Technische Universität München
Zentrum Mathematik
Wissenschaftliches Rechnen

Coherent structures and transfer operators

Andreas Denner

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Ulrich Bauer
Prüfer der Dissertation:	1.	Prof. Dr. Oliver Junge
	2.	Prof. Dr. Eric Sonnendrücker
	3.	Prof. Dr. Clarence Rowley, Princeton University (nur schriftliche Beurteilung)

Die Dissertation wurde am 16.01.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 19.05.2017 angenommen.

Acknowledgements

First of all, I would like to thank Oliver Junge, my supervisor, for his guidance. I appreciate his continuous interest in my progress, friendly encouragement and positive attitude. It has been a pleasure working with you.

Special thanks goes to Clarence Rowley for many helpful discussions, and for inviting me to Princeton. Also, I want to thank Péter Koltai for his continuous interest in my work and the helpful guidance, and Daniel Karrasch for the many discussions and posing the right questions.

During my work on this thesis I have been supported by the Helmholtz Graduate School in Plasma Physics. I want to thank Eric Sonnendrücker at the Max-Planck-Institut für Plasma Physik for accepting to be my second supervisor and for his support. I had the joy of working with Jakob Ameres, also at the Max-Planck-Institut für Plasma Physik, whom I want to thank for our fruitful collaboration.

I also want to thank Sonya Gzyl and Isabella Wiegand coordinating the International School of Applied Mathematics (ISAM), for patiently answering my many organizational questions over the years.

The members of the research unit M3 at the Technische Universität München deserve mentioning for creating a pleasant atmosphere to work in.

Last but not least, I am very grateful to my family and friends, especially my girlfriend, for their encouragement and unconditional support.

Zusammenfassung

Der Zustandsraum selbst komplizierter Dynamischer Systeme lässt sich häufig in Mengen unterteilen, die durch starke Transportbarrieren umgeben sind. Solche *kohärenten Mengen* beeinflussen das Verhalten des Systems und sind deshalb äußerst hilfreich für dessen Verständnis. Diese Arbeit widmet sich der formalen Beschreibung und mathematischen Berechnung solcher kohärenter Mengen. Wir entwickeln hierzu einen konzeptionellen Zugang und benutzen dessen Verbindung zu Transferoperatoren um effiziente Algorithmen für die numerische Berechnung kohärenter Mengen zu entwickeln. Wir wenden diese Algorithmen auf verschiedene Probleme der Fluiddynamik und Plasmaphysik an.

Abstract

Even for complicated dynamical systems, it is often possible to subdivide the state space into several sets that are separated by strong transport barriers. Such *coherent sets* greatly influence the behavior of the system and are therefore helpful for its comprehensive characterization. This thesis deals with the conceptional description and mathematical computation of coherent sets. To this end we develop a conceptual approach to coherence and use its connection to transfer operators to develop efficient algorithms for the numerical computation of coherent partitions. We apply these algorithms to various problems in fluid dynamics and plasma physics.

Contents

Chapter 1

Introduction

Time-dependent processes are mathematically modeled via *dynamical systems*. Their complexity ranges from the relatively simple motion of a pendulum to complex currents in the oceans or atmosphere. The laws covering the dynamics are seldom sufficiently simple to analytically obtain explicit solutions of the system. Therefore we depend on numerical methods to analyze such systems. Even if it is possible to obtain a numerical solution, it does not necessarily reveal valuable insight into the system. The purpose of an analysis then, if numerically or analytically, is a comprehensive characterization of the system, like

- topological or geometric information of invariant sets, e.g. attractors or invariant manifolds,
- statistical information, e.g. probability distributions of trajectories in some subsets of the state space,
- information on the stability of those objects with respect to small random perturbations of the deterministic system [Junge, 1999].

Subsets of state space which are stable with respect to small random perturbations during their evolution over time are called *coherent sets*. In other words, at least within a finite time horizon, trajectories initiating within such a set stay inside during its evolution with high probability even under small perturbations. The edges of those sets hence form persistent, albeit leaky transport barriers and play a fundamental role in the evolution of dynamical systems. In geophysical flows, coherent sets organize the fluid flow and obstruct transport between them. For example, vortices and currents influence the horizontal distribution of heat in the oceans, and atmospheric vortices can trap chemicals. In a plasma reactor, the confinement of the plasma is of uttermost interest

in order to lose as little as possible energy, and to protect the facility [Padberg et al., 2007].
Various techniques have been developed for the qualitative and quantitative study of transport problems and the computation of coherent sets: One big class of algorithms is formed by geometric approaches that mainly aim at detecting transport barriers, so-called Lagrangian coherent structures, see [Haller, 2000, Haller, 2001, Haller and Beron-Vera, 2012]. The concept of shape coherence is also of geometric nature [Ma and Bollt, 2013]. Another big group of approaches relies on the observation that coherent sets are closely connected to *transfer operators*, describing the evolution of densities of particles due to the given dynamics, and their (singular) spectrum. This fact is well-known in the autonomous setting, where *almost invariant sets* e.g. molecules are computed [Dellnitz and Junge, 1999, Deuflhard and Weber, 2005, Deuflhard et al., 2000, Deuflhard and Schütte, 2004, Koltai, 2010]. In the context of non-autonomous dynamical systems this connection was utilized first in [Froyland and Padberg, 2009] and various applications have developed from there [Froyland, 2013, Froyland et al., 2010a, Froyland et al., 2010b]. Methods based on transfer operators focus on the computation of the coherent sets themselves instead of their transport barriers. Recently, also purely data oriented algorithms [Hadjighasem et al., 2016, Banisch and Koltai, 2016] and differential geometric approaches [Froyland and Kwok, 2016, Karrasch and Keller, 2016] have been developed. Another group of approaches uses the Koopman operator and its spectrum for the computation of meaningful structures in a dynamical system [Rowley et al., 2009].

This thesis focuses on the computation of coherent sets, and contributes to the advancement of the set-oriented analysis of dynamical systems through the following:

1. *A rigorous definition of coherence and its connection to almost invariance.* Instead of relying on numerical dissipation we directly include diffusion into the dynamical system and in the definition of coherence. We establish a connection to the well understood concept of almost invariance. This allows a generalization of important theorems and high order numerical methods.

2. *Usage of transfer operator families.* We generalize the mathematical notion of coherence such that families of transfer operators can be used. In addition to conceptional enhancement, this allows very efficient discretizations and purely data-driven algorithms.

3. *Application to plasma physics.* We apply the not yet well-known concept and methods to various processes covering dynamics in plasma physics. This includes up to four dimensional problems.

In the following we give a detailed outline of this thesis.

In *Chapter 2*, we review on major concepts used in this manuscript, e.g. dynamical systems, transfer operators, Ulam's method and equations from plasma physics.

In *Chapter 3*, we derive a first mathematical notion of coherence based on previous works. We use the Frobenius-Perron operator to develop a heuristic algorithm computing a partition of the state space into a meaningful collection of n coherent sets. To this end, an existing method for the computation of coherent sets, focusing on partitions into a coherent sets and its complement, is generalized. As discretization Ulam's method, a Galerkin-projection onto the indicator functions of boxes, is used. The numerical dissipation added by *Ulam's method* can be interpreted as small diffusion, and consequently does not need to be included into the model. The algorithm is tested with the standard examples *Double gyre* and *Bickley jet*.

In *Chapter 4*, we include diffusion into the dynamical system – instead of relying on numerical dissipation. We establish a rigorous connection to the well understood concept of almost invariance. This allows a generalization of an important theorem [Huisinga and Schmidt, 2006] introducing a lower bound for the computed almost invariance to coherence. If white noise is used as diffusion, coherent sets may be computed by directly solving the Fokker-Planck equation. More precisely, instead of computing the evolution of the basis of our approximation space under the deterministic dynamics and then applying diffusion, we directly compute the evolution of this basis under the stochastic push forward operator given by the solution operator of the Fokker-Planck equation. This advection-diffusion equation can efficiently be discretised using spectral collocation (cf. also [Froyland et al., 2013]). In order to deal with aliasing in the case of dominating advection, a skew symmetric form of the advection term is used. In order to deal with stiffness in time due to the Laplace operator, an exponential time differentiation *(etd)* integrator is employed. As a key advantage of the new method, we only need to sample the vector field at each time instance on a fixed grid of rather coarse resolution. In particular, we do not need to integrate trajectories of (Lagrangian) particles and no interpolation of the vector field to points off the grid is needed.

In *Chapter 5*, we generalize the mathematical notion of coherence to not rely only on the initial and final time, but to include all intermediate times of the time interval observed. This results in the analysis of whole time-parameterized families of transfer operators. Those families are well-known and frequently used in optimization [Becker et al., 2007], [Tröltzsch, 2005]. They are also introduced in [Lasota and Mackey, 1993], Chapter 7.4, for the computation of invariant states. The motivation for this approach is twofold. First, common existing transfer operator methods, like Ulam's method introduced in Chapter 3, consider the dynamical system at initial and final time. They only implicitly know about what happens during the evolution. Our approach is a generalization considering the system at all time instants of interest. Second, especially in

applications, most of the time, only a limited set of observations (data) is available, but at many time instants. Again it is favorable to use all those data. We introduce a coarse meshfree discretization of the transfer operator family that leads to an algorithm only requiring data, which is comparable to recently developed, purely data-driven algorithms and hence connecting those to set oriented methods.

In *Chapter 6*, we close this thesis with an application of the developed methods to several problems in plasma physics. We identify coherent structures and transport barriers for several problems based on the the *Vlasov-Poisson* and *Vlasov-Maxwell* equations. This includes the *Two-stream instability*, the *Bump-on-tail instability* in two and four dimensions and the *Streaming Weibel instability*.

Chapter 2

Theory and background

2.1 Dynamical systems

In general a dynamical system is given as following.

Definition 1. *A **dynamical system** is a triple $(T, \Omega, \mathcal{T})$, where T is a monoid (e.g. $T \in \{\mathbb{N},\ \mathbb{R},\ \mathbb{R}_+\}$), Ω is a non-empty topological Hausdorff space and $\mathcal{T}$ is a function $\mathcal{T} : T \times \Omega \to \Omega$ with*

$$\mathcal{T}(0, x) = x \qquad \forall x \in \Omega, \tag{2.1}$$

$$\mathcal{T}(s, \mathcal{T}(t, x)) = \mathcal{T}(s + t, x) \qquad \forall s, t \in T, \tag{2.2}$$

and the mapping $(t, x) \to \mathcal{T}(t, x)$ from $T \times \Omega \to \Omega$ is continuous.

The function $\mathcal{T}(t, x) =: \mathcal{T}^{0,t}(x)$ is called the *flow map* or *evolution function* of the dynamical system. It associates to every point in the set Ω, and at initial time 0, a unique image, depending on the variable t, called the evolution parameter. Ω is called *phase space* or *state space*. The variable x represents an initial state of the system. If $T = \mathbb{N}$ we say $(T, \Omega, \mathcal{T})$ is a *discrete-time* dynamical system. We then write $\mathcal{T}(x) := \mathcal{T}^1(x)$ for one iteration of the map. For fixed $x \in \Omega$, $\Gamma_x := \{\mathcal{T}(t, x) : t \in T\}$ is called the *orbit* of x.

If the dynamics are non-autonomous, i.e. depend on the initial time, we write the flow map as $\mathcal{T} : T \times T \times \Omega \to \Omega$. Note that in this case we use the convention, that the first argument of $\mathcal{T}$ denotes the initial, and the second argument denotes the final time. The flow map hence associates to a point x in Ω at initial time t_0 a unique image

$\mathcal{T}(t_0, t, x) := \mathcal{T}^{t_0,t}(x)$, at time t. Conditions (2.1), (2.2) change to

$$\begin{aligned} \mathcal{T}(t_0, t_0, x) &= x && \forall x \in \Omega, \\ \mathcal{T}(t_0, s+t, x) &= \mathcal{T}(t, s+t, \mathcal{T}(t_0, t, x)) && \forall t_0, s, t \in T. \end{aligned}$$

2.2 Markov operators

In this section, we shortly introduce the concept of *Markov operators* and as special case the *Frobenius-Perron* and its adjoint, the *Koopman operator*. Originally developed for the analysis of chaotic systems and the computation of invariant states, those operators contain all information of a dynamical system and are hence naturally well suited for the analysis of those. A very readable and detailed introduction to Markov operators is given in [Lasota and Mackey, 1993], a nice motivation for the Frobenius-Perron operator is given in [Boyarsky and Gora, 2012].

Remark. *We denote with* $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$, $p \in \{1, 2, \ldots, \infty\}$ *the well-konwn Lebesgue spaces. When the meaning is clear, we sometimes omit the* σ*-algebra* $\mathcal{A}$ *or the measure* μ*. With* $\|\cdot\|_{\mathbb{L}^p(\Omega,\mathcal{A},\mu)}$ *we denote the corresponding* $\mathbb{L}^p$*-norm. If there is no danger of confusion, we write* $\|\cdot\|_{p,\mu}$ *or* $\|\cdot\|_p$ *for better readability.*

Definition 2. *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space with a* σ*-algebra* $\mathcal{A}$ *and a* σ*-finite measure* μ*. Any linear operator* $\mathcal{P} : \mathbb{L}^1(\Omega) \to \mathbb{L}^1(\Omega)$ *satisfying*

1. $\mathcal{P}f \geq 0$ *for* $f \geq 0, f \in \mathbb{L}^1(\Omega)$,
2. $\|\mathcal{P}f\|_1 = \|f\|_1$ *for* $f \geq 0, f \in \mathbb{L}^1(\Omega)$

is called a ***Markov operator****.*

Note that Markov operators are monotonic, i.e.

$$f(x) \geq g(x) \Rightarrow \mathcal{P}f(x) \geq \mathcal{P}g(x) \text{ for a.e. } x \in \Omega$$

because

$$f - g \geq 0 \Rightarrow \mathcal{P}(f-g) \geq 0 \Rightarrow \mathcal{P}f \geq \mathcal{P}g.$$

2.2.1 Frobenius-Perron operator

We motivate the Frobenius-Perron operator via considering a chaotic system, where the computation of trajectories is naturally ill-conditioned. To be precise we look at the *logistic map*

$$\begin{aligned} \mathcal{T} : [0,1] &\to [0,1] \\ x &\mapsto 4x(1-x), \end{aligned}$$

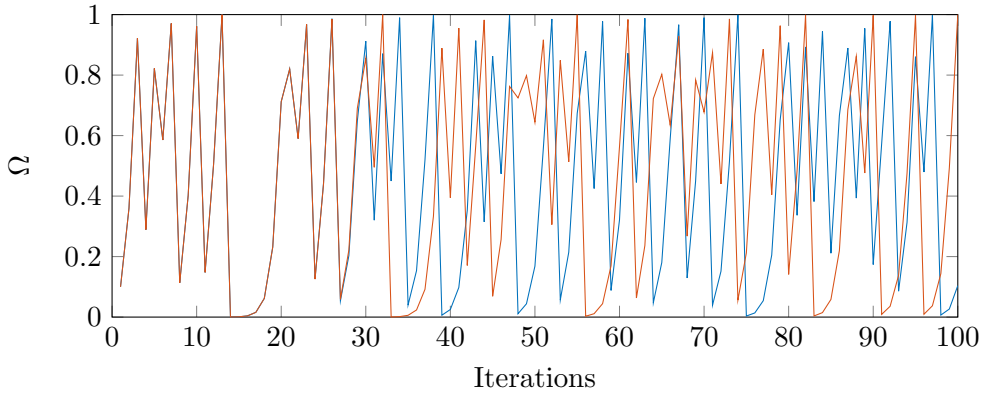

Figure 2.1: Logistic map: Two trajectories starting in 0.1 (blue) and $0.1 + 10^{-10}$ (red).

which is a standard example for a chaotic map. We consider two trajectories with initial values 0.1 and $0.1 + 10^{-10}$. In Figure 2.1 we show the first 100 iterations and see that they quickly diverge. In [Lasota and Mackey, 1993] it is observed that it is easier to predict the evolution of densities than trajectories (see also Gibb's original book [Gibbs, 1902] for the origins of statistical mechanics).

But how do the dynamics propagate densities? To see this we assume for a moment that μ is a probability measure and the random variable

$$X \sim f_0 \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu) \text{ , i.e.} \quad \mathbb{P}(X \in A) = \int_A f_0 \, d\mu \quad \forall A \in \mathcal{A},$$

where $\mathbb{P}(X \in A)$ denotes the probability that X is in A. We want to compute the distribution f_1 of $\mathcal{T}(X)$, $\mathcal{T}(X) \sim f_1$:

$$\mathbb{P}(\mathcal{T}(X) \in A) = \mathbb{P}(X \in \mathcal{T}^{-1}(A)) = \int_{\mathcal{T}^{-1}(A)} f_0 \, d\mu \overset{!}{=} \int_A f_1 \, d\mu.$$

When does f_1 exist? We note that for a σ-finite measure μ and a function $f \in \mathbb{L}^1_+(\Omega, \mu)$, the image measure

$$\nu_f(A) := \int_{\mathcal{T}^{-1}(A)} f(x)\mu(dx) \qquad \forall A \in \mathcal{A}$$

again is a measure: First, because $\mathcal{T}$ is measurable, $\mathcal{T}^{-1}(A) \in \mathcal{A}$. As

1. $\nu_f(\emptyset) = \int_\emptyset f(x)\mu(dx) = 0$,
2. $\nu_f(A) = \int_{\mathcal{T}^{-1}(A)} \underbrace{f(x)}_{\geq 0} \mu(dx) \geq 0 \qquad \forall A \in \mathcal{A}$,

3. $\nu_f(\bigcup_{k\geq 1} A_k) = \int_{\mathcal{T}^{-1}(\bigcup_{k\geq 1} A_k)} f(x)\mu(dx) = \int_{\bigcup_{k\geq 1}\mathcal{T}^{-1}(A_k)} f(x)\mu(dx)$
$= \sum_{k\geq 1}\int_{\mathcal{T}^{-1}(A_k)} f(x)\mu(dx) = \sum_{k\geq 1}\nu_f(A_k)$

   for all $A_1, A_2, \ldots \in \mathcal{A}$ disjoint,

ν_f is a measure on $\mathcal{A}$. If additionally ν_f is absolutely continuous with respect to the σ-finite measure μ, $\nu_f \ll \mu$, then by the Radon-Nikodym theorem (e.g. [Lasota and Mackey, 1993, Theorem 2.2.1]) there exists a unique function in $\mathbb{L}^1_+(\Omega, \mu)$, which we denote by $\mathcal{P}f$ such that

$$\nu_f(A) = \int_A \mathcal{P}f(x)\mu(dx) = \int_{\mathcal{T}^{-1}(A)} f(x)\mu(dx) \quad \forall A \in \mathcal{A}.$$

The question under which conditions the image measures ν_f are absolutely continuous w.r.t μ is answered by the following lemma:

Lemma 1. *$\nu_f \ll \mu$ for all $f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu)$ iff $\nu := \mu \circ \mathcal{T}^{-1} \ll \mu$.*

Proof. Assume $A \in \mathcal{A}$. We start with the first implication.

$$\begin{aligned}
&\mu(A) = 0 \\
&\Rightarrow \nu_f(A) = 0 \ \forall f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu) \\
\Rightarrow &\int_{\mathcal{T}^{-1}(A)} f(x)\ \mu(dx) = 0 \ \forall f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu) \\
\Rightarrow &\int_{\mathcal{T}^{-1}(A)} 1_\Omega(x)\ \mu(dx) = 0 \\
&\Rightarrow \mu(\mathcal{T}^{-1}(A)) = 0 \\
&\Rightarrow \nu(A) = 0 \\
&\Rightarrow \nu \ll \mu.
\end{aligned}$$

The second implication follows from

$$\begin{aligned}
&\mu(A) = 0 \Rightarrow \nu(A) = 0 \\
&\Rightarrow \mu(\mathcal{T}^{-1}(A)) = 0 \\
\Rightarrow &\int_{\mathcal{T}^{-1}(A)} f(x)\ \mu(dx) = 0 \quad \forall f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu) \\
&\Rightarrow \nu_f(A) = 0 \quad \forall f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu) \\
&\Rightarrow \nu_f \ll \mu \quad \forall f \in \mathbb{L}^1_+(\Omega, \mathcal{A}, \mu).
\end{aligned}$$

□

To sum up we have to demand from the flow map $\mathcal{T}$, that $\mu \circ \mathcal{T}^{-1} \ll \mu$ in order to define the image density $\mathcal{P}f$ on $\mathbb{L}^1_+$. This claim is formalized in the following definition:

Definition 3. *A measurable transformation $\mathcal{T} : \Omega \to \Omega$ on a measure space $(\Omega, \mathcal{A}, \mu)$ is called* ***non-singular*** *under $\mathcal{T}$ if $\mu(\mathcal{T}^{-1}(A)) = 0$ for all $A \in \mathcal{A}$ such that $\mu(A) = 0$.*

Now let $f \in \mathbb{L}^1(\Omega)$ be arbitrary, i.e. not necessarily non-negative. We write $f = f^+ - f^-$, where

$$f^+(x) = \max(0, f(x)), \qquad f^-(x) = \max(0, -f(x))$$

and define

$$\mathcal{P}f := \mathcal{P}f^+ - \mathcal{P}f^-.$$

Thus

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

and hence

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

As for integrable f, g holds that

$$\left[\int_A f(x)\mu(dx) = \int_A g(x)\mu(dx) \quad \forall A \in \mathcal{A}\right] \qquad \Rightarrow \qquad f = g \text{ a.e.}$$

and $\mathcal{T}$ is non-singular, $\mathcal{P}f$ is uniquely defined.
We use these considerations for the following definition.

Definition 4. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. If $\mathcal{T} : \Omega \to \Omega$ is a non-singular transformation, $f \in \mathbb{L}^1(\Omega)$, the unique operator $\mathcal{P} : \mathbb{L}^1(\Omega) \to \mathbb{L}^1(\Omega)$ defined by*

$$\int_A \mathcal{P}f(x)\mu(dx) = \int_{\mathcal{T}^{-1}(A)} f(x)\mu(dx) \quad \forall A \in \mathcal{A} \tag{2.3}$$

is called the ***Frobenius-Perron operator*** *corresponding to $\mathcal{T}$.*

The Frobenius-Perron operator hence describes the evolution of a given density f under the transformation $\mathcal{T}$, see Figure 2.2. It is consequently also called *push forward* operator. It follows directly from 2.3, that the Frobenius-Perron operator has the following properties:

Corollary 1. *1. $\mathcal{P}$ is a linear operator on $\mathbb{L}^1(\Omega)$,*

2. $f \geq 0 \Rightarrow \mathcal{P}f \geq 0$,

3. $\int_\Omega \mathcal{P}f(x)\mu(dx) = \int_\Omega f(x)\mu(dx)$.

