4.3.1 with the convex function $F(g) = g^2$.

A.4.4 Derivations

Definition 1: A non-zero operator $\mathbf{A}$ defined on a sub-algebra of $\mathcal{C}(\mathcal{Z})$ and satisfying $\mathbf{A}\mathbb{1} = 0$ and $\mathbf{A}(f \cdot g) = f \cdot \mathbf{A}g + g \cdot \mathbf{A}f$ is called an abstract derivation.
Theorem 2: The generator $\mathbf{A}$ of a strong continuous semigroup of deterministic Markov operators is an abstract derivation.
Theorem 3: Derivations are unbounded.
As a proof of this was hard to find, we give a proof of that important statement here:
We show that a derivation cannot be defined on any continuous function unless the derivation is the zero operator. We take $f \in \mathcal{C}$ and $z_0 \in \mathcal{Z}$ arbitrary and put $g = f - f(z_0)\mathbb{1}$. Because $\mathcal{C}$ is a Banach lattice, we can decompose $g = g_+ - g_-$. Then, we have $g(z_0) = 0$, $g_+(z_0) = 0$, $g_-(z_0) = 0$, $g_+^{\frac{1}{2}}(z_0) = 0$, $g_-^{\frac{1}{2}}(z_0) = 0$. Due to $\mathbf{A}\mathbb{1} = 0$ we have

$$\begin{aligned}
(\mathbf{A}f)(z_0) &= (\mathbf{A}g)(z_0) = (\mathbf{A}(g_+ - g_-))(z_0) = (\mathbf{A}g_+)(z_0) - (\mathbf{A}g_-)(z_0) = \\
&= (\mathbf{A}g_+^{\frac{1}{2}} \cdot g_+^{\frac{1}{2}})(z_0) - (\mathbf{A}g_-^{\frac{1}{2}} \cdot g_-^{\frac{1}{2}})(z_0) = \\
&= 2g_+^{\frac{1}{2}}(z_0)(\mathbf{A}g_+^{\frac{1}{2}})(z_0) - 2g_-^{\frac{1}{2}}(z_0)(\mathbf{A}g_-^{\frac{1}{2}})(z_0) = 0
\end{aligned}$$

Since z_0 was arbitrary, $\mathbf{A}$ maps f into 0. q.e.d.
Remark 4: The contradiction in the proof is that we assumed that any element of $D(\mathbf{A})$ can be decomposed as $g = g_+ - g_-$ with $g_+, g_- \in D(\mathbf{A})$. If $\mathcal{Z} \subset \mathbb{R}^m$ a differentiable function has not differentiable positive or negative parts, in general.
Remark 5: Theorem 3 implies that in a finite space, where every operator is bounded, there are no semigroups of deterministic Markov operators. Thus, any Lyapunov function has to decrease. This is important for approximations of first order derivatives (that have constant Lyapunov functions). As usual, approximating a continuous problem by a discrete one, one wants to preserve important properties of the continuous problem like positivity or mass conservation. But reversibility is a property that cannot be preserved keeping time continuous.

A.4.5 Markov semigroups in Lebesgue spaces

Theorem 1: $L_r(\mu)$ is the closure of $\mathcal{C}$ in the L_r-norm.
Theorem 2: A Markov semigroup $\mathbf{T}(t)$ is a strong continuous semigroup in $L_2(\mu)$ if $\mathbf{T}^*(t)\mu = \mu$.
Proof: Let $\mathbf{S}(t)$ the extension of $\mathbf{T}(t)$ in $L_2(\mu)$. $\mathbf{T}(t)$ is strong continuous and therefore weak continuous. We take $f \in \mathcal{C}$, then, because of A.3.3.1 with $F(x) = x^2$

$$\begin{aligned}
\|\mathbf{S}(t)f - f\|^2_{L_2(\mu)} &= \big\langle(\mathbf{T}(t)f - f)^2, \mu\big\rangle = \big\langle(\mathbf{T}(t)f)^2, \mu\big\rangle - 2\big\langle f \cdot \mathbf{T}(t)f, \mu\big\rangle + \big\langle f^2, \mu\big\rangle \leq \\
&\leq \big\langle \mathbf{T}(t)f^2, \mu\big\rangle - 2\big\langle f \cdot \mathbf{T}(t)f, \mu\big\rangle + \big\langle f^2, \mu\big\rangle = \\
&= 2\big\langle f^2, \mu\big\rangle - 2\big\langle f \cdot \mathbf{T}(t)f, \mu\big\rangle = \\
&= 2\big\langle f \cdot (\mathbf{T}(t)f - f), \mu\big\rangle \leq 2\|f\|_{\mathcal{C}}\big\langle|\mathbf{T}(t)f - f|, \mu\big\rangle \overset{t\to 0}{\to} 0
\end{aligned}$$

since $\mathbf{T}(t)f$ tends to f weakly.

References

[1] W. Arendt, A. Grabosch, G. Greiner, U. Groh, H.P. Lotz, U. Moustakas, R. Nagel, F. Neubrander, U. Schlotterbeck, *One-parameter Semigroups of Positive Operators*, Springer 1986.

[2] A. Arnold, P. Markowich, G. Toscani, A. Unterreiter, *On logarithmic Sobolev inequalities, Csiszar-Kullback inequalities, and the rate of convergence to equilibrium for Fokker–Planck type equations*, Preprint. Berlin, Techn. Univ., Fachbereich Mathematik (3), Preprint-Reihe Mathematik, 592, 1998.

[3] D. Bakry, *Functional inequalities for Markov semigroups*, Probability measures on groups, Mumbai : Inde (2004)

[4] R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics*, John Wiley & Sons, 1975.

[5] N. Dunford, J.T. Schwartz, *Theory of linear operators. Part I. General Theory*, Interscience Publishers 1958.

[6] C.W. Gardiner, *Handbook of Stochastic Methods*, Springer Series in Synergetics. Volume 13. Second Edition. 1985.

[7] M. Kac, *Probability and related topics in physical sciences*, Bull. Amer. Math. Soc. Volume 66, Number 6 (1960), 472-475.

[8] S. Kaplan, *The Bidual of* $C(X)$ *I*, North Holland, 1985.

[9] Z. Semadeni, *Banach Spaces of Continuous Functions*, Warszawa 1971.

[10] H. Stephan, *A dimension-reduced description of general Brownian motion by non-autonomous diffusion-like equations*, J. Mathematical Physics Analysis Geometry (MAG), 12 (2005) pp. 187-202.

[11] H. Stephan, *Modeling of drift-diffusion systems*, Zeitschrift fuer Angewandte Mathematik und Physik (ZAMP) 60 (2009) pp.33-53

[12] H. Stephan, *Lyapunov functions for positive linear evolution problems*, ZAMM Z. Angew. Math. Mech., 85 (2005) pp. 766-777

[13] C. Tzanakis, A.C. Grecos, *Classical Markovian Kinetic Equations: Explicit Form and H-Theorem*, arXiv:physics/9708031.