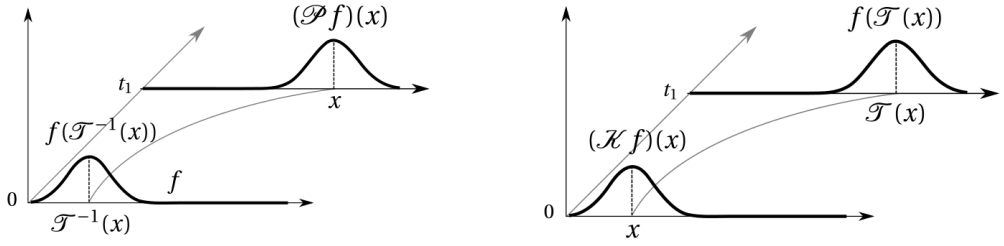

Figure 2.2: The Frobenius-Perron operator $\mathcal{P}$ pushes forward a density f, the Koopman operator $\mathcal{K}$ (see Chapter 2.2.2) pulls back a density f.

Remark. *Note that the Frobenius-Perron operator is hence a Markov operator.*

Next we want to explore if the Frobenius-Perron operator can also be defined on $\mathbb{L}^p,\ p = 1, 2, \ldots, \infty$. To this end we denote with $\mathbb{E}(f)$ the expected value of a random variable $X \sim f$, and with $\mathbb{E}(f|\mathcal{C})$ the conditional expectation of a random variable $X \sim f$ given $\mathcal{C}$. We will see, that this is possible if the measure μ is not affected by the transformation $\mathcal{T}$.

Definition 5. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $\mathcal{T}$ a non-singular transformation. We say that a measure μ is **invariant** with respect to $\mathcal{T}$, if*

$$\mu(\mathcal{T}^{-1}A) = \mu(A) \quad \forall A \in \mathcal{A}.$$

As

$$(1_A \circ \mathcal{T})(x) = \begin{cases} 1 & \mathcal{T}(x) \in A \\ 0 & \text{else} \end{cases} = \begin{cases} 1 & x \in \mathcal{T}^{-1}(A) \\ 0 & \text{else} \end{cases} = 1_{\mathcal{T}^{-1}(A)}(x),$$

in this case holds

$$\begin{aligned} \int_{\mathcal{T}^{-1}(B)} (1_A \circ \mathcal{T})(x)\ \mu(dx) &= \int_{\mathcal{T}^{-1}(B)} 1_{\mathcal{T}^{-1}(A)}(x)\mu(dx) = \mu(\mathcal{T}^{-1}(A) \cap \mathcal{T}^{-1}(B)) \\ &= \mu(\mathcal{T}^{-1}(A \cap B)) = \mu(A \cap B) = \int_B 1_A\ \mu(dx) \quad \forall A \in \mathcal{A}. \end{aligned}$$

With a standard argument from measure theory (see e.g. [Lasota and Mackey, 1993, Remark 2.2.6]), this also holds for simple functions which are dense in $\mathbb{L}^1(\Omega, \mathcal{A}, \mu)$. We can state for $f \in \mathbb{L}^1(\Omega, \mathcal{A}, \mu)$, $B \in \mathcal{A}$ that, if μ is invariant,

$$\int_{\mathcal{T}^{-1}B} (f \circ \mathcal{T})(x)\mu(dx) = \int_B f(x)\mu(dx). \tag{2.4}$$

Theorem 1. *If $(\Omega, \mathcal{A}, \mu)$ is a probability space, $\mathcal{T}$ a non-singular transformation and μ an invariant measure with respect to $\mathcal{T}$, then*

$$\mathcal{P}f \circ \mathcal{T} = \mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A}).$$

Proof. $\mathcal{P}f \circ \mathcal{T}$ is clearly $\mathcal{T}^{-1}\mathcal{A}$ measurable. We have for $A = \mathcal{T}^{-1}B \in \mathcal{T}^{-1}\mathcal{A}$:

$$\begin{aligned}\mathbb{E}(1_A(\mathcal{P}f \circ \mathcal{T})) &= \int_A \mathcal{P}f \circ \mathcal{T} d\mu \overset{\mu \text{ invariant, (2.4)}}{=} \int_B \mathcal{P}f \; d\mu \overset{\text{Definition } \mathcal{P}}{=} \int_{\mathcal{T}^{-1}B} f \; d\mu \\ &= \int_A f \; d\mu = \mathbb{E}(1_A f).\end{aligned}$$

□

Theorem 2. *If $\Omega \subset \mathbb{R}^d$ open, $d \in \mathbb{N}$ and $\mathcal{T} : \Omega \to \Omega$ is a Lebesgue-preserving homeomorphism, then*

$$\mathcal{P}f = f \circ \mathcal{T}^{-1} \text{ a.e.}$$

Proof. As $\mathcal{T}, \mathcal{T}^{-1}$ continuous, $\mathcal{T}^{-1}\mathcal{A} = \mathcal{A}$ and therefore $\mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A}) = f$. Using Theorem 1, we have that $\mathcal{P}f \circ \mathcal{T} = f$ and as $\mathcal{T}$ is invertible $\mathcal{P}f = f \circ \mathcal{T}^{-1}$. □

Theorem 3. *If $(\Omega, \mathcal{A}, \mu)$ is a probability space, $\mathcal{T}$ a non-singular transformation and μ an invariant measure with respect to $\mathcal{T}$, then $\mathcal{P}$ is a contraction for every $1 \leq p \leq \infty$.*

Proof. For $1 \leq p < \infty$

$$\begin{aligned}\|\mathcal{P}f\|_p^p &= \int_\Omega |\mathcal{P}f|^p d\mu = \int_\Omega |\mathcal{P}f \circ \mathcal{T}|^p d\mu = \int_\Omega |\mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A})|^p d\mu \\ &\leq \int_\Omega |\mathbb{E}(|f|^p|\mathcal{T}^{-1}\mathcal{A})| d\mu = \mathbb{E}(\mathbb{E}(|f|^p|\mathcal{T}^{-1}\mathcal{A})) = \mathbb{E}(|f|^p) = \int_\Omega |f|^p d\mu = \|f\|_p^p\end{aligned}$$

via using Jensen's inequality. For $p = \infty$ we have

$$\|\mathcal{P}f\|_\infty \overset{\mu \text{ invariant}}{=} \|\mathcal{P}f \circ \mathcal{T}\|_\infty = \|\mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A})\|_\infty \leq \|f\|_\infty,$$

as $\operatorname{ess\,sup}_{x\in\Omega}\mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A}) \leq \operatorname{ess\,sup}_{x\in\Omega} f$. □

Corollary 2. *If $(\Omega, \mathcal{A}, \mu)$ is a measure space with a finite measure μ, $\mu(\Omega) < \infty$. Then $\mathcal{P}$ is a contraction on $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$ for all $1 \leq p \leq \infty$.*

Proof. Via defining the probability measure $\hat{\mu} := \mu/\mu(\Omega)$, $\mathcal{P}_\epsilon$ is a contraction on $\mathbb{L}^p(\Omega, \mathcal{A}, \hat{\mu})$ for all $1 \leq p \leq \infty$. Furthermore

$$f \in \mathbb{L}^p(\Omega, \mathcal{A}, \hat{\mu}) \Leftrightarrow f \in \mathbb{L}^p(\Omega, \mathcal{A}, \mu).$$

□

Remark. *This allows us to e.g. consider the Frobenius-Perron operator on any Lebesgue space $\mathbb{L}^p(\Omega, \mathcal{B}, \lambda)$, $1 \leq p \leq \ldots \leq \infty$, with $\Omega \subset \mathbb{R}^d$ compact, and λ invariant under $\mathcal{T}$.*

Frobenius-Perron without invariant measure

Now we consider the case, when the measure μ is not invariant under $\mathcal{T}$. Then $\mathcal{P}$ is not stochastic in the sense that $\mathcal{P}1_\Omega = 1_\Omega$ and can not necessarily be defined on $\mathbb{L}^p,\ 1 \leq p \leq \infty$. However, if we remember $\nu := \mu \circ \mathcal{T}^{-1}$, the image measure of μ under the transformation $\mathcal{T}$ and assume $\mathcal{T}$ to be non-singular, i.e. $\nu \ll \mu$, we can introduce the transfer operator

$$\begin{aligned} \tilde{\mathcal{P}} : \mathbb{L}^1(\Omega, \mathcal{A}, \mu) &\longrightarrow \mathbb{L}^1(\Omega, \mathcal{A}, \nu) \\ \int_A \tilde{\mathcal{P}} f \ d\nu &= \int_{\mathcal{T}^{-1}(A)} f \ d\mu. \end{aligned} \tag{2.5}$$

If μ and ν are absolutely continuous with respect to the Lebesgue measure λ with Radon-Nikodym derivatives h_μ and $h_\nu = \mathcal{P}h_\mu$, respectively, we can compute

$$\begin{aligned} \int_A \tilde{\mathcal{P}} f \ d\nu = \int_{\mathcal{T}^{-1}(A)} f \ d\mu = \int_{\mathcal{T}^{-1}(A)} f h_\mu \ d\lambda = \int_A \mathcal{P}(f h_\mu) d\lambda \\ = \int_A \frac{\mathcal{P}(f h_\mu)}{\mathcal{P} h_\mu} \ d\nu. \end{aligned}$$

Hence, we can write for $\tilde{\mathcal{P}}$:

$$\tilde{\mathcal{P}} f = \frac{\mathcal{P}(f h_\mu)}{\mathcal{P} h_\mu} = \frac{\mathcal{P}(f h_\mu)}{h_\nu},$$

which can consequently be seen as normalized Frobenius-Perron operator, such that $\tilde{\mathcal{P}}$ is stochastic, i.e.

$$\tilde{\mathcal{P}} 1_\Omega = \frac{\mathcal{P}(1_\Omega h_\mu)}{\mathcal{P} h_\mu} = 1_\Omega,$$

see also [Froyland, 2013]. Furthermore note that, if μ is invariant, $\tilde{\mathcal{P}} = \mathcal{P}$.

Furthermore, for functions f in $\mathbb{L}^1(\Omega, \mathcal{A}, \nu)$ holds that

$$\int_B f \ d\nu = \int_{\mathcal{T}^{-1} B} f \circ \mathcal{T} \ d\mu, \tag{2.6}$$

as for $f = 1_A,\ A,\ B \in \mathcal{A}$ we can compute

$$\int_{\mathcal{T}^{-1}(B)} 1_A \circ \mathcal{T} \ d\mu = \int_{\mathcal{T}^{-1}(B)} 1_{\mathcal{T}^{-1}(A)} d\mu = \mu(\mathcal{T}^{-1}(A \cap B)) = \nu(A \cap B) = \int_B 1_A \ d\nu.$$

and with that (2.6) holds for simple functions and hence for all functions $f \in \mathbb{L}^1(\Omega, \mathcal{A}, \nu)$.

Theorem 4. *If $(\Omega, \mathcal{A}, \mu)$ is a probability space and $\mathcal{T}$ a non-singular transformation, then*

$$\tilde{\mathcal{P}} f \circ \mathcal{T} = \mathbb{E}_\mu(f|\mathcal{T}^{-1}\mathcal{A})$$

Proof. $\tilde{\mathcal{P}} f \circ \mathcal{T}$ is clearly $\mathcal{T}^{-1}\mathcal{A}$ measurable. We have for $A = \mathcal{T}^{-1}(B) \in \mathcal{T}^{-1}\mathcal{A}$:

$$\begin{aligned}\mathbb{E}_\mu(1_A(\tilde{\mathcal{P}} f \circ \mathcal{T})) &= \int_A \tilde{\mathcal{P}} f \circ \mathcal{T} d\mu \overset{(2.6)}{=} \int_B \tilde{\mathcal{P}} f \; d\nu \overset{(2.5)}{=} \int_{\mathcal{T}^{-1}(B)} f \; d\mu \\ &= \int_A f \; d\mu = \mathbb{E}_\mu(1_A f).\end{aligned}$$

□

Theorem 5. *If $(\Omega, \mathcal{A}, \mu)$ is a probability space and $\mathcal{T}$ a non-singular transformation, then $\tilde{\mathcal{P}}$ is a contraction for every $1 \leq p \leq \infty$.*

Proof. For $1 \leq p < \infty$

$$\begin{aligned}\|\tilde{\mathcal{P}} f\|_{\nu,p}^p &= \int_\Omega |\tilde{\mathcal{P}} f|^p d\nu \overset{(2.6)}{=} \int_\Omega |\tilde{\mathcal{P}} f \circ \mathcal{T}|^p d\mu = \int_\Omega |\mathbb{E}_\mu(f|\mathcal{T}^{-1}\mathcal{A})|^p d\mu \\ &\leq \int_\Omega |\mathbb{E}_\mu(|f|^p|\mathcal{T}^{-1}\mathcal{A})| d\mu = \mathbb{E}_\mu(\mathbb{E}_\mu(|f|^p|\mathcal{T}^{-1}\mathcal{A})) = \mathbb{E}_\mu(|f|^p) = \int_\Omega |f|^p d\mu = \|f\|_p^p.\end{aligned}$$

via using Jensen's inequality. For $p = \infty$ we have

$$\|\tilde{\mathcal{P}} f\|_{\nu,\infty} \overset{\text{Def. } \nu}{=} \|\tilde{\mathcal{P}} f \circ \mathcal{T}\|_{\mu,\infty} = \|\mathbb{E}(f|\mathcal{T}^{-1}\mathcal{A})\|_{\mu,\infty} \leq \|f\|_{\mu,\infty}.$$

□

Corollary 3. *If $(\Omega, \mathcal{A}, \mu)$ is a measure space with a finite measure μ. Then $\tilde{\mathcal{P}} : \mathbb{L}^p(\Omega, \mathcal{A}, \mu) \to \mathbb{L}^p(\Omega, \mathcal{A}, \nu)$ is a contraction for all $1 \leq p \leq \infty$.*

To sum up, even if a finite measure μ is not invariant, and hence $\mathcal{P}$ is no contraction on $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$, we can construct a normalized operator $\tilde{\mathcal{P}}$ that is a contraction on $\mathbb{L}^p(\Omega, \mathcal{A}, \mu) \to \mathbb{L}^p(\Omega, \mathcal{A}, \nu)$.

2.2.2 Koopman operator

We now introduce a third type of operator closely related to the Frobenius-Perron operator.

Definition 6. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $\mathcal{T} : \Omega \to \Omega$ a non-singular transformation and $f \in \mathbb{L}^\infty(\Omega, \mu)$. The operator $\mathcal{K} : \mathbb{L}^\infty(\Omega, \mu) \to \mathbb{L}^\infty(\Omega, \mu)$ defined by*

$$\mathcal{K} f = f \circ \mathcal{T}$$

*is called **Koopman operator** with respect to $\mathcal{T}$.*

Note that, as $\mathcal{T}$ is non-singular, $f_1(x) = f_2(x)$ a.e. implies that $f_1(\mathcal{T}(x)) = f_2(\mathcal{T}(x))$ a.e. Furthermore as $f(x) \leq \|f\|_\infty$ a.e. implies that $f(\mathcal{T}(x)) \leq \|f\|_\infty$ a.e. and therefore $\mathcal{K}$ is a contraction on $\mathbb{L}^\infty(\Omega)$, $\|\mathcal{K}f\|_\infty \leq \|f\|_\infty$. Hence $\mathcal{K}$ is well defined as an operator from $\mathbb{L}^\infty(\Omega,\mu) \to \mathbb{L}^\infty(\Omega,\mu)$.
$\mathcal{K}$ can be interpreted as going backward in time and assigning to a given density $f \in \mathbb{L}^\infty(\Omega,\mu)$ at final time t_1 the initial density $\mathcal{K}f$ evolving to become f. $\mathcal{K}$ is therefore also called *pull-back* operator, see Figure 2.2. $\mathcal{K}$ is also obviously linear.

Lemma 2 ([Lasota and Mackey, 1993, Section 3.3]). *For every $f \in \mathbb{L}^1(\Omega,\mu)$, $g \in \mathbb{L}^\infty(\Omega,\nu)$ holds*

$$\langle \mathcal{P}f, g\rangle = \langle f, \mathcal{K}g\rangle, \tag{2.7}$$

so that $\mathcal{K}$ is adjoint to the Frobenius-Perron operator $\mathcal{P} : \mathbb{L}^1(\Omega,\mu) \to \mathbb{L}^1(\Omega,\mu)$.

Proof. We first check (2.7) for $g = 1_A$, $A \in \mathcal{A}$:

$$\begin{aligned}\langle \mathcal{P}f, g\rangle &= \int_\Omega \mathcal{P}f\ 1_A\ d\mu = \int_A \mathcal{P}f\ d\mu = \int_{\mathcal{T}^{-1}(A)} f\ d\mu \\ &= \int_\Omega f\ 1_A \circ \mathcal{T}\ d\mu = \int_\Omega f\ \mathcal{K}1_A\ d\mu = \langle f, \mathcal{K}g\rangle\end{aligned}$$

Because (2.7) holds for $g = 1_A$, it is true for any simple function g and hence for all functions $g \in \mathbb{L}^\infty(\Omega)$, (see [Lasota and Mackey, 1993, Remark 2.2.6]). □

With the same argument as in Lemma 2.7, we can state

Corollary 4. *Let μ be invariant under $\mathcal{T}$. For every $f \in \mathbb{L}^p(\Omega,\mu)$, $g \in \mathbb{L}^q(\Omega,\mu)$, $\frac{1}{p}+\frac{1}{q} = 1$ holds*

$$\langle \mathcal{P}f, g\rangle = \langle f, \mathcal{K}g\rangle, \tag{2.8}$$

so that the Koopman operator $\mathcal{K} : \mathbb{L}^q(\Omega,\mu) \to \mathbb{L}^q(\Omega,\mu)$ is adjoint to the Frobenius-Perron operator $\mathcal{P} : \mathbb{L}^p(\Omega,\mu) \to \mathbb{L}^p(\Omega,\mu)$.

2.2.3 Semi groups of Frobenius-Perron operators

We next introduce the Frobenius-Perron operator for dynamical systems and therefore recapture some basics of semi group theory as developed for example in [Evans, 2010] Chapter 7.4.

Definition 7. *Let $(Y, \|\cdot\|)$ be a Banach space. A one parameter family $(\mathcal{P}^t)_{t\geq 0}$ of bounded linear operators $\mathcal{P}^t : Y \to Y$ is called a **semi group** on Y, if*

1. *$\mathcal{P}^0 = I$ (I denoting the identity on Y),*

2. $\mathcal{P}^{t+s} = \mathcal{P}^t \mathcal{P}^s \ \forall t, s \geq 0.$

Furthermore, if $\left\|\mathcal{P}^t\right\| \leq 1$*, then* $(\mathcal{P}^t)_{t\geq 0}$ *is called a semi group of* ***contractions****.*

Let Ω be a topological Hausdorff space and $\mathcal{A}$ be the Borel $\sigma-$algebra, i.e. the smallest $\sigma-$algebra containing all open subsets of Ω. Since, for any fixed $t \in \mathbb{R}^+$ in a dynamical system $(\mathcal{T}^{0,t})_{t\geq 0}$ the transformation $\mathcal{T}^{0,t}$ is measurable, we can adopt the discrete time definitions of the Frobenius-Perron operator directly for the continuous time case.

Let μ be a measure on Ω and let all transformations $\mathcal{T}^{0,t}$ of a dynamical system $(\mathcal{T}^{0,t})_{t\geq 0}$ be non-singular, that is

$$\mu(\mathcal{T}^{t,0}(A)) = 0 \quad \forall A \in \mathcal{A} \text{ such that } \mu(A) = 0.$$

Then analogously to Definition 2.3, the property

$$\int_A \mathcal{P}^t f(x)\mu(dx) = \int_{\mathcal{T}^{t,0}(A)} f(x)\mu(dx) \quad \forall A \in \mathcal{A}$$

for each fixed $t \geq 0$ uniquely defines the Frobenius-Perron operator $\mathcal{P}^t : \mathbb{L}^1(\Omega) \to \mathbb{L}^1(\Omega)$, corresponding to the transformation $\mathcal{T}^{0,t}$. Hence, for fixed $t \geq 0$, the operator $\mathcal{P}^t : \mathbb{L}^1(\Omega) \to \mathbb{L}^1(\Omega)$ is a Markov operator.

Remark. *The Frobenius-Perron operator* $\mathcal{P}$ *is always associated to a non-singular map* $\mathcal{T} : \Omega \to \Omega$*. If we want to make clear that this map is a flow map* $\mathcal{T}^{0,t} : \Omega \to \Omega$*, at fixed time* $t \geq 0$*, we write* $\mathcal{P}^t$ *or* $\mathcal{P}^{0,t}$ *for* $\mathcal{P}$*. If we want to make clear that this map is a flow map* $\mathcal{T}^{t_0,t_1} : \Omega \to \Omega$*, depending on the fixed initial time* t_0 *and the fixed final time* t_1*, we write* $\mathcal{P}^{t_0,t_1}$ *for* $\mathcal{P}$*.*

The Frobenius-Perron operator also fulfills two properties analogue to (2.1) and (2.2) in the definition of dynamical systems:
As $(\mathcal{T}^{0,t})_{t\geq 0}$ is a dynamical system it holds that $\mathcal{T}^{0,s+t} = \mathcal{T}^{0,s} \circ \mathcal{T}^{0,t}$ and hence $\mathcal{T}^{s+t,0} = \mathcal{T}^{t,0} \circ \mathcal{T}^{s,0}$. This property is inherited to $(\mathcal{P}^t)_{t\geq 0}$.

$$\begin{aligned}
\int_A \mathcal{P}^{s+t} f(x)\mu(dx) &= \int_{\mathcal{T}^{s+t,0}(A)} f(x)\mu(dx) \\
&= \int_{\mathcal{T}^{t,0}(\mathcal{T}^{s,0})(A)} f(x)\mu(dx) \\
&= \int_{\mathcal{T}^{s,0}(A)} \mathcal{P}^t f(x)\mu(dx) \\
&= \int_A \mathcal{P}^s(\mathcal{P}^t f(x))\mu(dx) \qquad \forall A \in \mathcal{A}.
\end{aligned}$$

Thus we can conclude

$$\mathcal{P}^{s+t}f = \mathcal{P}^s(\mathcal{P}^t f) \quad \forall f \in \mathbb{L}^1(\Omega), s,t \geq 0. \tag{2.9}$$

Furthermore, since $\mathcal{T}^{0,0}(x) = x$ we have $(\mathcal{T}^{0,0})^{-1}(A) = A$ for all $A \in \mathcal{A}$ and consequently

$$\int_A \mathcal{P}^0 f(x)\mu(dx) = \int_{(\mathcal{T}^{0,0})^{-1}(A)} f(x)\mu(dx) = \int_A f(x)\mu(dx) \tag{2.10}$$

which is equivalent to

$$\mathcal{P}^0 f = f \ \forall f \in \mathbb{L}^1(\Omega). \tag{2.11}$$

Hence $\mathcal{P}^t$ satisfies properties analogue to (2.1) and (2.2) in the definition of dynamical systems and therefore defines a semi group on $\mathbb{L}^1(\Omega)$.

Moreover it fulfills the following properties, too.

Definition 8. *Let $(\Omega,\mathcal{A},\mu)$ be a measure space. A family of operators $\mathcal{P}^t : \mathbb{L}^1(\Omega) \to \mathbb{L}^1(\Omega)$ $t \geq 0$, satisfying*

1. *$\mathcal{P}^t$ is a linear operator on $\mathbb{L}^1(\Omega)$,*
2. *$f \geq 0 \Rightarrow \mathcal{P}^t f \geq 0 \quad \forall f \in \mathbb{L}^1(\Omega)$,*
3. *$\int_\Omega \mathcal{P}^t f(x)\mu(dx) = \int_\Omega f(x)\mu(dx)$,*
4. *$\mathcal{P}^{s+t}f = \mathcal{P}^s(\mathcal{P}^t f) \quad \forall f \in \mathbb{L}^1(\Omega), s,t \geq 0$,*
5. *$\mathcal{P}^0 f = f \quad \forall f \in \mathbb{L}^1(\Omega)$.*

*is called a **stochastic semi group**.*

1.-3. are inherited by Corollary 1. *4.*, *5.* we showed above. Thus the family of Frobenius-Perron operators on $\mathbb{L}^1(\Omega)$ is also a stochastic semi group. Note that *3.* holds iff $\mathcal{T}^{t,0}(\Omega) = \Omega$.

2.3 Galerkin projections

In this section let $(\Omega, \mathcal{A}, \mu)$ be a measure space, Ω a compact metric space and μ a σ-finite measure. Let $l_1, \dots l_n \in (\mathbb{L}^p)'(\Omega)$ be elements of the dual space $(\mathbb{L}^p)'(\Omega)$ of $\mathbb{L}^p(\Omega)$, $p \in \{1, 2, \dots, \infty\}$. Let $V_n := \text{span}(\varphi_1, \dots, \varphi_n)$, where $\varphi_i : \Omega \to \mathbb{R}$ are bounded, piecewise continuous and linearly independent functionals.

Remark. *Note that the φ_i may also depend on n but we omit the index n for a clearer notation.*

We proceed analogous to [Koltai, 2010], Chapter 3. It holds that

$$V_n \subset \mathbb{L}^\infty(\Omega), \qquad \dim V_n = n.$$

We define now the projection $\pi_n : \mathbb{L}^p(\Omega) \to V_n$ such that

$$l_i(f - \pi_n f) = 0 \qquad \forall f \in \mathbb{L}^p(\Omega)\ \forall i = 1, \dots, n. \tag{2.12}$$

Lemma 3. *The projection π_n exists and is unique.*

Proof. Let be $f \in \mathbb{L}^p(\Omega)$. Then $\pi_n f = \sum_{j=1}^n c_j \varphi_j,\ c_j \in \mathbb{R}$ and hence

$$\begin{aligned} l_i(f) = l_i(\pi_n f) = l_i(\sum_{j=1}^n c_j \varphi_j) = \sum_{j=1}^n c_j l_i(\varphi_j) \quad \forall i = 1, \dots, n. \\ \Leftrightarrow Lc = l \end{aligned}$$

where $L = (L_{ij}) \in \mathbb{R}^{n \times n}$, $L_{ij} = l_i(\varphi_j)$, $l = (l_i) \in \mathbb{R}^n$, $l_i = l_i(f) \in \mathbb{R}$, $c = (c_j) \in \mathbb{R}^n$. Because the functionals l_i are linear independent, L is injective and as $\mathbb{R}^n$ is a finite dimensional vector space also surjective. Hence there is a unique solution to $Lc = f$ and a unique projection $\pi_n : \mathbb{L}^p(\Omega) \to V_n$.

□

As $(\mathbb{L}^p(\Omega))' \cong \mathbb{L}^q(\Omega)$, $\frac{1}{p} + \frac{1}{q} = 1$, see for example [Evans, 2010], there are $\psi_1, \dots, \psi_n \in \mathbb{L}^q(\Omega)$, such that

$$l_i(f) = \int_\Omega f \psi_i d\mu \qquad \forall f \in \mathbb{L}^p(\Omega) \quad \forall i = 1, \dots, n.$$

The ψ_i are called *test functions*.
For general ψ_i, π_n is called *Petrov-Galerkin projection*.
If $\psi_i = \varphi_i\ \forall i = 1, \dots, n$, π_n is called *Galerkin projection*.
We are mainly interested in Galerkin projections and hence can write for $f \in \mathbb{L}^p(\Omega)$:

$$\pi_n f = \sum_{i=1}^n c_i \varphi_i \qquad \in V_n, \quad \varphi_i \in \mathbb{L}^\infty(\Omega).$$

Set

$$b_j := \int_\Omega f \varphi_j d\mu = l_j(f) = l_j(\pi_n f) = \int_\Omega \pi_n f \varphi_j d\mu = \sum_{i=1}^n c_i \int_\Omega \varphi_i \varphi_j \quad \forall j = 1, \dots, n. \tag{2.13}$$

If we define $A_n^{ij} := \int_\Omega \varphi_i \varphi_j d\mu(x)$, $b = (b_1, \dots, b_n)'$ and $c = (c_1, \dots, c_n)'$ we can write equation (2.13) as a Matrix vector equation:

$$c = A_n^{-1} b. \tag{2.14}$$

With $\Phi_n := (\varphi_1, \dots, \varphi_n)'$ holds

$$A_n = \int_\Omega \Phi_n \Phi_n^T d\mu(x), \qquad\qquad b = \int_\Omega \Phi_n f d\mu(x)$$

and thus

$$\pi_n f = \Phi_n^T A_n^{-1} \int_\Omega \Phi_n f d\mu(x).$$

2.4 Discretization of the Frobenius-Perron operator

In this section we present a common method to discretize the Frobenius-Perron operator

$$\begin{aligned} &\mathcal{P} : \mathbb{L}^p(\Omega, \mathcal{A}, \lambda) \to \mathbb{L}^p(\Omega, \mathcal{A}, \lambda) \\ &\int_A \mathcal{P} f(x) \lambda(dx) = \int_{\mathcal{T}^{-1}(A)} f(x) \lambda(dx) \quad \forall A \in \mathcal{A} \ \ \forall\ f \in \mathbb{L}^p(\Omega, \mathcal{A}, \lambda) \end{aligned}$$

with $\Omega \subset \mathbb{R}^d$ compact, and $\mathcal{T}$-invariant Lebesgue measure λ. Therefor we use Ulam's well known method (see e.g. [Koltai, 2010] or originally [Ulam, 1960]). We partition the state space Ω into finitely many disjoint subsets $B_1, \dots B_n$, i.e. $\Omega = \bigcup_{i=1}^n B_i = \Omega$, where each set B_i has a piecewise smooth boundary ∂B_i, such that the unit outer normal vector n_i exists almost everywhere. In all our considerations the B_i will be hyper rectangles and are called *boxes*. The size of the boxes will decrease at least linearly in $\frac{1}{n}$, i.e.

$$\exists c \geq 0 \ \forall i = 1 \dots n : \lambda(B_i) \leq \frac{c}{n} \ \text{ and } d(B_i) \leq \frac{c}{n}, \tag{2.15}$$

where d denotes the longest side length of the box, which we call diameter. Let $\chi_i, i = 1, \dots, n$ denote the characteristic function on B_i, i.e. $\chi_i : \Omega \to \mathbb{R}$,

$$\chi_i(x) = \begin{cases} 1 & x \in B_i, \\ 0 & x \in \Omega \backslash B_i. \end{cases}$$

We construct the approximation space $V_n := \operatorname{span}(\chi_1, \dots, \chi_n)$ and represent functions in and operators on V_n with respect to the basis $B_n' = \left\{\chi_1^1, \dots, \chi_n^1\right\}$, where $\chi_i^1 = \frac{\chi_i}{\lambda(B_i)}$. Next, we define a projection $\pi_n : \mathbb{L}^p(\Omega) \to V_n,\ p \in \{1, 2, \dots\}$, by

$$\pi_n f = \sum_{i=1}^n a_i \chi_i^1 \qquad \text{with} \qquad a_i = \int_{B_i} f(x) d\lambda(x). \tag{2.16}$$

We define the approximate Frobenius-Perron operator

$$\mathcal{P}_n^t := \pi_n \mathcal{P}^t \pi_n, \tag{2.17}$$

yielding $\mathcal{P}_n^t : \mathbb{L}^p(\Omega) \to V_n^1$. We are also in the position to compute the matrix representation P_n^t for the operator $\mathcal{P}_n^t|_{V_n}$ with respect to the basis B_n':

$$\mathcal{P}_n^t|_{V_n}\chi_j^1 = \pi_n\mathcal{P}^t\pi_n\chi_j^1 = \sum_{i=1}^n \int_{B_i} \mathcal{P}^t\chi_j^1 d\lambda \cdot \chi_i^1 = \sum_{i=1}^n \left(\frac{1}{\lambda(B_j)}\int_{B_i}\mathcal{P}^t\chi_j d\lambda\right)\chi_i^1$$

as $\pi_n\chi_j^1 = \chi_j^1$ and therefore P_n^t reads as

$$P_n^{t,ij} = \frac{1}{\lambda(B_j)}\int_{B_i}\mathcal{P}^t\chi_j d\lambda.$$

If we use the defining property of the Frobenius-Perron operator we can compute this to

$$\frac{1}{\lambda(B_j)}\int_{B_i}\mathcal{P}_n^t\chi_j d\lambda = \frac{1}{\lambda(B_j)}\int_{\mathcal{T}^{t,0}(B_i)}\chi_j d\lambda = \frac{\lambda(B_j\cap\mathcal{T}^{t,0}(B_i))}{\lambda(B_j)},$$

and hence

$$P_n^{t,ij} = \frac{\lambda(B_j\cap\mathcal{T}^{t,0}(B_i))}{\lambda(B_j)}.$$

2.5 Ulam's method as Galerkin projection and stochastic interpretation

Ulam's method above corresponds to a Galerkin projection $\pi_n : \mathbb{L}^p(\Omega) \to V_n$, $p \in \{1,2,\dots\}$, with basis $B_n' := \left\{\frac{\chi_1}{\lambda(B_1)},\dots,\frac{\chi_n}{\lambda(B_n)}\right\}$ and functionals

$$\begin{aligned} l_i(f) &:= \int_\Omega \chi_i^1 f d\lambda \\ &= \frac{1}{\lambda(B_i)}\int_\Omega \chi_i f d\lambda(x) \\ &= \frac{1}{\lambda(B_i)}\int_{B_i} f d\lambda(x) \qquad \forall f\in\mathbb{L}^p(\Omega)\ \forall i = 1,\dots n. \end{aligned}$$

Because with

$$\pi_n f = \sum_{i=1}^n c_i\chi_i \ \in V_n,$$

$c_i \in \mathbb{R}$ constant, and due to (2.12),

$$l_i(f) = l_i(\pi_n f),$$

we have

$$\begin{aligned}\frac{1}{\lambda(B_i)}\int_{B_i} f(x)d\lambda(x) &= l_i(f) = l_i(\pi_n f)\\ &= \frac{1}{\lambda(B_i)}\int_{B_i} \pi_n f d\lambda(x)\\ &= \frac{1}{\lambda(B_i)}\int_{B_i} \sum_{j=1}^{n} c_j\chi_j(x)d\lambda(x)\\ &= \frac{1}{\lambda(B_i)}\int_{B_i} c_i d\lambda(x)\\ &= c_i \qquad \forall f \in \mathbb{L}^p(\Omega)\ \forall i = 1,\dots n.\end{aligned} \tag{2.18}$$

Hence the projection π_n in (2.16) in Ulam's method corresponds to the Galerkin projection (2.18).
Furthermore Ulam's discretization has a stochastic interpretation, too. If we use the basis B' given above, the transition matrix P_n^t of $\mathcal{T}^{0,t}$ with respect to B' is given by

$$\begin{aligned}P_n^{t,ij} &= \frac{\lambda(B_j \cap \mathcal{T}^{t,0}(B_i))}{\lambda(B_j)}\\ &= \int_{\mathcal{T}^{t,0}(B_i)} \frac{\chi_j}{\lambda(B_j)} d\lambda\\ &= \int_{B_i} \mathcal{P}_n^t \frac{\chi_j}{\lambda(B_j)} d\lambda\\ &= \int_{B_i} \mathcal{P}_n^t \chi_j^1 d\lambda.\end{aligned}$$

$\mathcal{P}_n^t$ describes the probability that a point $x \in B_j$ chosen randomly via a uniform distribution (with respect to λ) on B_j is mapped to B_i by $\mathcal{T}^{0,t}$. Hence $P_n^{t,ij}$ is the transition rate from B_j to B_i. $\mathcal{P}_n^t$ is obviously positive and

$$\begin{aligned}\sum_{i=1}^{n} P_n^{t,ij} &= \sum_{i=1}^{n} \frac{\lambda(B_j \cap \mathcal{T}^{t,0}(B_i))}{\lambda(B_j)}\\ &= \frac{1}{\lambda(B_j)} \sum_{i=1}^{n} \lambda(B_j \cap \mathcal{T}^{t,0}(B_i))\\ &= \frac{1}{\lambda(B_j)} \lambda\left(\bigcup_{i=1}^{n} (B_j \cap \mathcal{T}^{t,0}(B_i)\right), \text{ as the } B_j \text{ are disjoint},\\ &= \frac{1}{\lambda(B_j)} \lambda\left(B_j \cap \bigcup_{i=1}^{n} \mathcal{T}^{t,0}(B_i)\right) = 1 \qquad \forall j = 1,\dots n,\end{aligned} \tag{2.19}$$

as $\mathcal{T}^{t,0}(\Omega) = \Omega$. Hence P_n^t is a stochastic matrix (column wise) and thus Ulam's method defines a Markov jump process on Ω_n.

In summary the Markov operator $\mathcal{P}^t$ is approximated by an operator $\mathcal{P}_n^t$ whose representation on V_n is a stochastic matrix. Ulam's discretization of the Frobenius-Perron operator leads to an approximation of the deterministic process

$$Z^{0,t} = \mathcal{T}^{0,t}$$

with probability 1. It is approximated by a stochastic process $(Y_n^t)_{t\geq 0}$ such that if $Y_n^0 = y \in \Omega$ ($Y_n^0 \sim \delta_y$) then

$$Y_n^t \sim \sum_{i=1}^{n} P_n^{t,ij_y} \chi_i^1$$

where j_y is the unique index of B_{j_y} with $y \in B_{j_y}$.

Convergence

Theorem 6 ([Koltai, 2010])**.** *The discrete Frobenius-Perron operator $\mathcal{P}_n^t$ converges to the Frobenius-Perron operator $\mathcal{P}^t$ for $n \to \infty$ point-wise in $\mathbb{L}^p(\Omega, \mathcal{B}, \lambda),\ p \in \{1, 2, \dots\}$.*

Proof. First we show, that for all $f \in \mathbb{L}^p(\Omega, \mathcal{B}, \lambda)$:

$$\pi_n f \to f\ (n \to \infty). \tag{2.20}$$

Therefor let $f \in C^0(\Omega)$. As $\mathbb{L}^p(\Omega, \lambda) \subset \mathbb{L}^q(\Omega, \lambda),\ 1 \leq q \leq p \leq \infty$, there is a $c > 0$, such that

$$\begin{aligned}
\|\pi_n f - f\|_{\mathbb{L}^p(\Omega)} &\leq c \|\pi_n f - f\|_{\mathbb{L}^\infty(\Omega)} \\
&= c \left\| \left(\sum_{i=1}^{n} \left(\int_{B_i} f(y) d\lambda(y) \right) \chi_i^1(x) \right) - f(x) \right\|_{\mathbb{L}^\infty(\Omega)} \\
&= c \left\| \left(\sum_{i=1}^{n} \frac{1}{\lambda(B_i)} \int_{B_i} f(y) d\lambda(y) \chi_i(x) \right) - f(x) \right\|_{\mathbb{L}^\infty(\Omega)}.
\end{aligned}$$

According the mean value theorem for integration $\exists \xi_i \in \overline{B}_i$:

$$f(\xi_i) = \frac{1}{\lambda(B_i)} \int_{B_i} f(y) d\lambda(y)$$

for all $i = 1 \dots n$. Hence

$$\begin{aligned}
\|\pi_n f - f\|_{\mathbb{L}^p(\Omega)} &\leq c \left\| \left(\sum_{i=1}^{n} f(\xi_i) \chi_i(x) \right) - f(x) \right\|_{\mathbb{L}^\infty(\Omega)} \\
&= c \max_{i \in \{1,\dots,n\}} \|f(\xi_i) - f\|_{\mathbb{L}^\infty(B_i)},
\end{aligned}$$

as $\Omega = \bigcup_{i=1,\dots,n} B_i$. Furthermore

$$\begin{aligned} & c \max_{i\in\{1,\dots,n\}} \|f(\xi_i) - f\|_{\mathbb{L}^\infty(B_i)} \\ =& c \max_{i\in\{1,\dots,n\}} \sup_{x\in B_i} |f(\xi_i) - f(x)| \\ \leq& c' \max_{i\in\{1,\dots,n\}} \sup_{x\in B_i} \|\xi_i - x\| \to 0 \ (n\to\infty) \end{aligned}$$

as $f \in C^0(\Omega)$, and according to (2.15), with $\lambda(B_i)$ also the diameter $d(B_i) \to 0$ $(n \to \infty)$.

As $C^0(\Omega)$ is dense in $\mathbb{L}^p(\Omega)$ and π_n is continuous we have proven (2.20).

Due to Theorem 3, the operator $\mathcal{P}^t : \mathbb{L}^p(\Omega) \to \mathbb{L}^p(\Omega)$ is bounded and

$$\begin{aligned} \left\|\mathcal{P}^t f - \mathcal{P}^t_n f\right\|_{\mathbb{L}^p(\Omega)} &= \left\|\mathcal{P}^t f - \pi_n \mathcal{P}^t \pi_n f\right\|_{\mathbb{L}^p(\Omega)} \\ &= \left\|\mathcal{P}^t f - \mathcal{P}^t \pi_n f + \mathcal{P}^t \pi_n f - \pi_n \mathcal{P}^t \pi_n f\right\|_{\mathbb{L}^p(\Omega)} \\ &\leq \|\mathcal{P}^t (f - \pi_n f)\|_{\mathbb{L}^p(\Omega)} + \|(Id - \pi_n) \underbrace{\mathcal{P}^t \pi_n f}_{:= g \in \mathbb{L}^1(\Omega)} \|_{\mathbb{L}^p(\Omega)} \\ &\leq \|\mathcal{P}^t\|_{\text{op}} \|(Id - \pi_n) f\|_{\mathbb{L}^p(\Omega)} + \|(Id - \pi_n) g\|_{\mathbb{L}^p(\Omega)} \ \to 0 \ \forall f \in \mathbb{L}^p(\Omega) \end{aligned}$$

for $n \to \infty$ in $\mathbb{L}^p(\Omega)$ as $\pi_n \to Id$ point-wise in $\mathbb{L}^p(\Omega)$, $\mathcal{P}^t$ bounded $\forall t$.

□

2.6 Bochner spaces

2.6.1 Integration of Banach space valued functions

In Chapter 5 we will consider the whole evolution of a special set $A \subset \Omega$ over a fixed time interval $[t_0, t_1]$. The indicator function of such a time parameterized family of sets $(A_t)_{t\in[t_0,t_1]}$ is a mapping $1_{(A_t)} : [t_0, t_1] \longrightarrow \mathbb{L}^2(\Omega)$ and hence a Banach space valued function. We now shortly extend the notions of measurability, integrability, etc. to functions

$$f : [t_0, t_1] \to X, \quad t_0, t_1 \in \mathbb{R}^+, \ t_0 < t_1,$$

where X is a real Banach space with norm $\|.\|_X$. The approach is similar to the one for real valued functions and for example given in [Lasota and Mackey, 1993] (see also the references therein).

Definition 9. *1. A function $s : [t_0, t_1] \to X$ is called **simple**, if it is of the form*

$$s(t) = \sum_{i=1}^{m} \chi_{E_i}(t) u_i \quad t \in [t_0, t_1]$$

whereat E_i are Lebesgue measurable subsets of $[t_0, t_1]$ and $u_i \in X$ for $i = 1 \dots m$

2. *A function $f : [t_0, t_1] \to X$ is called* ***strongly measurable****, if there exists a sequence of simple functions $s_k : [t_0, t_1] \to X,\ k = 1, 2, \ldots$ so that it holds:*

$$s_k(t) \to f(t) \text{ for almost all } t \in [t_0, t_1].$$

Now, analogue to the Lebesgue integral, integration of strongly measurable functions can be specified.

Definition 10. *1. For a simple function $s : [t_0, t_1] \to X$ it is defined:*

$$\int_{[t_0,t_1]} s(t)dt := \sum_{i=1}^{m} |E_i| u_i.$$

2. *A strongly measurable function $f : [t_0, t_1] \to X$ is called* ***integrable****, if there exists a sequence $(s_k)_{k\in\mathbb{N}}$ of simple functions, so that it holds:*

$$\int_{[t_0,t_1]} \|s_k(t) - f(t)\|_X \, dt \to 0 \quad (k \to \infty)$$

3. *For integrable f we define:*

$$\int_{[t_0,t_1]} f(t)dt = \lim_{k\to\infty} \int_{[t_0,t_1]} s_k(t)dt.$$

Theorem 7 (Bochner theorem)**.** *A strongly measurable function $f : [t_0, t_1] \to X$ is integrable if and only if $t \to \|f(t)\|$ is integrable. In this case*

$$\left\| \int_{[t_0,t_1]} f(t)dt \right\|_X \leq \int_{[t_0,t_1]} \|f(t)\|dt$$

$$\left\langle u^*, \int_{[t_0,t_1]} f(t)dt \right\rangle = \int_{[t_0,t_1]} \langle u^*, f(t)\rangle dt.$$

for each $u^* \in X^*$.

See e.g. [Yosida, 1995, Chapter V] for a proof. We use this to define the following $\mathbb{L}^p$ spaces.

Definition 11. *The space*

$$\mathbb{L}^p([t_0, t_1], X)$$

consists of all strongly measurable functions $f : [t_0, t_1] \to X$ with

$$\|f\|_{\mathbb{L}^p([t_0,t_1],X)} := \left(\int_{[t_0,t_1]} \|f(t)\|^p dt \right)^{\frac{1}{p}} < \infty$$

for $1 \leq p < \infty$ and

$$\|f\|_{\mathbb{L}^\infty([t_0,t_1],X)} := \operatorname*{ess\,sup}_{t\in[t_0,t_1]} \|f(t)\| < \infty.$$

In particular the space $\mathbb{L}^2([t_0,t_1],\mathbb{L}^2(\Omega))$ is a Hilbert space with norm

$$\|f\|_{\mathbb{L}^2([t_0,t_1],\mathbb{L}^2(\Omega))} = \left(\int_{[t_0,t_1]} \|f(t)\|^2_{\mathbb{L}^2(\Omega)} dt\right)^{\frac{1}{2}} = \left(\int_{[t_0,t_1]} \left(\int_\Omega f(t,x)^2 dx\right) dt\right)^{\frac{1}{2}}$$

and inner product

$$\langle f,g\rangle_{\mathbb{L}^2([t_0,t_1],\mathbb{L}^2(\Omega))} = \int_{[t_0,t_1]} \langle f(t),g(t)\rangle_{\mathbb{L}^2(\Omega)}\ dt.$$

2.7 Functional analysis

Essential for the computation of coherent sets is the *Courant-Fischer theorem*, that gives a variational characterization of eigenvalues of (in our case) compact, self-adjoint operators on Hilbert spaces.
We consider a compact, self-adjoint operator $\mathcal{H} : \mathbb{L}^2(\Omega,\mu) \to \mathbb{L}^2(\Omega,\mu)$. $\mathcal{H}$ then exhibits eigenvalues $\ldots \leq \lambda_n \leq \ldots \leq \lambda_2 \leq \lambda_1$ counted according to their multiplicity. The corresponding set of μ-orthonormal eigenvectors, associated to the n largest eigenvalues, is denoted by $\{v_n,\ldots,v_1\}$. Recall, that by the Rayleigh principle (see e.g. [Huisinga and Schmidt, 2006]), the $k-th$ largest eigenvalue λ_k of $\mathcal{H}$ is given by the variational formula

$$\lambda_k = \max\left\{\langle \mathcal{H}w,w\rangle_\mu : w \in \mathbb{L}^2(\Omega,\mu),\ \|w\|_\mu = 1,\ w \perp v_1,\ldots,v_{k-1}\right\}$$

where $\perp$ denotes orthogonality with respect to the inner product $\langle\cdot,\cdot,\rangle_\mu$.
We consider a finite dimensional subspace $U \subset \mathbb{L}^2(\Omega,\mu)$ with orthonormal basis $\varphi_1,\ldots,\varphi_n$, $\dim(U)= n$. Then for a self adjoint operator $\mathcal{H}$ on $\mathbb{L}^2(\Omega,\mu)$ the *Rayleigh trace* is defined as

$$\mathrm{Tr}_U\,\mathcal{H} = \sum_{i=1}^n \langle \mathcal{H}\varphi_i,\varphi_i\rangle_\mu.$$

Note, that this choice is independent of the particular choice of the orthonormal basis. We now state the Courant-Fischer theorem and give a short proof, following [Huisinga and Schmidt, 2006].

Theorem 8 (Courant-Fischer theorem, [Huisinga and Schmidt, 2006, Theorem 1])**.** *Assume that $\mathcal{H} : \mathbb{L}^2(\Omega,\mu) \to \mathbb{L}^2(\Omega,\mu)$ is a self adjoint, compact operator. Then*

$$\begin{aligned}\sum_{i=1}^n \lambda_k &= \max\left\{\mathrm{Tr}_U\,\mathcal{H} :\ U \textit{ n-dim subspace of } \mathbb{L}^2(\Omega,\mu)\right\}\\ &= \max\left\{\sum_{i=1}^n \langle \mathcal{H}\varphi_i,\varphi_i\rangle_\mu : (\varphi_1,\ldots,\varphi_n) \textit{ is orthonormal system in } \mathbb{L}^2(\Omega,\mu)\right\}\end{aligned}$$

Proof. The second equality is clear. For $k \leq n$ denote by v_k the normalized eigenvector of $\mathcal{H}$ corresponding to λ_k. Set $\varphi_i = v_i$, then

$$\langle \mathcal{H}\varphi_i, \varphi_i \rangle_\mu = \langle \mathcal{H}v_i, v_i \rangle_\mu = \lambda_i,$$

and therefore

$$\sum_{i=1}^{n} \lambda_k \leq \max \left\{ \mathrm{Tr}_U \, \mathcal{H} : \; U \text{ } n\text{-dim subspace of } \mathbb{L}^2(\Omega, \mu) \right\}.$$

Now let U be an arbitrary n-dimensional subspace. We choose w_n to be a normalized vector $w_n \in U$ orthogonal to span $\{v_1, \ldots, v_{n-1}\}$. We now inductively define $w_{n-1}, \ldots, w_1$:

If $w_n, \ldots, w_{k+1}$ have been defined, we choose w_k to be a normalized vector in the k-dimensional subspace $U \cap \text{span}\{w_{k+1}, \ldots, w_n\}^\perp$, which is additionally perpendicular to span $\{v_1, \ldots, v_{k-1}\}$.
Like that, for the constructed vectors $w_k, \; k = 1, \ldots, n$ holds, that $w_k \perp \text{span}\{v_1, \ldots, v_{k-1}\}$. Hence, with the Rayleigh principle, $\lambda_k \geq \langle \mathcal{H}w_k, w_k \rangle_\mu$, and thus

$$\sum_{i=1}^{n} \lambda_i \geq \sum_{i=1}^{n} \langle \mathcal{H}w_k, w_k \rangle_\mu = \mathrm{Tr}_U \, \mathcal{H}.$$

□

Let $\mathcal{L} : \mathbb{L}^2(\Omega, \mu) \to \mathbb{L}^2(\Omega, \mu)$ be a compact operator. Then $\mathcal{L}^*\mathcal{L}$ is self adjoint and compact, and the k-th singular value of $\sigma_k(\mathcal{L}) = \lambda_k(\mathcal{L}^*\mathcal{L})^{\frac{1}{2}}$. As the Rayleigh principle also holds for singular values

$$\sigma_k(\mathcal{L}) = \max \left\{ \langle \mathcal{L}^*\mathcal{L}w, w \rangle_\mu^{\frac{1}{2}} : w \in \mathbb{L}^2(\Omega, \mu), \; \|w\|_\mu = 1, \; w \perp v_1, \ldots, v_{k-1} \right\},$$

where $v_1, \ldots, v_{k-1}$ are the corresponding right singular vectors of $\mathcal{L}$, we can directly state a version of the Courant-Fischer theorem for singular values:

Corollary 5 (Courant-Fischer theorem for singular values)**.** *Assume that $\mathcal{L} : \mathbb{L}^2(\Omega, \mu) \to \mathbb{L}^2(\Omega, \mu)$ is a compact operator. Then*

$$\sum_{i=1}^{n} \sigma_k = \max \left\{ \sum_{i=1}^{n} \langle \mathcal{L}^*\mathcal{L}\varphi_i, \varphi_i \rangle_\mu^{\frac{1}{2}} : (\varphi_1, \ldots, \varphi_n) \textit{ is orthonormal system} \right\}.$$

2.8 Heuristic clustering

During this thesis, we will encounter a classic problem in computational geometry:

We want to compute $k \in \mathbb{N}$ clusters within a given data set $\mathcal{X} \subset \mathbb{R}^d$.

To this end, we identify each cluster with its center $c \in \mathbb{R}^d$, and assign each point to the *center* closest to it (with respect to a given metric). Like that we can restate the problem as:

Given an integer k and a set of n data points in $\mathbb{R}^d$, we aim at choosing k centers so as to minimize a given cost function ϕ.

The cost function ϕ is usually the total squared distance between each point and its closest center. Solving this problem exactly is NP-hard. To nevertheless get an approximate solution to this problem, we shortly introduce a widely used, heuristic clustering algorithm introduced in [Lloyd, 1982] and usually referred to as *k-means*.
We wish to choose the k centers $c \in \mathcal{C} \subset \mathbb{R}^d$ so as to minimize

$$\phi = \sum_{x \in \mathcal{X}} \min_{c \in \mathcal{C}} \|x - c\|_2^2.$$

From these centers, we can simply define a clustering by grouping data points according to which center each point is closest to (in euclidian distance). The k-means algorithm is a two-phase iterative algorithm to minimize the sum of point-to-centroid distances, summed over all k clusters:

1. Arbitrarily choose k initial centers $\mathcal{C} = \{c_1, \ldots, c_k\}$.
2. For each $i \in \{1, \ldots, k\}$, set the cluster C_i to be the set of points in $\mathcal{X}$ that are closer to c_i than they are to c_j for all $j \neq i$.
3. For each $i \in \{1, \ldots, k\}$, set c_i to be the center of mass of all points in C_i:
$$c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.$$
4. Repeat Steps 2 and 3 until C no longer changes.

In its original form the initial centers are chosen uniformly at random in $\mathcal{X}$. In step 2, ties may be broken arbitrarily. Very common nowadays is a version with a modified initialization step 1, called *k-means++*, see [Arthur and Vassilvitskii, 2007]:
Let $D(x)$ denote the shortest distance from a data point to the closest center already chosen. Then, step 1 is replaced by

1a. Take one center c_1, chosen uniformly at random from $\mathcal{X}$.

1b. Take a new center c_i, choosing $x \in \mathcal{X}$ with probability
$$\frac{D(x)^2}{\sum_{x \in \mathcal{X}} D(x)^2}.$$

1c. Repeat Step 1b. until all k centers are chosen.

With this modification the algorithm both, speed and the accuracy are improved [Arthur and Vassilvitskii, 2007].

As k-means(++) massively depends on the chosen initial centers, k-means(++) is usually run several times and the result with the lowest within cluster sums of point to centroid distances $\sum_{i=1:k} \sum_{x \in C_i} \|x - c_i\|_2$ is taken. k-means++ is implemented in Matlab. The Matlab command

```
C=kmeans(X, k, 'Replicates', 500);
```

calls the kmeans++ algorithm for the data matrix $X \in \mathbb{R}^{n \times d}$, performs it 500 times, and chooses the result with the lowest within cluster sums.

2.9 Some basics of plasma physics

A plasma is usually described as the fourth fundamental state of matter. The first three states are solid, liquid and gas, which are separated by the strength of the bonds that hold their constituent particles together. These binding forces are relatively strong in a solid, weak in a liquid, and essentially almost absent in the gaseous state. Depending on pressure, a substance can transform from solid via fluid to gaseous state with increasing thermal energy of its atoms or molecules, i.e. with increasing temperature. If the substance is heated up further, an increasing fraction of atoms possesses enough kinetic energy to overcome, by collisions, the binding energy of the outermost orbital electrons. The result is a (fully) ionized gas, also known as *plasma*.

In this section we derive two basic sets of equations describing the particle dynamics of a plasma, the *Vlasov-Maxwell equations* and the *Vlasov-Poisson equations*, see e.g. [Bittencourt, 2013]. First we consider the *Lorentz force law* and the *Maxwell equations*, which together form the foundation of classical electrodynamics. The Lorentz force describes the combination of electric and magnetic force on a point charge due to electromagnetic fields: If a particle of charge q moves with velocity $\mathbf{v}$ in the presence of an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$, then it will experience a force

$$\mathbf{F} = q\left[\mathbf{E} + (\mathbf{v} \times \mathbf{B})\right]. \tag{2.21}$$

A positively charged particle hence will be accelerated in the same linear orientation as the $\mathbf{E}$ field and will curve perpendicularly to both the instantaneous velocity vector v and the magnetic field $\mathbf{B}$ according to the right-hand rule. The term $q\mathbf{E}$ is called the electric force, while the term $q\mathbf{v} \times \mathbf{B}$ is called the magnetic force.

The (macroscopic) Maxwell equations describe how the electric and magnetic fields are generated and altered by each other and by charges and currents. They read

$$\begin{aligned} \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{B} &= \mu_0 \mathbf{I} + \mu_0 \epsilon_0 \partial_t \mathbf{E} \\ \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0, \end{aligned} \tag{2.22}$$

where ρ and $\mathbf{I}$ describe the electric charge density and the current density, respectively. The appearing universal constants are the permittivity of free space, ϵ_0, and the permeability of free space, μ_0.

Each of the four Maxwell equations describes a physical concept:

Faraday's law

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B}$$

describes how a time varying magnetic field induces an electric field. This dynamically induced electric field has closed field lines just as the magnetic field, if not superposed by a static (charge induced) electric field.

Ampere's law

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{I} + \mu_0 \epsilon_0 \partial_t \mathbf{E}$$

states that magnetic fields can be generated either by electric current (Ampere) or by changing electric fields (Maxwell's addition).

Gauss's Law

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$

describes the relationship between a static electric field and the electric charges that cause it. The static electric field points away from positive charges and towards negative charges.

Gauss's law for magnetism

$$\nabla \cdot \mathbf{B} = 0$$

states that there are no magnetic charges, analogous to electric charges.

Poisson equation

The Poisson equation is obtained from the Maxwell equations when the electric and magnetic fields are not, or only very little, time dependent. In this case we obtain the stationary Maxwell equations

$$\begin{aligned}\nabla \times \mathbf{E} &= 0\\ \nabla \times \mathbf{B} &= \mu_0 I\\ \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0}\\ \nabla \cdot \mathbf{B} &= 0.\end{aligned}$$

Here the electric and magnetic fields are decoupled. Furthermore in many cases, because $\mathbf{B}$ or its contribution to the Lorentz force is small we only consider the equations for the electric field

$$\begin{aligned}\nabla \times \mathbf{E} &= 0\\ \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0}.\end{aligned}$$

The first equation implies that that the electric field $\mathbf{E}$ is rotation free and hence there exists a scalar potential $\Phi : \Omega \to \mathbb{R}$ such that $\mathbf{E} = -\nabla\Phi$ so that the Maxwell equations simplify to the *Poisson equation*

$$\Delta\Phi = \frac{\rho}{\epsilon_0}. \tag{2.23}$$

Vlasov equation

In a *kinetic model*, each particle species s in the plasma is characterized by a distribution function $f_s(t, \mathbf{x}, \mathbf{v})$ which corresponds to a statistical mean of the repartition of particles in phase space for a large number of realizations of the considered physical system. This distribution function f_s contains more information than a fluid description as it also includes information on the distributions of particle velocities at each position. In the non-relativistic case assuming that collective effects are dominant on Coulomb (binary) collisions between particles, the kinetic equation for the plasma is the Vlasov equation which reads

$$\frac{df(t, \mathbf{x}, \mathbf{v})}{dt} = 0,$$

which can be reformulated as

$$\begin{aligned}\frac{df}{dt} &= f_t + f_x\dot{\mathbf{x}} + f_v\dot{\mathbf{v}}\\ &= f_t + f_x\mathbf{v} + \frac{\mathbf{F}}{m}f_v\\ &= f_t + f_x\mathbf{v} + \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B})f_v = 0,\end{aligned}$$

where we used *Newton's second law* $\mathbf{F} = m \cdot \dot{\mathbf{v}}$ and the Lorentz equation (2.21). This equation expresses that the distribution function f_s is conserved along the trajectories of the particles which are determined by the mean electric field. The Vlasov equation, when it takes into account the self-consistent electromagnetic field generated by the particles, is coupled to the Maxwell equations (2.22) or the Poisson equation (2.23). Note that in the kinetic model the electric charge density and the current density can be expressed via the distribution functions of the different species of particles f_s using the relations

$$\rho = \rho(t, \mathbf{x}) = \sum_s q_s \int f(t, \mathbf{x}, \mathbf{v}) dv, \qquad \mathbf{I} = \mathbf{I}(t, \mathbf{x}) = \sum_s q_s \int f(t, \mathbf{x}, \mathbf{v}) \mathbf{v} \; d\mathbf{v},$$

where q_s is the charge of the particles species s.

If the magnetic field $\mathbf{B}$ and the electric field $\mathbf{E}$ are created by the particles we can combine the Vlasov equation and the Maxwell equations (2.22) to the *Vlasov-Maxwell system of equations*

$$\begin{aligned} f_t + f_x \mathbf{v} + \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_v &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{B} &= \mu_0 \mathbf{I} + \mu_0 \epsilon_0 \partial_t \mathbf{E} \\ \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{2.24}$$

and obtain a closed model for a plasma. If the magnetic field is negligible we can combine the Vlasov equation and the Poisson equation (2.23) to the *Vlasov-Poisson equation*

$$\begin{aligned} f_t + f_x \mathbf{v} + \frac{q}{m} \mathbf{E} f_v &= 0 \\ \Delta \Phi &= \frac{\rho}{\epsilon_0}. \end{aligned} \tag{2.25}$$

Chapter 3

Coherent structures

In this chapter, we turn to the main topic of this thesis, *coherence*, and its connection to transfer operators. First, we give a short overview over the preceding concept of metastability (almost invariance) and use this to motivate a mathematical notion of coherence for non-autonomous dynamics. We derive a numerical method based on Ulam's discretization to partition the state space of a given system into n coherent sets and apply this method on two examples: The Double gyre and the Bickley jet.

3.1 Almost invariant sets

The effective dynamics of many biomolecules is generally understood to be governed by statistically rare transitions between so-called conformations of the biomolecule. In a conformation, the large scale geometric structure of the molecule is conserved, whereas on smaller scales the system may well rotate, oscillate or fluctuate. While the microscopic state of a molecule varies quickly, a change between two conformations is a comparatively rare event. Hence, from the dynamical point of view, the subset of the state space associated with a conformation is metastable. The resulting macroscopic dynamical behavior can be described as a flipping process between the metastable subsets. Consequently, it is of interest to decompose the state space of the molecular motion into some main metastable sets and evaluate the transition probabilities between them. This is of further importance because computing the full (microscopic) system over longer timespans is infeasable [Deuflhard and Schütte, 2004]. In Figure 3.1, we show the three conformations of the molecule Butane. The conformation of the molecule mainly depends on the central dihedral angle $\theta \in [0, 2\pi)$, so the state space in this model is one-dimensional.

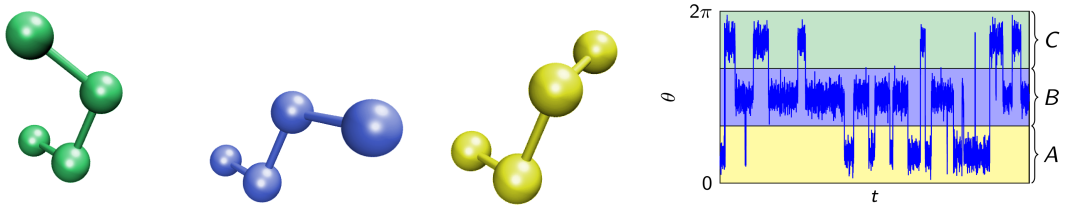

Figure 3.1: The three metastable conformations C, B, A for the molecule butane. On the right, a possible trajectory in state space is shown.

Additionally a typical trajectory in state space is shown.[1]

The setting usually used in molecular dynamics is the following: Let $X = (X)_{n\in\mathbb{N}}$ denote a homogeneous, reversible Markov chain on the state space Ω with transition kernel

$$p(x, A) = \mathbb{P}[X_1 \in A | X_0 = x]$$

for all $x \in \Omega$ and all subsets $A \subset \Omega$ contained in the $\sigma-$algebra $\mathcal{A}$. Consider a probability measure ν on Ω and assume that the Markov chain is initially distributed according to ν, i.e. $X_0 \sim \nu$, $\mathbb{P}[X_0 \in A] = \nu(A)$ for all $A \in \mathcal{A}$. The Markov chain at time $k > 0$ is then distributed via

$$\mathbb{P}[X_k \in A | X_0 \sim \nu] =: \mathbb{P}_\nu[X_k \in A] =: \nu_k(A).$$

The evolution of probability measures ν_k can be described by a transfer operator $\mathcal{P}$ acting on the space of bounded measures on $(\Omega, \mathcal{A})$ via

$$\mathcal{P}\nu(A) = \mathbb{P}_\nu[X_1 \in A] = \int_\Omega p(x, A)\ \nu(dx).$$

Assume, that μ is an invariant probability measure of $\mathcal{P}$, then $\nu_0 \ll \mu$ implies $\nu_k \ll \mu$ (see e.g. [Revuz, 2008] Chapter 4). Hence, usually $\mathcal{P}$ is considered as operator on $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ acting on probability measures, which are absolutely continuous w.r.t. μ according to

$$\int_A \mathcal{P}f(x)\mu(dx) = \int_\Omega p(x, A)f(x)\mu(dx).$$

As the Markov chain X is reversible, the transfer operator $\mathcal{P}$ is self-adjoint on $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$. In order to identify the conformations of a given molecule, the concept of almost invariant (metastable) sets in state space is introduced in [Dellnitz and Junge, 1999]:

Definition 12. *Let* $A, B \in \mathcal{A}$

[1] For a more detailed investigation of Butane and its metastable conformations, see [Bittracher, 2016, Griebel et al., 2007].

- *The transition probability is defined to be the conditional probability*

$$p(A,B) = \mathbb{P}_\mu[X_1 \in B | X_0 \in A] = \frac{1}{\mu(A)} \int_A p(x,B)\mu(dx)$$

  if $\mu(A) > 0$ *and* $p(A,B) = 0$ *if* $\mu(A) = 0$.

- *A subset is called* ***invariant*** *if* $p(A,A) = 1$.

- *A subset is called* δ***-almost invariant*** *(metastable) if* $p(A,A) = \delta$.

Usually, a set is hence considered to be *almost invariant*, if it is $\delta-$almost invariant with $\delta \approx 1$. This transition probability of approximately one is of course a vague statement. In most applications, however, one is interested in a partition of the state space into the most metastable sets. Therefore, it is not necessary to interpret "approximately one", but to determine the number of subsets we are looking for.
The connection between the notion of almost invariance and the transfer operator $\mathcal{P}$ is the following:

$$\begin{aligned} p(A,B) = \frac{1}{\mu(A)} \int_A p(x,B)\mu(dx) &= \frac{1}{\mu(A)} \int_\Omega p(x,B) 1_A(x) \mu(dx) \\ &= \frac{1}{\mu(A)} \int_B \mathcal{P} 1_A \mu(dx) = \frac{\langle \mathcal{P} 1_A, 1_B \rangle_\mu}{\mu(A)}. \end{aligned}$$

In [Huisinga and Schmidt, 2006] it is shown that under Assumption 1, Theorem 9 holds.

Assumption 1. *The transfer operator* $\mathcal{P}$ *is self-adjoint and exhibits* n *eigenvalues* $\lambda_n \leq \ldots \leq \lambda_2 < \lambda_1 = 1$ *counted according to their multiplicity. The corresponding set of* n *orthonormal eigenvectors are denoted by* $\{v_n, \ldots, v_1\}$. *Furthermore the spectrum* $\sigma(\mathcal{P})$ *satisfies*

$$\sigma(\mathcal{P}) \subset [a,b] \cup \{\lambda_n \leq \ldots \leq \lambda_2 < \lambda_1\}$$

for some constants $a,\ b \in (-1,1)$ *satisfying* $-1 < a \leq b < \lambda_n$. *In this sense, the eigenvalues* $\lambda_1, \ldots, \lambda_n$ *are called* ***dominant****.*

Theorem 9 ([Huisinga and Schmidt, 2006])**.** *Consider a reversible homogeneous continuous-time Markov process* $X = (X_t)_{t \in [0,\infty)}$ *and its corresponding semigroup of transfer operators* $\mathcal{P}^t : \mathbb{L}^2(\Omega, \mathcal{A}, \mu) \to \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$. *If* $\mathcal{P} = \mathcal{P}^{t_1}$ *satisfies Assumption 1 for some fixed time* $t_1 > 0$, *then the metastability of an arbitrary decomposition* $D = \{A_1, \ldots, A_n\}$ *of the state space can be bounded from above by*

$$p(A_1,A_1) + p(A_2,A_2) + \ldots + p(A_n,A_n) \leq 1 + \lambda_2 + \ldots \lambda_n$$

and bounded from below by

$$1 + \rho_2\lambda_2 + \ldots \rho_n\lambda_n \leq p(A_1,A_1) + \ldots + p(A_n,A_n)$$

where $\rho_j = \|Qv_j\|^2$ *and* λ_k *denote eigenvalues of* $\mathcal{P}$.

Here

$$Q : \mathbb{L}^2(\Omega, \mathcal{A}, \mu) \longrightarrow \text{span}(1_{A_1}, \ldots, 1_{A_n})$$
$$Qf = \sum_{k=1}^{n} \langle f, \chi_{A_k} \rangle \chi_{A_k}$$

where $\chi_{A_k} = 1_{A_k}/\|1_{A_k}\|_2,\ k = 1 \ldots n$, is the orthogonal projection from $\mathbb{L}^2$ onto the span of the indicator functions of the metastable sets.
In other words, the lower bound depends on how close the eigenfunctions of $\mathcal{P}$ are to indicator functions of metastable sets. Via a spectral analysis of the transfer operator $\mathcal{P}$ we can find the metastable sets for a given molecule: If we can identify a gap in the (eigen-) spectrum of the transfer operator, this tells us how many metastable sets exist in the system. The first n eigenfunctions then are approximations to linear combinations of the indicator functions of the n metastable sets.

3.2 Coherent sets

After having examined the concept of metastability in molecule dynamics, we turn to the main topic of this thesis: The analysis of complex non-autonomous systems, such as fluids (see Chapters 3, 4, 5) or plasmas (see Chapter 6), via finding partitions into meaningful high level structures, so called *coherent sets*.

Intuitively spoken, coherent sets are subsets of the domain Ω that are surrounded by a strong, albeit leaky transport barrier. As a result, they keep most of their mass while being moved around by the flow $\mathcal{T}$ over a finite time interval $[t_0, t_1]$. They have a massive influence on the behavior of the system, and usually a regular surface shape.

This setting is slightly different than for metastable sets. The dynamics are usually deterministic. As they might be non-autonomous, we are interested in a finite time interval. Furthermore, we are also interested in the shape of a coherent set.

To get a first mathematical grasp, we proceed analogously to the pioneering work in [Froyland et al., 2010b], where the notion of coherent sets was first connected with transfer operators. To keep the computations simple, we consider a dynamical system induced by a smooth vector field $b : [t_0, t_1] \times \Omega \to \Omega$ and its flow map $\mathcal{T}^{t_0,t_1}$ from initial time $t_0 = 0$ to final time t_1. Inspired by the associated notion of a metastable set, we start formalizing this request by asking for a pair of sets $A_0,\ A_1 \subset \Omega$, such that A_0 will approximately be carried to A_1 by $\mathcal{T}^{t_0,t_1}$ in the sense that

$$\rho(A_0, A_1) = \frac{\mu(A_0 \cap \mathcal{T}^{t_1,t_0}(A_1))}{\mu(A_0)} \approx 1. \tag{3.1}$$

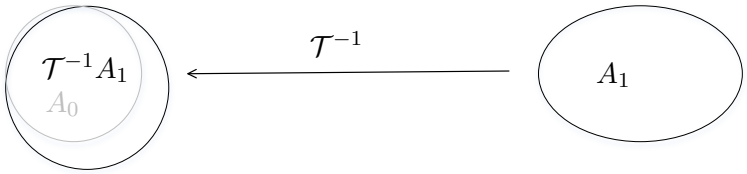

Figure 3.2: The preimage of the set A_1 is almost equal to the set A_0.

In other words, a particle starting in the set A_0 at time $t_0 = 0$ is very likely to end up in the set A_1, see Figure 3.2. A_0 and A_1 in the end will represent the coherent set A at times t_0 and t_1, respectively. Evidently, with $A_1 = \mathcal{T}^{t_0,t_1}(A_0)$, we obtain $\rho(A_0, A_1) = 1$ for any $A_0 \subset \Omega$. This is not yet a well defined problem. In fact, (3.1) does not impose any condition on the geometries of the sets A_0 and A_1. In particular, the image set $A_1 = \mathcal{T}^{t_0,t_1}(A_0)$ might be stretched and folded all over the domain Ω – but this is not the type of coherent set we have in mind. Froyland observed that $\rho(A_0, \mathcal{T}^{t_0,t_1}(A_0))$ is not close to 1 *for every set* A_0 any more, as soon as some random perturbation is artificially added to the dynamics (cf. [Dellnitz and Junge, 1999, Dellnitz et al., 2001] for related ideas in the autonomous context). This perturbation can be realized via e.g. diffusion.

3.3 Computing coherent n-partitions

With this first mathematical intuition of coherence, we next want to compute a *coherent partition* of the domain $\Omega \subset \mathbb{R}^d$ into n *coherent sets*. We never consider a coherent set on its own, but always a partition, because our problem formulation is intrinsically symmetric. Mathematically we want to partition Ω into n coherent sets $A_j^k,\ k = 1, \dots n$, at times $t_j = t_0, t_1$, $\Omega = \dot{\bigcup} A_0^k = \dot{\bigcup} A_1^k$, such that

1. $\mathcal{P}^{t_0,t_1} 1_{A_0^k} \approx 1_{A_1^k}\ , k = 1, \dots n$
2. $\mu(A_1^k) = \mu(A_0^k),\ k = 1, \dots n$
3. $\mathcal{P}^{t_0,t} A_0^k,\ t \in [t_0, t_1],\ k = 1 \dots n$, are stable under small random perturbations.

Condition 1 ensures, that A_1^k is approximately the image of A_0^k under the dynamics. Condition 2 means the coherent sets A_0^k do not lose mass at all, while being mapped to A_1^k. This is a very strong claim, which will be relaxed later. The stability under small perturbations in Condition 3 helps to prefer nicely shaped sets, i.e. sets that do not become filaments: Ill-shaped sets have a longer boundary and, hence, loose more mass due to the perturbation. To phrase our problem mathematically, we consider a symmetric version of (3.1). For a clearer representation let $\mathcal{P} := \mathcal{P}^{t_0,t_1}$. As the number n of those sets is not known a priori, we try to find a formulation that reveals the partition

into n coherent sets for several $n \geq 1$ and choose a meaningful number from this. To this end, we formulate the optimization problem

$$\max_{\Omega=\dot{\bigcup} A_0^k,\ \mu(A_0^k)=\nu(A_1^k)} \sum_{k=1}^{n} \frac{\mu(A_0^k \cap \mathcal{T}^{t_1,t_0} A_1^k)}{\mu(A_0^k)}, \tag{3.2}$$

which we will interpret as a problem in $\mathbb{L}^2(\Omega,\mu)$. We demand the coherent sets A_k to form a partition of the domain and to keep their mass during the evolution. ν denotes the evolution of the reference measure ν, i.e. $\nu = \mu \circ \mathcal{T}^{t_1,t_0}$. The connection between (3.2) and the Frobenius-Perron operator $\tilde{\mathcal{P}} : \mathbb{L}^2(\Omega,\mu) \to \mathbb{L}^2(\Omega,\nu)$ is given via

$$\begin{aligned}\mu(A_0^k \cap \mathcal{T}^{t_1,t_0} A_1^k) &= \int_{\mathcal{T}^{t_1,t_0}(A_1^k)} 1_{A_0^k}\, d\mu = \int_{A_1^k} \tilde{\mathcal{P}} 1_{A_0^k}\, d\nu \\ &= \langle \tilde{\mathcal{P}} 1_{A_0^k}, 1_{A_1^k} \rangle_\nu,\end{aligned}$$

see (2.5). We can reformulate (3.2) as

$$\begin{aligned}&\max_{\Omega=\dot{\bigcup} A_0^k,\ \mu(A_0^k)=\nu(A_1^k)} \sum_{k=1}^{n} \frac{\mu(A_0^k \cap \mathcal{T}^{t_1,t_0} A_1^k)}{\mu(A_0^k)} \\ =& \max_{\Omega=\dot{\bigcup} A_k,\ \mu(A_0^k)=\nu(A_1^k)} \sum_{k=1}^{n} \frac{\langle \tilde{\mathcal{P}} 1_{A_0^k}, 1_{A_1^k} \rangle_\nu}{\|1_{A_0^k}\|_\mu\ \|1_{A_1^k}\|_\nu} \\ \leq& \max_{f_1 \perp \cdots \perp f_n \in \mathbb{L}^2(\Omega,\mu),\ g_k \in \mathbb{L}^2(\Omega,\nu)} \sum_{k=1}^{n} \frac{\langle \tilde{\mathcal{P}} f_k, g_k \rangle_\nu}{\|f_k\|_\mu\ \|g_k\|_\nu} \\ =& \max_{f_1 \perp \cdots \perp f_n,\ g_k} \sum_{k=1}^{n} \frac{\langle \tilde{\mathcal{P}} f_k, g_k/\|g_k\|_\nu \rangle_\nu}{\|f_k\|_\mu} \\ =& \max_{f_1 \perp \cdots \perp f_n} \sum_{k=1}^{n} \frac{\langle \tilde{\mathcal{P}} f_k, \tilde{\mathcal{P}} f_k \rangle_\nu^{\frac{1}{2}}}{\|f_k\|_\mu},\end{aligned}$$

where we simply used the Cauchy Schwartz inequality,

$$\begin{aligned}&\max_{f_1 \perp \cdots \perp f_n} \sum_{k=1}^{n} \frac{\langle \tilde{\mathcal{P}} f_k, \tilde{\mathcal{P}} f_k \rangle_\nu^{\frac{1}{2}}}{\|f_k\|_\mu} \\ =& \max_{f_1,\ldots,f_n \text{ ONB of } V_n \leq \mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \|\tilde{\mathcal{P}} f_k\|_\nu \\ =& \max_{f_1,\ldots,f_n \text{ ONB of } V_n \leq \mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \langle \tilde{\mathcal{P}}^* \tilde{\mathcal{P}} f_k, f_k \rangle_\mu^{\frac{1}{2}}.\end{aligned} \tag{3.3}$$

We have not yet considered the demand for stability under small random perturbations (which, e.g. can be realized via diffusion). To this end, we add some suitable small

random perturbation to the problem. We obtain a slightly perturbed transfer operator $\tilde{\mathcal{P}}_\epsilon$ for the slightly perturbed dynamics $\mathcal{T}_\epsilon^{t_0,t_1}$. We can compute[2]

$$\begin{aligned} &\approx \max_{f_1,\dots,f_n \text{ ONB of } V_n \leq \mathbb{L}^2(\Omega)} \sum_{k=1}^{n} \langle \tilde{\mathcal{P}}_\epsilon^* \tilde{\mathcal{P}}_\epsilon f_k, f_k \rangle_\mu^{\frac{1}{2}} \\ &= \sum_{k=1}^{n} \lambda_k (\tilde{\mathcal{P}}_\epsilon^* \tilde{\mathcal{P}}_\epsilon)^{\frac{1}{2}} \\ &= \sum_{k=1}^{n} \sigma_k (\tilde{\mathcal{P}}_\epsilon) \end{aligned} \tag{3.4}$$

due to the *Courant-Fischer theorem* 5. In order to apply this theorem, we have to assume that the diffusion added to the problem makes the transfer operator $\tilde{\mathcal{P}}_\epsilon$ compact. We do not need to assume the leading singular value is isolated. The maximizing functions are the corresponding first n singular functions of $\tilde{\mathcal{P}}_\epsilon$. As a basis of the vector space $V_n = \text{span}(v_1, \dots, v_n)$ is not unique, the computed singular functions $v_1, \dots, v_n$ are not necessarily approximations to the indicator functions of the coherent sets $1_{A_0^1}, \dots, 1_{A_0^n}$. They are linear combination of these, if our relaxation step from indicator functions to functions in $\mathbb{L}^2$ was not too strong. To "change" the basis to indicator functions and actually extract the coherent partition, some heuristic algorithms like k-means (see Section 2.8) can be used (see also [Banisch and Koltai, 2016], [Hadjighasem et al., 2016]). The question of how many sets a partition should contain is answered by the singular spectrum of $\tilde{\mathcal{P}}_\epsilon$, analogous to [Huisinga and Schmidt, 2006, Proposition 2.3]. For a partition into n coherent sets should hold:

1. the first n singular values are close to one,
2. $v_1, \dots, v_n$ are good approximations of linear combinations of n(!) indicator functions partitioning Ω.

The first condition ensures that the partition is actually coherent. The second is an a posteriori check of the relaxation to functions in $\mathbb{L}^2$ was not strong. Useful choices for n are usually encoded in the spectrum of $\tilde{\mathcal{P}}_\epsilon$, in which gaps occur showing where to cut the spectrum, see [Huisinga and Schmidt, 2006]. In Chapter 4, we will formalize this request and give a lower bound on the approximation of the singular values to the coherence ratios.

3.4 Time-discrete diffusion and Ulam's method

In the last Section 3.2, we asserted that we can compute an approximation to (3.2) via a spectral analysis of the slightly perturbed operator $\tilde{\mathcal{P}}_\epsilon$, which is rendered compact by a small diffusion. We now have to do two things:

[2] Here we make a modeling choice and we have only an approximate equality in this step. However, for application purposes this is good enough. In chapter 4 we will establish a rigorous connection here.

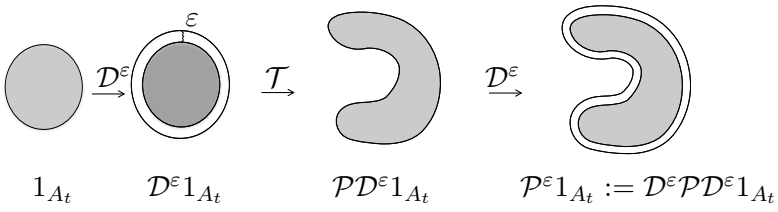

Figure 3.3: Perturbed operator $\mathcal{P}^\epsilon$ via time-discrete diffusion.

- Choose a diffusion such that $\tilde{\mathcal{P}}_\epsilon$ is compact,
- Discretize the operator $\tilde{\mathcal{P}}_\epsilon$.

We first assume $\mu = \lambda$, the Lebesgue measure, to be invariant under $\mathcal{T}^{t_0,t_1}$. In this case $\tilde{\mathcal{P}}_\epsilon = \mathcal{P}_\epsilon$. In [Froyland, 2013] it is observed, that adding a small spatial diffusion at initial and final time is enough to make the such perturbed transfer operator $\mathcal{P}_\epsilon$ compact, see Figure 3.3. To achieve this, we add a random perturbation to the flow map $\mathcal{T}^{t_0,t_1}$ by a diffusion operator

$$\mathcal{D}_\epsilon : \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega) \qquad \mathcal{D}_\epsilon f(y) = \int_\Omega \alpha_\epsilon(y-x) f(x) dx,$$

where $\alpha_\epsilon : \Omega \to [0,\infty)$ is a bounded kernel with $\int_\Omega \alpha_\epsilon(y-x)dx = 1$ for all $y \in \Omega$. With the Frobenius-Perron operator $\mathcal{P}$ and the perturbation operator $\mathcal{D}_\epsilon$, we define the disturbed transfer operator

$$\mathcal{P}_\epsilon : \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega) \qquad \mathcal{P}_\epsilon := \mathcal{D}_\epsilon \mathcal{P} \mathcal{D}_\epsilon,$$

see also Figure 3.3. Note that $\mathcal{P}_\epsilon$ is still stochastic, i.e. $\mathcal{P}_\epsilon 1_\Omega = 1_\Omega$. Further, as α_ϵ is bounded, $\mathcal{P}_\epsilon$ is a Hilbert-Schmidt operator. Hence, $\mathcal{P}_\epsilon$ is compact and suited for the computation of coherent sets. In Section 2.4, we introduced an accessible discretization of the Frobenius-Perron operator, *Ulam's method*, as a Galerkin-Projection onto the indicator functions of boxes $B_1, \dots, B_N$, $\Omega = \bigcup_{i=1}^N B_i$. The approximate Frobenius-Perron operator $P \in \mathbb{R}^{N\times N}$ then reads

$$P^{ij} = \frac{\lambda(B_j \cap \mathcal{T}^{t_1,t_0}(B_i))}{\lambda(B_j)}. \tag{3.5}$$

In Section 2.5 we already established an interpretation of Ulam's method as stochastic process. Here we interpret the method as two stochastic perturbations (in space) at initial and final time. To numerically estimate (3.5), one may use a Monte Carlo method and sample test points x_{i_k}, $k = 1, \dots, K$ uniformly distributed over B_i and then compute $\mathcal{T}^{t_0,t_1}(x_{i_k})$ and count how many fall in B_j. That is

$$P^{ij} \approx \frac{\#(k : \mathcal{T}^{t_0,t_1}(x_{i_k}) \in B_j)}{K}. \tag{3.6}$$

As the indicator functions of the boxes can be interpreted as density functions of uniform distributions on boxes, the discretization of the Frobenius-Perron operator via Ulam's method usually induces enough numerical dissipation to regularize the problem. In this case we do not need to add explicit diffusion in order to compute coherent sets. Once the approximate Frobenius-Perron operator $P = \mathcal{P}_\epsilon \in \mathbb{R}^{N\times N}$ is computed, the spectral analysis

$$P = USV' = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_N \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{pmatrix} \begin{pmatrix} - & v_1' & - \\ & \vdots & \\ - & v_N' & - \end{pmatrix},$$

where $u_1, \dots, u_N$ are the left singular values, $v_1, \dots, v_N$ are the right singular values, and $\sigma_1, \dots \sigma_N$ are the singular values, can be performed by standard methods. One problem with this approach is, that Ulam's method does not conserve the stochasticity of the Frobenius-Perron operator, i.e. $P1 \neq 1$ in general, even if $\mathcal{P}1_\Omega = 1_\Omega$. The discretized transfer operator P is column stochastic, but not row stochastic, see (2.19). As a result the leading singular value is only approximately equal to 1. In the next section we introduce a way to deal with that.

3.4.1 Ulam's method for varying domains and reference measures

We now take a look at how to compute an approximation to $\tilde{\mathcal{P}}$ via Ulam's method. This means we want to compute coherent sets with respect to a reference measure μ, that is not invariant under $\mathcal{T}^{t_0,t_1}$. We also directly cover the situation, in which the flow does not stay in the initial domain Ω_1 but evolves to a different domain Ω_2. We therefore not consider a flow $\mathcal{T}^{t_0,t_1} : (\Omega, \lambda) \circlearrowleft$, but a still non-singular flow $\mathcal{T}^{t_0,t_1} : (\Omega_1, \mu) \to (\Omega_2, \nu)$, where $\nu = \mu \circ \mathcal{T}^{t_1,t_0}$ is the image of μ under the dynamics. We assume both μ and ν to be absolutely continuous with respect to the Lebesgue measure λ, and μ and ν to have Radon-Nikodym derivatives h_μ and $h_\nu = \mathcal{P}h_\mu$, respectively. Hence we have to consider the operator

$$\begin{aligned} \tilde{\mathcal{P}} &: \mathbb{L}^2(\Omega_1, \mu) \to \mathbb{L}^2(\Omega_2, \nu) \\ \tilde{\mathcal{P}}f &= \frac{\mathcal{P}(fh_\mu)}{h_\nu}, \end{aligned}$$

see Section 3.3. Via defining $\tilde{\mathcal{P}}$ like that, we obtain again a stochastic operator in the sense that $\tilde{P}1_{\Omega_1} = \mathcal{P}(h_\mu 1_{\Omega_1})/h_\nu = 1_{\Omega_2}$..

For discretizing the problem we partition the domain Ω_1 into N_1 boxes $B_1, \dots, B_{N_1}$, $\Omega = \bigcup_{i=1}^{N_1} B_i$ and the domain Ω_2 into N_2 boxes $C_1, \dots, C_{N_2}$, $\Omega = \bigcup_{i=1}^{N_2} C_i$. We then project the initial density h_μ to the approximation space V_{N_1}. V_{N_1} is the space spanned by the indicator functions of the boxes $B_1, \dots, B_{N_1}$, V_{N_2} is the space spanned by the indicator functions of the boxes $C_1, \dots, C_{N_2}$. h_μ is hence represented by a vector $p \in \mathbb{R}^{N_1}$. An

approximation q to h_ν is computed via $q = Pp \in \mathbb{R}^{N_2}$. Approximations to the inner products $\langle\cdot,\cdot\rangle_\mu$ and $\langle\cdot,\cdot\rangle_\nu$ hence read

$$\langle x, y\rangle_p = \langle \Pi_p^{\frac{1}{2}} x, \Pi_p^{\frac{1}{2}} y\rangle_2 \qquad \forall x, y \in \mathbb{R}^{N_1},$$
$$\langle x, y\rangle_q = \langle \Pi_q^{\frac{1}{2}} x, \Pi_q^{\frac{1}{2}} y\rangle_2 \qquad \forall x, y \in \mathbb{R}^{N_2}.$$

$\Pi_p^{\frac{1}{2}} = \mathrm{diag}(p)$ and $\Pi_q^{\frac{1}{2}} = \mathrm{diag}(q)$ are diagonal matrices with p, q on their diagonals, respectively. $\langle\cdot,\cdot\rangle_2$ denotes the Euclidean inner product:

$$\langle x,y\rangle_2 = \sum_{k=1}^{N} x_i \cdot y_i \qquad \forall x, y \in \mathbb{R}^N.$$

We can compute an approximation to $\tilde{\mathcal{P}}$ via Ulam's method as

$$\tilde{P} : (\mathbb{R}^{N_1}, \langle\cdot,\cdot\rangle_p) \to (\mathbb{R}^{N_2}, \langle\cdot,\cdot\rangle_q)$$
$$\tilde{P} := \Pi_q^{-1} P \Pi_p.$$

We then have to compute the singular value decomposition of $\tilde{P}$. The singular values of $\tilde{P}$ are the eigenvalues of $\tilde{P}^*\tilde{P}$. We note that the adjoint $\tilde{P}^*$ of $\tilde{P}$ (with respect to the inner products $\langle .,.\rangle_p$, $\langle .,.\rangle_q$) is not $\tilde{P}^T$, the transposed matrix of $\tilde{P}$, but the Koopman operator, which is the adjoint of P with respect to the standard 2-norm, i.e. P^T. A short computation shows, that the connections are:

$$\langle Pf, g\rangle_2 = \langle f, Kg\rangle_2,$$

and as

$$\langle \tilde{P}f, g\rangle_q = \langle \Pi_q^{-1} P \Pi_p f, g\rangle_q = \langle P\Pi_p f, g\rangle_2 = \langle \Pi_p f, Kg\rangle_2 = \langle f, Kg\rangle_p$$

we have

$$\langle \tilde{P}f, g\rangle_q = \langle f, Kg\rangle_p,$$

where $K = P^T$ is the approximation to the Koopman operator. Hence

$$\langle \tilde{P}f, g\rangle_q = \langle f, P^T g\rangle_p.$$

To sum up, we have to compute the singular value decomposition of $\tilde{P}$, i.e an eigenvalue decomposition of $\tilde{P}^*\tilde{P} = P^T\tilde{P}$. For this discretization holds, that $\sigma_1 = 1$ is the leading singular value:

Lemma 4. *The leading singular value of* $\tilde{P} : (\mathbb{R}^n, \langle\cdot,\cdot,\rangle_p) \to (\mathbb{R}^n, \langle\cdot,\cdot,\rangle_q)$ *is equal to 1.*

Proof. The squares of the singular values of $\tilde{P}$ are the eigenvalues of $\tilde{P}^*\tilde{P}$. We have seen that $\tilde{P}^* = P^T$. $\tilde{P}$ is row-stochastic by construction, and P is column stochastic, see (2.19). Hence, P^T is row-stochastic and $P^T\tilde{P}$ is row-stochastic. It follows, that the leading eigenvalue of $P^T\tilde{P}$ is equal to 1. □

This discretization works fine, but naively we have to compute the matrix-matrix multiplication of two matrices, that are hopefully sparse. This is computationally very expensive. In order to avoid that, we perform a transformation from $(\mathbb{R}^d, \langle\cdot,\cdot\rangle_p)$ to $(\mathbb{R}^d, \langle\cdot,\cdot\rangle_2)$. Like that, we get an optimization problem in $(\mathbb{R}^d, \langle\cdot,\cdot\rangle_2)$, where adjungating is simply transposing. To this end, we consider the discrete version of (3.3) and adapt [Froyland et al., 2010b, Lemma 1] for our purposes:

Lemma 5. *The solution to*

$$\max_{x_1,\dots,x_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\|\tilde{P}x_k\|_q}{\|x_k\|_p} \;:\; \langle x_i, x_j\rangle_p = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\} \tag{3.7}$$

is given by the sum $\sum_{k=1}^{n} \sigma_k(\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}})$ *of the n largest singular values* σ_k, $k = 1,\dots,n$, *of* $\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}}$. *The maximizing* x_k *are given by* $v_k = \Pi_p^{\frac{1}{2}}\hat{v}_k$, *where* $\hat{v}_k$ *is the corresponding* $k-th$ *right singular vector. Furthermore the leading singular value is* $\sigma_1 = 1$.

Proof. We convert (3.7) to a maximization in the standard euclidean norm via

$$\begin{aligned}
&\max_{x_1,\dots,x_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\|\tilde{P}x_k\|_q}{\|x_k\|_p} \;:\; \langle x_i, x_j\rangle_p = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\} \\
&= \max_{x_1,\dots,x_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\langle \Pi_q^{\frac{1}{2}}\tilde{P}x_k, \Pi_q^{\frac{1}{2}}\tilde{P}x_k\rangle_2}{\langle \Pi_p^{\frac{1}{2}}x_k, \Pi_p^{\frac{1}{2}}x_k\rangle_2} \;:\; \langle \Pi_p^{\frac{1}{2}}x_i, \Pi_p^{\frac{1}{2}}x_j\rangle_2 = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\} \\
&= \max_{x_1,\dots,x_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\|\Pi_q^{\frac{1}{2}}\tilde{P}x_k\|_2}{\|\Pi_p^{\frac{1}{2}}x_k\|_2} \;:\; \langle \Pi_p^{\frac{1}{2}}x_i, \Pi_p^{\frac{1}{2}}x_j\rangle_2 = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\}.
\end{aligned}$$

We substitute $\hat{x} = \Pi_p^{\frac{1}{2}}x$ and obtain

$$\begin{aligned}
&= \max_{\hat{x}_1,\dots,\hat{x}_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\|\Pi_q^{\frac{1}{2}}\tilde{P}\Pi_p^{-\frac{1}{2}}\hat{x}_k\|_2}{\|\hat{x}_k\|_2} \;:\; \langle \hat{x}_i, \hat{x}_j\rangle_2 = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\} \\
&= \max_{\hat{x}_1,\dots,\hat{x}_n\in\mathbb{R}^{N_1}} \left\{ \sum_{k=1}^{n} \frac{\|\Pi_q^{-\frac{1}{2}}P\Pi_p^{\frac{1}{2}}\hat{x}_k\|_2}{\|\hat{x}_k\|_2} \;:\; \langle \hat{x}_i, \hat{x}_j\rangle_2 = \delta_{ij} \quad \forall i,j = 1,\dots,n \right\}
\end{aligned}$$

Again, due to theCourant-Fischer theorem, the maximum of (3.7) is $\sum_{k=1}^{n} \sigma_k(\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}})$, and the maximizing vectors are the corresponding singular vectors $\hat{v}_k,\ k = 1, \dots, n$. Via re-substituting, we obtain that the maximizing vectors v_k of (3.7) are $v_k = \Pi_p^{\frac{1}{2}} \hat{v}_k$.
The claim, that the leading singular value σ_1 is equal to one directly follows from Lemma 4. □

Remark. *Note that we also directly obtain an approximation of the indicator functions of the coherent sets at final time. These are given by* $u_k = \Pi_q^{-\frac{1}{2}} \hat{u}_k$, *where* $\hat{u}_k$ *is the k-th left singular vector of* $\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}}$. *This holds, because*

$$u_k = \tilde{P} v_k = \Pi_q^{-1} P \Pi_p v_k = \Pi_q^{-\frac{1}{2}} \left(\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}} \right) \hat{v}_k = \Pi_q^{-\frac{1}{2}} \hat{u}_k.$$

To sum up, with Lemma 5, we obtain a discretization of (3.2) that preserves the stochasticity of the operator $\mathcal{P}$ and can be solved via a singular value decomposition of the matrix $\Pi_q^{-\frac{1}{2}} P \Pi_p^{\frac{1}{2}} \in (\mathbb{R}^{N_1 \times N_2}, \langle \cdot, \cdot \rangle_2)$. We can avoid the matrix matrix multiplications. Let

$$\begin{pmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_{N_2} \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{N_1 \times N_2}, \quad \begin{pmatrix} - & b_1' & - \\ & \vdots & \\ - & b_{N_1}' & - \end{pmatrix} \in \mathbb{R}^{N_1 \times N_2}, \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_N \end{pmatrix} \in \mathbb{R}^{N \times N}.$$

We realize that the multiplication with diagonal matrix from the right hand side is scaling the columns with the diagonal entries. Multiplication from the left hand side is scaling the rows with the diagonal entries:

$$\begin{pmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_{N_2} \\ | & | & & | \end{pmatrix} \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_{N_2} \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \alpha_1 a_1 & \alpha_2 a_2 & \dots & \alpha_{N_2} a_{N_2} \\ | & | & & | \end{pmatrix}$$

$$\begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_{N_1} \end{pmatrix} \begin{pmatrix} - & b_1' & - \\ & \vdots & \\ - & b_{N_1}' & - \end{pmatrix} = \begin{pmatrix} - & \alpha_1 b_1' & - \\ & \vdots & \\ - & \alpha_{N_1} b_{N_1}' & - \end{pmatrix}.$$

We can implement this very efficiently via directly using the *bsxfun* (Binary Singelton eXpansion) function in matlab:

```
1 Π_q^{-1/2} P Π_p^{1/2} =bsxfun(@times,q.^(-0.5), bsxfun(@times,P,p.^(0.5)));
```

where $P \in \mathbb{R}^{N_1 \times N_2}$ is the Frobenius-Perron operator computed via Ulam's method, $p \in \mathbb{R}^{N_1}$, and $q \in \mathbb{R}^{N_2}$ are the approximations of h_μ, h_ν, respectively. We use this approach for the following numerical experiments.

Remark. *When we use Ulam's method in practice, we often do not know the image domain Ω_2, but start with an initial domain Ω_1, which some particles possibly leave during their evolution over time. We can easily deal with this if we initially define $\mathcal{P}$ mapping into a greater domain $\Omega \supseteq \Omega_2$, and delete all boxes not reached by any particles.*

3.5 Numerical experiments

3.5.1 Double gyre

First we look at the benchmark example of the Double gyre. The model was introduced in [Shadden et al., 2005] and also considered in [Froyland and Padberg-Gehle, 2014], [Froyland and Padberg, 2009], [Williams et al., 2014b], [Ma and Bollt, 2013]. The model describes a flow

$$\Psi(t, x, y) := A \sin(\pi f(t, x)) \sin(\pi y),$$

where $f(t, x) = \delta \sin(\omega t)x^2 + (1 - 2\delta \sin(\omega t))x$, in a domain $\Omega = [0, 2] \times [0, 1]$. The velocity field is given by

$$b(t, x, y) = \begin{pmatrix} -\frac{\partial \Psi}{\partial y} \\ \frac{\partial \Psi}{\partial x} \end{pmatrix} = \begin{pmatrix} -\pi A \sin(\pi f(t, x)) \cos(\pi y) \\ \pi A \cos(\pi f(t, x)) \sin(\pi y) \frac{\partial f}{\partial x}(t, x) \end{pmatrix}. \tag{3.8}$$

This model is not intended to describe a real fluid flow but a simplification of a double-gyre pattern seen in geophysical flows [Shadden et al., 2005].

The example describes two counter-rotating vortices separated by a periodically moving leaky transport barrier. For $\delta = 0$ the flow is autonomous. For $\delta > 0$ the flow is non-autonomous and the gyres conversely expand and contract periodically in the x-direction, such that the domain Ω enclosing the gyres remains invariant. A determines the magnitude of the vector field b, $\omega/2\pi$ is the frequency of oscillation, and δ is approximately how far the line separating the gyres moves to the left and to the right, see [Shadden et al., 2005]. We fix the parameter values to

$$A = 0.25, \qquad \delta = 0.25, \qquad \omega = 2\pi.$$

In Figure 3.4, we show the velocity field of the Double gyre at times 0, 0.25 and 0.75. The Double gyre is 1-periodic in our configuration . We compute an approximation of the maximally coherent sets via Ulam's method using $2^8 \times 2^7$ boxes and 100 test points per box. We use $\mu = \lambda$ as reference measure, which is invariant under the flow. For the time integration, we use a standard Runge-Kutte 4 scheme with step size $h = 0.1$. As initial and final time, we choose $t_0 = 0.25$, $t_1 = 10.75$. In Figure 3.5 we show the singular spectrum of the transfer operator. There we spot a gap after the third singular value.

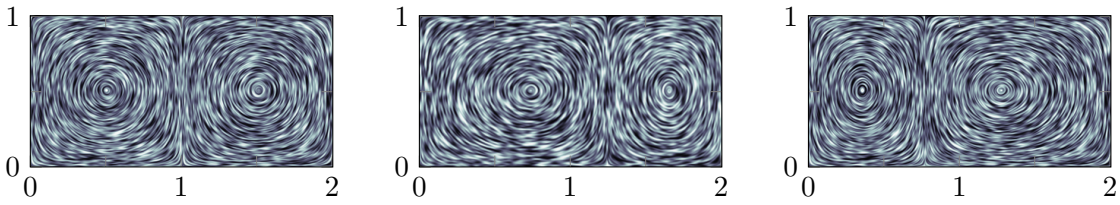

Figure 3.4: Field lines of the vectorfield of the Double gyre at times 0, 0.25 and 0.75. The vectorfield is 1-periodic.

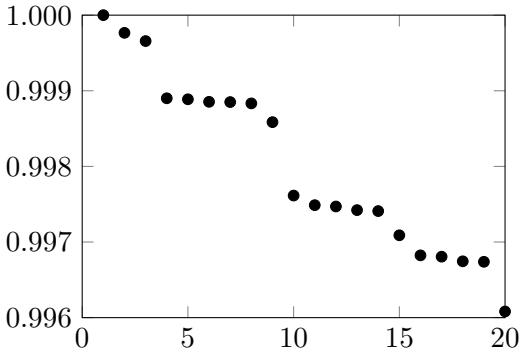

Figure 3.5: Singular values for the Double gyre.

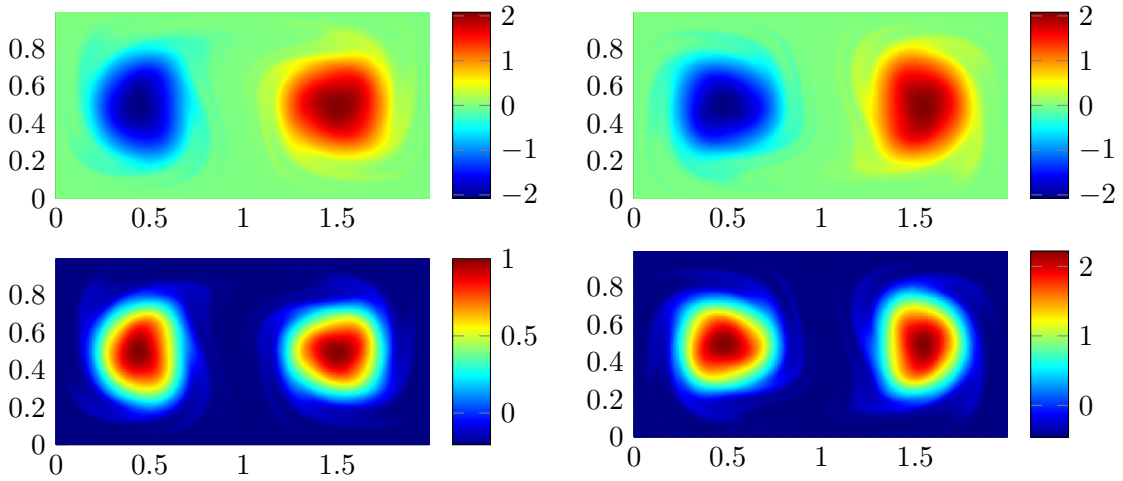

Figure 3.6: First row: Second left and right singular vectors of the Frobenius-Perron operator for the Double gyre, Second row: Third singular vector at initial and final time.

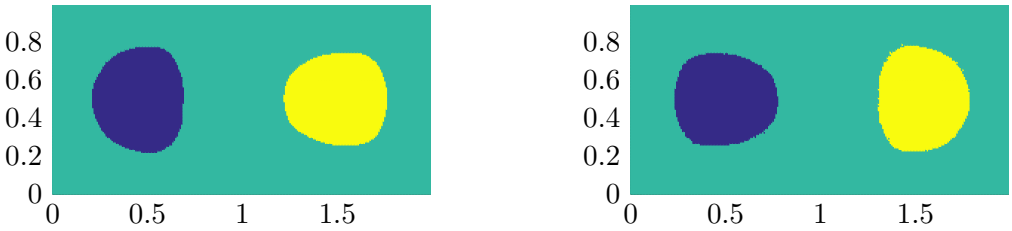

Figure 3.7: Extracted partition into three coherent sets at initial and final time.

In Figure 3.6, we show the second and third singular vectors of the such computed Frobenius-Perron operator. They nicely indicate the two vortices at initial and final time.

We use k-means, see Section 2.8, to actually extract the coherent sets. The partition at initial and final time is shown in Figure 3.7.

3.5.2 Quasiperiodic Bickley jet

The Bickley jet, an idealized model of a meandering zonal jet, see [del Castillo-Negrete and Morrison, 1993] [Rypina et al., 2007] [Hadjighasem et al., 2016], will serve as a test problem throughout the thesis. This model consists of a steady background flow subject to a time-dependent perturbation. The time-dependent Hamiltonian reads

$$\begin{aligned}\psi(x,y,t) &= \psi_0(y) + \psi_1(x,y,t),\\ \psi_0(y) &= -U_0 L_0 \tanh(\frac{y}{L_0}),\\ \psi_1(x,y,t) &= U_0 L_0 \operatorname{sech}^2(\frac{y}{L_0})\mathcal{R}\left(\sum_{n=1}^{3} f_n(t)\exp(ik_n x)\right)\end{aligned}$$

where ψ_0 is the steady background flow and ψ_1 is the perturbation. The constants U_0 and L_0 are the characteristic velocity and the characteristic length scale, respectively. For the following analysis, we apply the set of parameters used in [Rypina et al., 2007] and [Hadjighasem et al., 2016]:

$$U_0 = 62.66 \text{ ms}^{-1}, \qquad L_0 = 1770 \text{ km}, \qquad k_n = \frac{2n}{r_0},$$

where $r_0 = 6371$ km is the mean radius of the earth. For $f_n(t) = \epsilon_n exp(ik_n c_n t)$, the time-dependent part of the Hamiltonian consists of three Rossby waves with wave numbers k_n traveling at speeds c_n. The amplitude of each Rossby wave is determined by the parameters ϵ_n. The parameters we use are the following: $c_1 = 0.1446\ U_0$, $c_2 = 0.205\ U_0$, $c_3 = 0.461\ U_0$, $ly = 1.77 \cdot 10^6$, $\epsilon_1 = 0.0075$, $\epsilon_2 = 0.04$, $\epsilon_3 = 0.3$, $l_x = 6.371 \cdot 10^6\pi$, $k_n = 2n\pi/l_x$. The time interval we consider is from $t_0 = 0$ to $t_1 = 40$ days $= 3.456{\cdot}10^6$ seconds. The initial domain is $\Omega_0 = [0, 20] \times [-3, 3]$ at time t_0. As the flow does not stay in Ω_0,

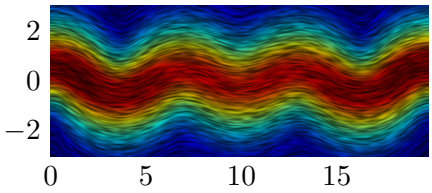

Figure 3.8: Field lines of the vector field of the Bickley jet at time t_0. The red color indicates high velocity.

we define the initial density as $\mu = 1/|\Omega| 1_\Omega$, project μ onto the approximation space V_N and obtain the constant vector $p = 1/(n \cdot m) \in \mathbb{R}^N$. We let the approximate transfer operator P_n map to a greater domain $\Omega = [-5, 5] \times [0, 20]$. Then we only consider the subdomain $\Omega_2 \subset \Omega$ where $q = pP > 0$, i.e. delete all boxes that are not reached by any particle.
A picture of the vector field at time $t_0 = 0$ and t_1 is given in Figure 3.8. We utilize Ulam's method to compute an approximation of the Frobenius-Perron operator for the system, see Section 3.4. Therefore we use $2^7 \times 2^5$ boxes with 100 testpoints per box.
In Figure 3.9 we show the singular values of the transfer operator and see several gaps occurring after the second, the eighth, the ninth and the fifteenth singular value. This indicates that there is not only one meaningful partition but several, and it is not clear a priori which to choose. In Figure 3.10 we show the second singular vector at initial and final time. We see that the 2-partition will simply separate the northern and the southern hemisphere. In Figure 3.11 we show partitions into two, eight, nine and fifteen coherent sets together with the corresponding last singular vectors contributing to the partition:

- The 2-partition indeed separates the northern from the southern hemisphere.
- The 8-partition identifies the six vortices and the two hemispheres.
- The 9-partition separates the zonal jet, too.
- The 15-partition additionally divides the vortices in an inner and outer part.

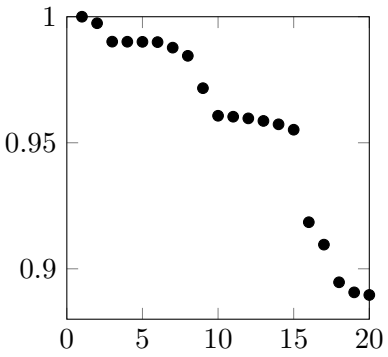

Figure 3.9: Singular values for the Bickley jet.

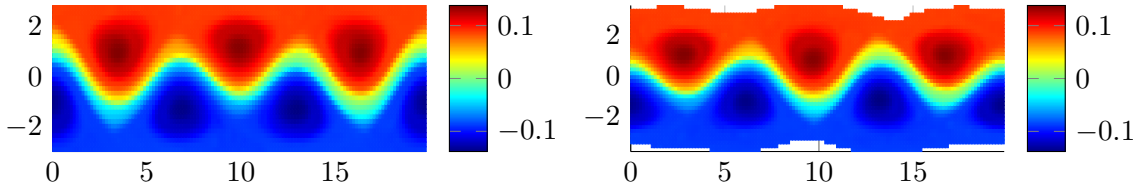

Figure 3.10: Left and right second singular vectors indicating the coherent sets at times t_0 and t_1.

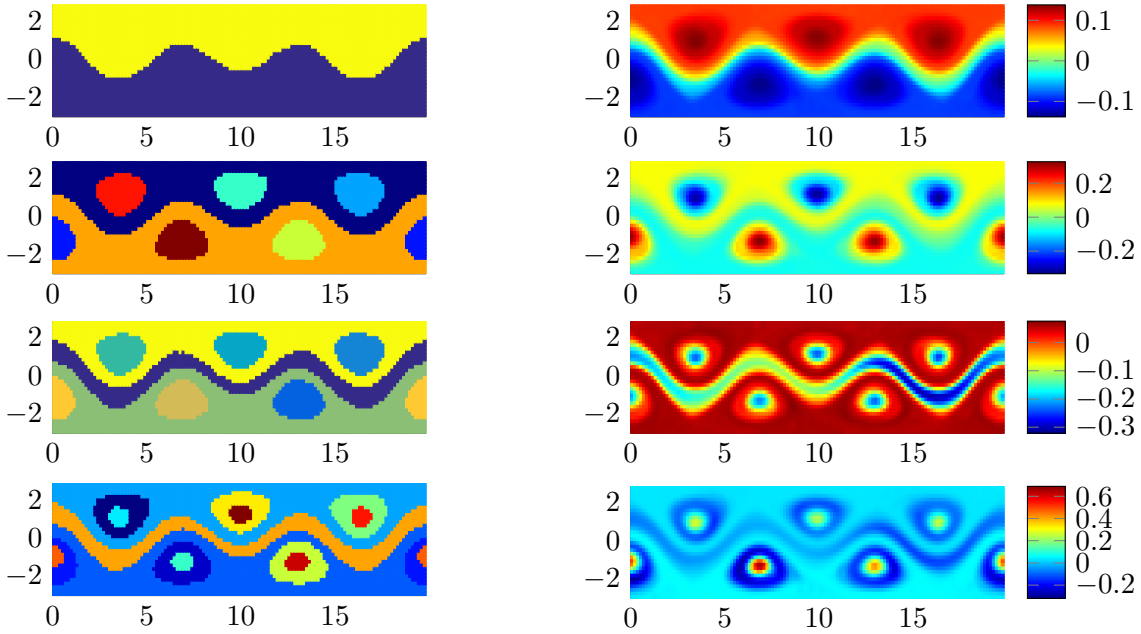

Figure 3.11: First column: Partition into two, eight, nine and fifteen coherent sets. Second column: Second, eighth, ninth and fifteenth singular vectors.

Chapter 4

Coherence via stochastic process

In this chapter, we directly add the diffusion needed for the computation of a coherent partition into the modeling of the dynamical system. We introduce a stochastic flow map, and construct the transfer operator for this system. This approach has the advantage that we can derive some rigorous lower bounds for the approximation of the coherence ratio of the partition. In the second part of the chapter, we choose a stochastic flow map induced by white noise added to the deterministic process. The associated transfer operator solves a Fokker-Planck equation. For the numerical implementation, we employ spectral collocation methods and an exponential time differentiation scheme. We compare our approach experimentally with the more classical method by Ulam.

4.1 Stochastic flow maps

In the previous chapter we called a set $A \subset \Omega$ coherent, if

$$\frac{\mu(A_0 \cap \mathcal{T}^{t_1,t_0} A_1)}{\mu(A_0)} \approx 1$$

subject to some small random perturbation. Here A_0 denotes the coherent set at initial time t_0, A_1 at final time t_1. Using this expression, the connection to transfer operators can be directly established via

$$\mu(A_0 \cap \mathcal{T}^{t_1,t_0} A_1) = \langle \mathcal{P} 1_{A_0}, 1_{A_1} \rangle_\mu.$$

On the other hand, this formulation is challenging because the perturbation is not formulated initially but added "on the way". To overcome this, we aim at including the diffusion directly in the mathematical definition of coherence. We introduce a *stochastic*

flow map of homeomorphisms as in [Kunita, 1997, Chapter 4], and introduce the transfer operators for these. We always consider ϵ-small perturbations of the deterministic system. When talking about stochastic flows, we equip the stochastic flow map with an ϵ to distinguish it from the deterministic one.

Definition 13. *Let $\mathcal{T}_\epsilon^{s,t}(x,\xi),\ s,t\in[t_0,t_1],\ x\in\mathbb{R}^d$ be a continuous, $\mathbb{R}^d$-valued random field defined on the probability space $(\Sigma,\mathcal{F},\mathbb{P})$, $\mathcal{T}_\epsilon^{s,t}(x,.):\Sigma\to\mathbb{R}^d$. Then for almost all $\xi,\ \mathcal{T}_\epsilon^{s,t}(\xi)=\mathcal{T}_\epsilon^{s,t}(\cdot,\xi)$ defines a continuous map from $\mathbb{R}^d$ into itself for any s,t. It is called a* ***stochastic flow of homeomorphisms*** *if there exists a null set N of Σ such that for any $\xi\in N^c$, the family of continuous maps $\left\{\mathcal{T}_\epsilon^{s,t}(\xi):s,t\in[t_0,t_1]\right\}$ defines a* ***flow of homeomorphisms****, i.e. it satisfies the following properties:*

1. *$\mathcal{T}_\epsilon^{s,u}(\xi)=\mathcal{T}_\epsilon^{t,u}(\xi)\circ\mathcal{T}_\epsilon^{s,t}(\xi)=$ holds for all $s,\ t,\ u$,*
2. *$\mathcal{T}_\epsilon^{s,s}(\xi)$ is the identity map for all s,*
3. *the map $\mathcal{T}_\epsilon^{s,t}(\xi):\mathbb{R}^d\to\mathbb{R}^d$ is a homeomorphism for all s,t.*

The definition can, of course, be generalized to suitable manifolds (see [Kunita, 1997, Chapter 4.8]). However, as for our use the domain considered is usually a subspace of $\mathbb{R}^d$, we stay in this setting for now and deal with other domains as they occur.

4.2 Transfer operators for stochastic flows

We now introduce the push forward (Frobenius-Perron) operator for stochastic flow maps. Again we fix an initial time t_0 and a final time t_1. For a σ-finite, invariant measure μ on $\mathbb{R}^d$ (i.e. for all $\xi\in\Sigma$ holds that $\mu\circ\mathcal{T}_\epsilon^{t_0,t_1}(A,\xi)=\mu(A)$ for all $A\in\mathcal{A}$) we define the Frobenius-Perron operator as

$$\begin{aligned}\mathcal{P}_\epsilon&:\mathbb{L}^1(\mathbb{R}^d,\mathcal{A},\mu)\to\mathbb{L}^1(\mathbb{R}^d,\mathcal{A},\mu)\\(\mathcal{P}_\epsilon f)(x)&=\mathbb{E}_{\mathbb{P}}[f\circ\mathcal{T}_\epsilon^{t_1,t_0}(x,\cdot)].\end{aligned}\tag{4.1}$$

This definition is similar to the deterministic one but now we take the mean over all realizations of the stochastic process. The operator itself is deterministic. We again omit the fixed initial and final time $t_0,\ t_1$, respectively.
The *Koopman* operator can be defined in the same way as

$$\begin{aligned}\mathcal{K}_\epsilon&:\mathbb{L}^\infty(\mathbb{R}^d,\mathcal{A},\mu)\to\mathbb{L}^\infty(\mathbb{R}^d,\mathcal{A},\mu)\\(\mathcal{K}_\epsilon f)(x)&=\mathbb{E}_{\mathbb{P}}[f\circ\mathcal{T}_\epsilon^{t_0,t_1}(x,\cdot)].\end{aligned}\tag{4.2}$$

Analogous to the deterministic case, $\mathcal{P}_\epsilon$ and $\mathcal{K}_\epsilon$ can be restricted to $\mathbb{L}^2$ and it can easily be seen, that they are dual operators:

$$\langle\mathcal{P}_\epsilon f,g\rangle_\mu=\langle f,\mathcal{K}_\epsilon g\rangle_\mu.$$

4.3 Coherence via stochastic flows

We next use the stochastic flow for stochastic flow maps to introduce a more rigorous notion of the coherence of sets. We subsequently connect this notion to transfer operators and use these for the computation of coherent partitions. To this end, we assume the stochastic flow map to be induced by ϵ-small perturbations of the deterministic dynamics, and the reference measure μ to be a probability measure.

Definition 14. *We call a set $A \in \mathcal{A}$ (at initial time) $\delta-$**coherent**, if*

$$\mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot)\Big] \;\Big|\; A\Big) = \delta. \tag{4.3}$$

We furthermore call a set coherent, if it is δ-coherent, with $\delta \approx 1$. In other words, a set A is called coherent if A is mapped on average approximately onto itself via the perturbed forward and backward evolution. The expected value $\mathbb{E}_{\mathbb{P}\times\mathbb{P}}$ in (4.3) denotes simply the average over all realizations of the stochastic process in forward ($\mathcal{T}_\epsilon^{t_0,t_1}$) and backward time ($\mathcal{T}_\epsilon^{t_1,t_0}$). This is the same intuition as before but now we specify the small random perturbation as stochastic process.
In the deterministic setting holds $\mu(A_0 \cap \mathcal{T}^{t_1,t_0}A_1) = \langle \mathcal{P}1_{A_0}, 1_{A_1}\rangle$. We can establish a similar connection here:

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[\mu(A_0 \cap \mathcal{T}_\epsilon^{t_1,t_0}(A_1,\cdot))] \\
=&\mathbb{E}_{\mathbb{P}}\left[\int_{\mathcal{T}_\epsilon^{t_1,t_0}(A_1,\cdot)} 1_{A_0}\; d\mu(x)\right] \\
=&\int_\Sigma \int_{\mathcal{T}_\epsilon^{t_1,t_0}(A_1,\cdot)} 1_{A_0}\; d\mu(x)\; d\mathbb{P}(\xi) \\
\overset{\mu \text{ invariant}}{=}&\int_\Sigma \int_{A_1} 1_{A_0}\circ\mathcal{T}_\epsilon^{t_1,t_0}(x,\cdot)\; d\mu(x)\; d\mathbb{P}(\xi) \\
\overset{\text{Fubini}}{=}&\int_{A_1} \int_\Sigma 1_{A_0}\circ\mathcal{T}_\epsilon^{t_1,t_0}(x,\cdot)\; d\mathbb{P}(\xi)\; d\mu(x) \\
=&\int_{A_1} \mathbb{E}_{\mathbb{P}}\left[1_{A_0}\circ\mathcal{T}_\epsilon^{t_1,t_0}(x,\cdot)\right]\; d\mu(x) \\
=&\int_{\mathbb{R}^d} \mathcal{P}_\epsilon 1_{A_0} 1_{A_1}\; d\mu(x) \\
=&\langle \mathcal{P}_\epsilon 1_{A_0}, 1_{A_1}\rangle_\mu.
\end{aligned}$$

In the connection between the spectrum of the deterministic transfer operator and the maximally coherent partition in Section 3.3, we saw that the optimal choice for 1_{A_1} is

$\mathcal{P}_\epsilon 1_{A_0}$. Hence we can directly include this into the considerations:

$$
\begin{aligned}
&\langle \mathcal{P}_\epsilon 1_{A_0}, \mathcal{P}_\epsilon 1_{A_0} \rangle \\
=&\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon 1_{A_0}, 1_{A_0} \rangle_\mu \\
=&\int_{\mathbb{R}^d} \mathcal{K}_\epsilon \mathcal{P}_\epsilon 1_{A_0}(x) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \mathbb{E}_{\mathbb{P}} \left[\mathcal{P}_\epsilon 1_{A_0} \circ \mathcal{T}_\epsilon^{t_0,t_1}(x,\cdot) \right] 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_\Sigma \mathcal{P}_\epsilon 1_{A_0} \circ \mathcal{T}_\epsilon^{t_0,t_1}(x,\xi_1)\ d\mathbb{P}(\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_\Sigma \mathbb{E}_{\mathbb{P}} \left[1_{A_0} \circ \mathcal{T}_\epsilon^{t_1,t_0}(\cdot,\cdot) \right] \circ \mathcal{T}_\epsilon^{t_0,t_1}(x,\xi_1)\ d\mathbb{P}(\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_\Sigma \left(\int_\Sigma 1_{A_0} \circ \mathcal{T}_\epsilon^{t_1,t_0}(\cdot,\xi_2)\ d\mathbb{P}(\xi_2) \right) \circ \mathcal{T}_\epsilon^{t_0,t_1}(x,\xi_1)\ d\mathbb{P}(\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_\Sigma \left(\int_\Sigma 1_{A_0}(\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(x,\xi_1),\xi_2))\ d\mathbb{P}(\xi_2) \right) d\mathbb{P}(\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_{\Sigma\times\Sigma} 1_{A_0}(\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(x,\xi_1),\xi_2))\ d(\mathbb{P}\times\mathbb{P})(\xi_2,\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\int_{\mathbb{R}^d} \left(\int_{\Sigma\times\Sigma} 1_{\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A_0,\xi_2),\xi_1)}(x)\ d(\mathbb{P}\times\mathbb{P})(\xi_2,\xi_1) \right) 1_{A_0}(x)\ d\mu(x) \\
=&\mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}} \big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A_0,\cdot),\cdot) \big] \cap A_0 \Big).
\end{aligned}
\tag{4.4}
$$

If we use this for a definition of coherence, we can motivate (4.3) and call a set $A = A_0 \in \mathcal{A}$ (at initial time) *coherent*, if

$$
\frac{\mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}} \big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot) \big] \cap A \Big)}{\mu(A)} = \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}} \Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot) \Big] \ \Big|\ A \Big) \approx 1.
$$

which tells us that coherence is nothing but almost invariance under forward-backward evolution of the stochastic flow.

The following considerations are then analogous to the previous chapter: We now want to compute an actual *coherent partition* of the domain $\Omega \subset \mathbb{R}^d$ into n *coherent sets*. We never consider a coherent set on its own but always a partition. We use as definition of a coherent set

$$
\rho(A) := \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}} \Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot) \Big] \ \Big|\ A \Big) \approx 1. \tag{4.5}
$$

The number n of those sets however again is not known a priori. So we try to find a formulation which shows the partition into n coherent sets for several $n \geq 1$, and

choose a meaningful number from that. We formulate the optimization problem for $A_1, \dots, A_n \in \mathcal{A}$

$$\max_{\Omega=\dot{\bigcup} A_k} \sum_{k=1}^{n} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A_k,\cdot),\cdot)\Big] \;\Big|\; A_k\Big) \tag{4.6}$$

which we interpret in $\mathbb{L}^2(\Omega,\mathcal{A},\mu)$, demanding that the coherent sets A_k form a partition of the domain and that they keep their mass during the evolution. We can reformulate (4.6) as

$$\begin{aligned}
&\max_{\Omega=\dot{\bigcup} A_k} \sum_{k=1}^{n} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A_k,\cdot),\cdot)\Big] \;\Big|\; A_k\Big) \\
=&\max_{\Omega=\dot{\bigcup} A_k} \sum_{k=1}^{n} \frac{\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon 1_{A_k}, 1_{A_k}\rangle_\mu}{\|1_{A_k}\|_\mu^2} \\
\leq&\max_{f_1\perp\cdots\perp f_n\in\mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \frac{\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon f_k, f_k\rangle_\mu}{\|f_k\|_\mu^2} \\
=&\max_{f_1\perp\cdots\perp f_n\in\mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \left\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \left(\frac{f_k}{\|f_k\|_\mu}\right), \frac{f_k}{\|f_k\|_\mu}\right\rangle_\mu \\
=&\max_{f_1,\dots,f_n \text{ ONB of } V_n\leq\mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon f_k, f_k\rangle_\mu \\
=&\sum_{k=1}^{n} \lambda_k(\mathcal{K}_\epsilon\mathcal{P}_\epsilon) \\
=&\sum_{k=1}^{n} \sigma_k(\mathcal{P}_\epsilon)^2
\end{aligned}$$

due to the Courant-Fischer theorem 8. In order to apply the theorem, we again have to assume the diffusion in the system to make the transfer operator $\mathcal{P}_\epsilon$ compact (or at least, that $\mathcal{P}_\epsilon$ has n singular values below the bottom of the essential spectrum). The maximizing functions are the corresponding first n singular functions of $\mathcal{P}_\epsilon$.

Remark. *Note that instead of the singular values σ_k of $\mathcal{P}_\epsilon$, we obtain σ_k^2. This happens because this notion of coherence considers directly forward and backward evolution. We can also normalize the coherence to $\sqrt{\rho(A)}$ and obtain the same algorithms as before.*

Upper and lower bound

In Chapter 3.3 we observed that, because a basis of the vector space $V_n = \operatorname{span}(v_1, \dots, v_n)$ is not unique, the computed singular functions $v_1, \dots, v_n$ are not necessarily (approximations to) the indicator functions of the coherent sets $1_{A_1}, \dots, 1_{A_n}$. We examined that the computed singular functions are approximations to linear combinations of these if

our relaxation step was not too strong. For the spectral decomposition of $\mathcal{P}_\epsilon$ should hold:

1. the first n singular values are close to one,

2. $v_1, \ldots, v_n$ are good approximations of linear combinations of n indicator functions partitioning Ω.

We now take a closer look at this statement. The first condition is obvious by our initial motivation, and ensures that the partition is actually coherent. The second is as mentioned an a posteriori check if the relaxation to functions in $\mathbb{L}^2$ was not strong. We now derive a theorem which rigorously links the quality of this approximation with the coherence of the indicated sets. To this end, we assume that we already have computed a partition of the domain Ω into coherent sets $A_1, \ldots, A_n$. We consider the first n singular values $\sigma_1, \cdots, \sigma_n$ and right singular vectors $v_1, \ldots, v_n$ of $\mathcal{P}_\epsilon$. We choose the singular vectors to be orthonormal and introduce the following projections:

$$\begin{aligned} Q &: \mathbb{L}^2(\Omega, \mathcal{A}, \mu) \longrightarrow \operatorname{span}(1_{A_1}, \ldots, 1_{A_n}) \\ Qf &= \sum_{k=1}^{n} \langle f, \chi_{A_k} \rangle_\mu \chi_{A_k} \end{aligned}$$

where $\chi_{A_k} = 1_{A_k} / \|1_{A_k}\|_2,\ k = 1 \ldots n$ and

$$\begin{aligned} \Pi &: \mathbb{L}^2(\Omega, \mathcal{A}, \mu) \longrightarrow \operatorname{span}(v_1, \ldots, v_n) \\ \Pi f &= \sum_{k=1}^{n} \langle f, v_k \rangle_\mu v_k. \end{aligned}$$

Further we define $\Pi^\perp := \mathrm{Id} - \Pi$. We can now state a theorem inducing a lower bound for the coherence ratios depending on how good the singular vectors of $\mathcal{P}_\epsilon$, v_i approximate linear combinations of indicator functions of the coherent sets χ_{A_i}, i.e. depending on how strong our relaxation step was:

Theorem 10. *It holds that*

$$\sum_{k=1}^{n} \rho_k \sigma_k^2 \leq \sum_{k=1}^{n} \rho(A_k) \leq \sum_{k=1}^{n} \sigma_k^2,$$

where $\rho_k = \|Q v_k\|_2^2 \in [0, 1]$.

Proof. The proof is strongly related to the one elaborated in [Huisinga and Schmidt, 2006]. The second inequality directly follows from the derivation because we used a

relaxation from indicator functions to general functions in $\mathbb{L}^2$ in order to end up with the singular value decomposition. For the first inequality we look at

$$\begin{aligned}
\sum_{j=1}^{n} \rho(A_j) &= \sum_{j=1}^{n} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot)\Big] \ \cap \ A\Big) \\
&= \sum_{j=1}^{n} \langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \chi_{A_j}, \chi_{A_j} \rangle_\mu \\
&= \sum_{j=1}^{n} \langle \Big(\mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi + \mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi^\perp\Big) \chi_{A_j}, \Big(\Pi + \Pi^\perp\Big) \chi_{A_j} \rangle_\mu \\
&= \sum_{j=1}^{n} \langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi \chi_{A_j}, \Pi \chi_{A_j} \rangle_\mu + \sum_{j=1}^{n} \langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi^\perp \chi_{A_j}, \Pi^\perp \chi_{A_j} \rangle_\mu.
\end{aligned}$$

The last equality holds, because Π is a projection and, hence, $\Pi^* = \Pi$, $\Pi^2 = \Pi$ and

$$\begin{aligned}
\langle \Pi f, \Pi^\perp f \rangle_\mu &= \langle \Pi f, f \rangle_\mu - \langle \Pi f, \Pi f \rangle_\mu = \langle \Pi f, f \rangle_\mu - \langle \Pi^* \Pi f, f \rangle_\mu \\
&= \langle \Pi f, f \rangle_\mu - \langle \Pi^2 f, f \rangle_\mu = \langle \Pi f, f \rangle_\mu - \langle \Pi f, f \rangle_\mu = 0.
\end{aligned}$$

and the fact that the action of $\mathcal{K}_\epsilon \mathcal{P}_\epsilon$ does not leave the span of singular vectors of $\mathcal{P}_\epsilon$. Like that the mixed terms vanish. For the first summand we compute

$$\begin{aligned}
\sum_{j=1}^{n} \langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi \chi_{A_j}, \Pi \chi_{A_j} \rangle_\mu &= \sum_{j=1}^{n} \left\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \sum_{k=1}^{n} \langle \chi_{A_j}, v_k \rangle_\mu v_k, \sum_{l=1}^{n} \langle \chi_{A_j}, v_l \rangle_\mu v_l \right\rangle_\mu \\
&= \sum_{j=1}^{n} \left\langle \sum_{k=1}^{n} \sigma_k^2 \langle \chi_{A_j}, v_k \rangle_\mu v_k, \sum_{l=1}^{n} \langle \chi_{A_j}, v_l \rangle_\mu v_l \right\rangle_\mu \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_k^2 \langle \chi_{A_j}, v_k \rangle_\mu^2 \langle v_k, v_k \rangle_\mu \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \sigma_k^2 \langle \chi_{A_j}, v_k \rangle_\mu^2 \\
&= \sum_{k=1}^{n} \sigma_k^2 \langle Q v_k, Q v_k \rangle_\mu \\
&= \sum_{k=1}^{n} \sigma_k^2 \|Q v_k\|_2^2.
\end{aligned}$$

To sum up we have

$$\begin{aligned}\sum_{j=1}^{n} \rho(A_j) &= \sum_{j=1}^{n} \sigma_j^2 \|Qv_j\|_2^2 + \sum_{j=1}^{n} \left\langle \mathcal{K}_\epsilon \mathcal{P}_\epsilon \Pi^\perp \chi_{A_j}, \Pi^\perp \chi_{A_j} \right\rangle_\mu \\ &= \sum_{j=1}^{n} \sigma_j^2 \|Qv_j\|_2^2 + \underbrace{\sum_{j=1}^{n} \|\mathcal{P}_\epsilon \Pi^\perp \chi_{A_j}\|_2^2}_{\geq 0}\end{aligned}$$

□

4.4 The Fokker-Planck equation and coherent sets

This section is based on [Denner et al., 2016]. We now want to employ a stochastic flow map and actually compute coherent sets. We assume that the dynamics are induced by a time-dependent ordinary differential equation $\dot{x} = b(t, x)$, $b : \mathbb{R} \times \Omega \to \mathbb{R}^d$, on some bounded domain $\Omega \subset \mathbb{R}^d$. We fix some initial and final time $t_0, t_1 \in \mathbb{R}$, and assume that the vector field b is continuous and locally Lipschitz w.r.t. x for all $t \in [t_0, t_1]$ such that the associated flow map $\mathcal{T}^{t_0,t_1} : \Omega \to \Omega$ is uniquely defined. In order not to obscure the key ideas, we restrict to the case of Ω being a hyperrectangle, and b being periodic in x and divergence free, i.e. the flow map $\mathcal{T}^{t_0,t_1}$ being volume preserving.
We incorporate a small random perturbation continuously in time, i.e. we now use the stochastic differential equation

$$dx = b(t, x)dt + \epsilon dB_t \tag{4.7}$$

in order to define the stochastic flow map $\mathcal{T}_\epsilon^{t_0,t_1}$. Here, $(B_t)_{t\geq 0}$ is d-dimensional Brownian motion and $\epsilon > 0$. Since we assume $b(t, \cdot)$ to be Lipschitz, Ω to be bounded and b to be periodic in x, for any initial condition $\xi \in \Omega$, (4.7) has a unique continuous solution x in the sense of [Oksendal, 2003, Theorem 5.2.1]. The associated flow map forms a stochastic flow map [Kunita, 1997]. Due to Feynman-Kac formula, the transfer operator $\mathcal{P}_\epsilon$ associated to this stochastic differential equation is given by the solution operator of the parabolic *Fokker-Planck equation*

$$\partial_t u = L_\epsilon u := \tfrac{\epsilon^2}{2} \Delta u - \operatorname{div}(ub), \tag{4.8}$$

see also [Lasota and Mackey, 1993]. Appropriate boundary conditions are chosen (e.g., periodic or homogeneous Neumann boundary conditions), so that for all $u, w \in L^2(\Omega)$ in the domain of L_ϵ holds:

$$\langle w, L_\epsilon u \rangle = -\frac{\epsilon^2}{2} \int_\Omega \nabla u \cdot \nabla w + \int_\Omega u\, b \cdot \nabla w. \tag{4.9}$$

More precisely, $\mathcal{P}_\epsilon u_0 = u(t_1, \cdot)$, where u is the solution to (4.8) with initial condition $u(t_0, \cdot) = u_0$. We then have to ensure the transfer operator $\mathcal{P}_\epsilon$ to be compact and stochastic.

Lemma 6. *If*

$$\|b\|_{C^1} = \sup_{s\in[t_0,t_1]} \sup_{x\in\Omega} \max\{|b(s,x)|, |\partial_{x_1} b(s,x)|, \dots, |\partial_{x_d} b(s,x)|\} < \infty,$$

then $\mathcal{P}_\epsilon : L^2(\Omega) \to L^2(\Omega)$ *is compact.*

Proof. We prove the lemma in the slightly more general case that $\operatorname{div}(b) \neq 0$. $\|b\|_{C^0} := \sup_{s\in[t_0,t_1],\ x\in\Omega}\{|b(t,x)| < \infty, i = 1,\dots,d\}$. Let $u = u(t,x)$ be the solution of (4.8) with initial condition $f = u_0 \in L^2(\Omega)$. Without loss of generality we set $\epsilon = \sqrt{2}$ and $t_0 = 0$. We first note that $\|u\|_2$ is bounded by $\|u_0\|_2$:

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|u\|_2^2 &= \langle u, \partial_t u\rangle = \langle u, \Delta u + \operatorname{div}(ub)\rangle = \langle u, \Delta u\rangle + \langle u, \operatorname{div}(ub)\rangle \\ &= -\|\nabla u\|_2^2 - \langle \nabla u, ub\rangle = -\|\nabla u\|_2^2 - \langle \frac{1}{2}\nabla(u^2), b\rangle \\ &= -\|\nabla u\|_2^2 - \frac{1}{2}\langle u^2, \operatorname{div}(b)\rangle \leq -\|\nabla u\|_2^2 + \frac{1}{2}\|u\|_2^2\|b\|_{C^1}.\end{aligned} \tag{4.10}$$

Gronwall's inequality thus implies that for all $t > 0$

$$\|u\|_2^2 \leq e^{t\|b\|_{C^1}}\|u_0\|_2^2. \tag{4.11}$$

We now show that also $\|\nabla u\|_2$ is bounded by $\|u_0\|_2$. Because of (4.10) we have

$$\|\nabla u\|_2^2 \leq -\frac{1}{2}\frac{d}{dt}\|u\|_2^2 + \frac{1}{2}\|u\|_2^2\|b\|_{C^1}.$$

Integrating from $t = t_0 = 0$ to $t = t_1$ we obtain

$$\begin{aligned}\int_0^{t_1} \|\nabla u\|_2^2\, dt &\leq \frac{1}{2}\left(\|u_0\|_2^2 - \|u(t_1)\|_2^2\right) + \frac{1}{2}\|b\|_{C^1}\int_0^{t_1}\|u\|_2^2\, dt \\ &\overset{(4.11)}{\leq} \frac{1}{2}\|u_0\|_2^2 + \frac{1}{2}\|b\|_{C^1}\int_0^{t_1} e^{t\|b\|_{C^1}}\|u_0\|_2^2 ds \\ &= \frac{1}{2}e^{t_1\|b\|_{C^1}}\|u_0\|_2^2.\end{aligned} \tag{4.12}$$

Therefore there is at least one $t^* \in [0, t_1]$, such that

$$\|\nabla u(t^*,\cdot)\|_2^2 \leq \frac{1}{2t_1}e^{t_1\|b\|_{C^1}}\|u_0\|_2^2, \tag{4.13}$$

and we finally get

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|\nabla u\|_2^2 &= \frac{1}{2}\frac{d}{dt}\langle\nabla u, \nabla u\rangle = \frac{1}{2}\langle\frac{d}{dt}\nabla u, \nabla u\rangle + \frac{1}{2}\langle\nabla u, \frac{d}{dt}\nabla u\rangle \\ &= -\langle\Delta u, \partial_t u\rangle = -\langle\Delta u, \Delta u + \nabla(ub)\rangle \\ &= -\|\Delta u\|_2^2 - \langle\Delta u, \nabla ub\rangle - \langle\Delta u, u\operatorname{div}(b)\rangle \\ &\overset{C-S}{\leq} -\|\Delta u\|_2^2 + \|\Delta u\|_2\|\nabla u\|_2\|b\|_{C^0} + \|\Delta u\|_2\|u\|_2\|b\|_{C^1} \\ &\leq \frac{1}{2}\|\nabla u\|_2^2\|b\|_{C^0}^2 + \frac{1}{2}\|u\|_2^2\|b\|_{C^1}^2,\end{aligned}$$

where we used that $\frac{1}{2}(a^2c^2 + d^2e^2) \geq -b^2 + bac + bde$ for $a, b, c, d, e \in \mathbb{R}$.

Integration from $t = t^*$ to $t = t_1$ yields

$$\begin{aligned}\|\nabla u(t_1)\|_2^2 &\leq \|\nabla u(t^*)\|_2^2 + \int_{t^*}^{t_1} \|\nabla u\|_2^2 \|b\|_{C^0}^2 \, dt + \int_{t^*}^{t_1} \|u\|_2^2 \|b\|_{C^1}^2 \, dt \\ &\overset{(4.11)(4.12)(4.13)}{\leq} \frac{1}{2t_1} e^{t_1\|b\|_{C^1}} \|u_0\|_2^2 + \|b\|_{C^0}^2 \frac{1}{2} e^{t_1\|b\|_{C^1}} \|u_0\|_2^2 \\ &\quad + \|b\|_{C^1}^2 \int_{t^*}^{t_1} e^{t\|b\|_{C_1}} \|u_0\|_2^2 \, dt \\ &\leq \left(\frac{1}{2t_1} + \frac{1}{2}\|b\|_{C^0}^2 + \|b\|_{C^1}\right) \|u_0\|_2^2 e^{t_1\|b\|_{C^1}}.\end{aligned} \tag{4.14}$$

Combining equations (4.11) and (4.14) we obtain

$$\|\mathcal{P}_\epsilon u_0\|_{H^1} = \|u(t_1, \cdot)\|_{H^1} \leq \left(1 + \frac{1}{2t_1} + \frac{1}{2}\|b\|_{C^0}^2 + \|b\|_{C^1}\right) e^{t_1\|b\|_{C^1}} \|u_0\|_2^2.$$

Hence $\mathcal{P}_\epsilon$ maps bounded sets in $L^2(\Omega)$ onto bounded sets in $H^1(\Omega)$. As the embedding of $H^1(\Omega)$ onto $L^2(\Omega)$ is compact by Rellich's theorem, $\mathcal{P}_\epsilon$ is a compact operator. □

Since we assumed b to be divergence free, we also have:

Lemma 7. *$\mathcal{P}_\epsilon : L^2(\Omega) \to L^2(\Omega)$ and $\mathcal{K}_\epsilon : L^2(\Omega) \to L^2(\Omega)$ are stochastic.*

Proof. We set $\epsilon = \sqrt{2}$ without loss of generality. Since $L_\epsilon 1_\Omega = \Delta 1_\Omega - \operatorname{div}(1_\Omega b) = \operatorname{div} b = 0$, it follows that 1_Ω is a steady state of (4.8), and consequently $\mathcal{P}_\epsilon 1_\Omega = 1_\Omega$.

For the adjoint operator $\mathcal{K}_\epsilon$ we test with $u \in L^2(\Omega)$:

$$\langle \mathcal{K}_\epsilon 1_\Omega, u\rangle = \langle 1_\Omega, \mathcal{P}_\epsilon u\rangle = \int_\Omega \mathcal{P}_\epsilon u = \int_\Omega u = \int_\Omega u 1_\Omega = \langle 1_\Omega, u\rangle,$$

where we have used that $\mathcal{P}_\epsilon$ is integral conserving:

$$\partial_t \int u = \int L_\epsilon u = \langle 1_\Omega, L_\epsilon u\rangle = -\int_\Omega \nabla u \cdot \nabla 1_\Omega + \int_\Omega u\, b \cdot \nabla 1_\Omega = 0,$$

thanks to the integration-by-parts rule (4.9). Hence $\int_\Omega u = \int_\Omega \mathcal{P}_\epsilon u$ for all $u \in L^2(\Omega)$, cf. [Lasota and Mackey, 1993]. By the Riesz representation theorem, we can conclude that also $\mathcal{K}_\epsilon 1_\Omega = 1_\Omega$.

For proving positivity of $\mathcal{P}_\epsilon$, we consider the evolution of the negative part $u_-(s, x) = -\min(u(s, x), 0)$. Since $\partial_t(u_-^2) = 2u_-\partial_t u$ almost everywhere on $\Omega \times (t_0, t_1)$, using the

integration-by-parts rule (4.9), we have

$$\begin{aligned}\frac{1}{2}\partial_t \int_\Omega u_-^2 &= \int_\Omega u_- \partial_t u = \int u_- L_\epsilon u \\ &= -\int_\Omega \nabla u \cdot \nabla u_- + \int_\Omega u\, b \cdot \nabla u_- = -\int_\Omega |\nabla u_-|^2 + \frac{1}{2}\int_\Omega b \cdot \nabla(u_-^2) \\ &= -\int_\Omega |\nabla u_-|^2 - \frac{1}{2}\int_\Omega \operatorname{div}(b)\; u_-^2 \\ &= -\int_\Omega |\nabla u_-|^2 \leq 0.\end{aligned} \tag{4.15}$$

Hence if $u(t_0, \cdot)$ is non-negative, $u(t, \cdot)$ is non-negative for all $t > t_0$, as the norm of its negative part does not increase. To show positivity of $\mathcal{K}_\epsilon$, let two non-negative functions $u, w \in L^2(\Omega)$ be given. Then

$$\langle \mathcal{K}_\epsilon w, u\rangle = \langle w, \mathcal{P}_\epsilon u\rangle \geq 0$$

by positivity of $\mathcal{P}_\epsilon$. For any fixed non-negative w, this relation holds for all non-negative u. This implies that $\mathcal{K}_\epsilon w$ is non-negative. □

Hence, $\mathcal{P}_\epsilon$ is compact and stochastic.

4.5 Discretization of the Fokker-Planck equation

In order to approximate the transfer operator $\mathcal{P}_\epsilon$, we choose a finite dimensional approximation space $V_N \subset L^2(\Omega)$, and use collocation. As the Fokker-Planck equation (4.8) is parabolic, and since we assume the vector field $b(t, \cdot)$ to be smooth for all t, its solution $u(t, \cdot)$ is smooth for all $t > t_0$ (see e.g. [Evans, 2010], Chapter 7, Theorem. 7). To exploit this, we choose V_N as the span of the Fourier basis

$$\varphi_{\boldsymbol{k}}(x) = e^{i\langle \boldsymbol{k}, x\rangle}, \quad \boldsymbol{k} \in \mathbb{Z}^d, \|\boldsymbol{k}\|_\infty \leq (N-1)/2, N \text{ odd}.$$

Note that $\dim(V_N) = N^d$. Choosing a corresponding set $\{x_1, \ldots, x_M\} \subset \Omega$ of collocation points (typically on an equidistant grid), the entries of the matrix representation P_ϵ of $\mathcal{P}_\epsilon$ are given by

$$P_\epsilon^{j\boldsymbol{k}} = \mathcal{P}_\epsilon \varphi_{\boldsymbol{k}}(x_j), \tag{4.16}$$

where $\boldsymbol{k} \in \mathbb{Z}^d, \|\boldsymbol{k}\|_\infty \leq (N-1)/2$ and $j = 1, \ldots, M$. For P_ϵ, we then compute singular values and vectors via standard algorithms. Note that we might choose $M \geq N$, i.e. more collocation points than basis functions. This turns out to be useful, because we expect the maximally coherent sets to be comparatively coarse structures which can be captured with a small number of basis functions.

Solving the Fokker-Planck equation. In order to compute $\mathcal{P}_\epsilon \varphi_{\boldsymbol{k}}$ in (4.16) for some basis function $\varphi_{\boldsymbol{k}} \in V_N$ we need to solve the Fokker-Planck equation (4.8) with initial condition $u(t_0, \cdot) = \varphi_{\boldsymbol{k}}$. This can efficiently be done in Fourier space via integrating the Cauchy problem

$$\partial_t \hat{u} = \tfrac{\epsilon^2}{2} \hat{\Delta} \hat{u} - \hat{\operatorname{div}} \mathcal{F}(\mathcal{F}^{-1}(\hat{u}) b), \quad \hat{u}(t_0, \cdot) = \hat{\varphi}_{\boldsymbol{k}},$$

in time, where $\hat{v} = \mathcal{F}(v)$ is the Fourier transform of $v \in V_N$. Note that the differential operators in Fourier space reduce to multiplications with diagonal matrices, while $\mathcal{F}$ and $\mathcal{F}^{-1}$ can efficiently be computed by the (inverse) fast Fourier transform.

Aliasing. One problem with this formulation is the possible occurence of aliasing. As $\hat{u}$ and $\hat{b}$ are trigonometric polynomials of degree N, the multiplication $\mathcal{F}^{-1}(\hat{u})b$ in the advection term leads to a polynomial $\mathcal{F}^{-1}(\hat{u})b$ of degree $2N$, which cannot be represented in our approximation space V_N. The coefficients of degree $\geq N$ of this polynomial act on the coefficients of lower degree leading to unphysical contributions in these. This shows up in high oscillations and blow ups (see [Boyd, 2001], Chapter 11) in the computed solution. One way to deal with this problem is to use the advection term

$$\operatorname{div}(bu) = \frac{1}{2} \operatorname{div}(bu) + \frac{1}{2} (b \nabla u).$$

The spectral discretization of the left-hand side is not skew symmetric but the discretization of the right-hand side is [Zang, 1991]. This leads to purely imaginary eigenvalues of the resulting discretization matrix, and hence to mass conservation. For the unperturbed operator ($\epsilon = 0$), the resulting matrix consequently has eigenvalues on the unit circle. However, this approach has to be used carefully: Even though the solution does not blow up, it might still come with a large error. For example small scale structures may be suppressed. If the system produces such small scale structures the grid has to be chosen fine enough to resolve them.

Time integration. For low resolutions, the time integration of the space discretized system can be performed by a standard explicit scheme. For higher resolutions, the stiffness of the system due to the Laplacian becomes problematic. and a more sophisticated method must be employed. Here, we use the *exponential time differentiation scheme* [Cox and Matthews, 2002] for the space discretized system. The etd-scheme separates the diffusion term $\mathcal{L} = \frac{\epsilon^2}{2} D$, where D is the discretized Laplacian, from the advection term $\mathcal{N}(u, t) = -\operatorname{div} \mathcal{F}(\mathcal{F}^{-1}(\hat{u}) b(\cdot, t))$, where b is evaluated via spectral collocation. The system can thus be written as

$$u_t = \mathcal{L} u + \mathcal{N}(u, t). \tag{4.17}$$

Via multiplying (4.17) with $e^{-t\mathcal{L}}$ and integrating from t_0 to t_1 we obtain

$$u(t_1) = e^{h\mathcal{L}} u(t_0) + e^{h\mathcal{L}} \int_0^h e^{-\tau\mathcal{L}} \mathcal{N}(u(t_0 + \tau), t_0 + \tau) d\tau, \tag{4.18}$$

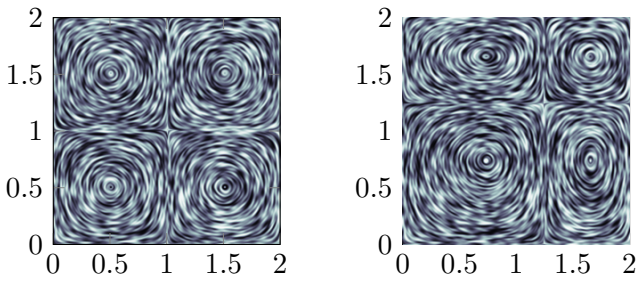

Figure 4.1: Quadruple gyre vector field at $t = 0$ and $t = 10.25$. The four gyres are separated by a horizontal and a vertical line, such that their intersection point moves on the diagonal.

with $h = t_1 - t_0$. A numerical scheme is derived by approximating the integral in (4.18), e.g. by a Runge-Kutta 4 type rule, resulting in a scheme called *etdrk4*. Note that D is a diagonal matrix. We use the version in [Kassam and Trefethen, 2005], which elegantly treats a cancellation problem occurring in a naive formulation of *etdrk4* by means of a contour integral approximated by the trapezoidal rule. (See Appendix for a motivation of the *etdrk4* scheme and the Matlab Code used).

4.6 Numerical experiments

4.6.1 Quadruple gyre

The first numerical example is a two dimensional flow (cf. Fig. 4.1), which is an extension of the double gyre flow (see Chapter 3.5.1), given by

$$\begin{aligned}\dot{x} &= -g(t, x, y)\\ \dot{y} &= g(t, y, x)\end{aligned}$$

on the 2-torus $[0, 2] \times [0, 2]$, where

$$g(t, x, y) = \pi \sin(\pi f(t, x)) \cos(\pi f(t, y)) \partial_x f(t, y)$$

and $f(t, x) = \delta \sin(\omega t)x^2 + (1 - 2\delta \sin(\omega t))x$. We fix $\delta = 0.25$, $\omega = 2\pi$, $t_0 = 0$, $t_1 = 10.25$, $h = 0.205$ (i.e. 50 time steps) and choose $\epsilon = 0.02$ in such a way, that the six largest singular values of P^ϵ roughly equal those obtained by Ulam's method (without explicit diffusion).

We use $M = 15$ collocation points and $N = 5$ basis functions in each direction, and compute the first four singular values and right singular vectors of P^ϵ. As shown in Figure 4.2 (top row), they reveal the gyres in their sign structures nicely. The computation of P^ϵ takes less than a second, the computation of all singular values and -vectors less than

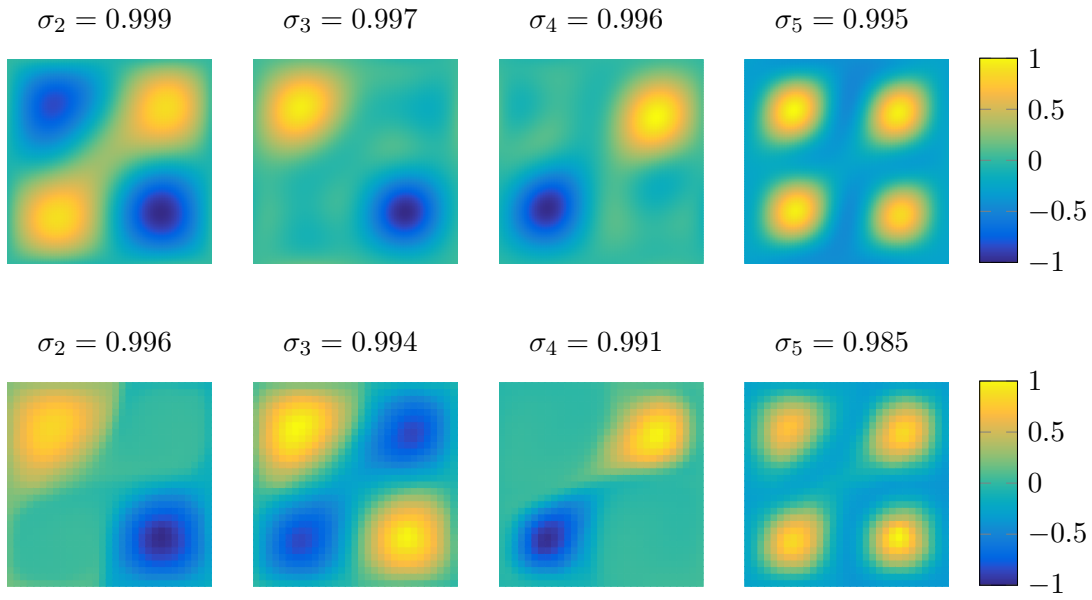

Figure 4.2: 2-th to 5-th right singular vector. Top row: computed via Fokker-Planck approach. Bottom row: computed via Ulam's method.

0.01 seconds[1]. For comparison, in the bottom row of Fig. 4.2, we show the same singular vectors computed via Ulam's method (without explicit diffusion) on a 32×32 box grid using 100 sample points per box. Here, the computation of the transition matrix takes around 5 seconds, the computation of the six largest singular values resp. vectors less than 0.2 seconds. In Figure 4.3, we show the singular values of the discretized transfer operator, and the coherent partition into five coherent sets.

4.6.2 A fluid flow

We now turn to a case where the vector field is only given on a discrete grid. Here, the approach proposed in Section 4.5 is particularly appealing, if we choose the grid points as collocation points (resp. a subset of them): In contrast to methods based on explicit integration of individual trajectories (such as Ulam's method), no further interpolation of the vector field is necessary. Depending on the initial point of a trajectory, the small scale structure of the turbulent vector field might enforce very small step sizes of the time integrator and, hence, makes the computation expensive.

For the experiment, we consider the incompressible Navier-Stokes equation with con-

[1]Computation times are measured on an 2.6 GHz Core i5 running Matlab R2015b.

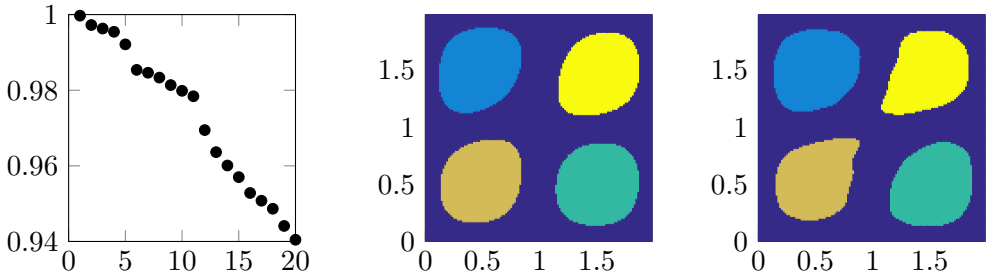

Figure 4.3: Singular values for the quadruple gyre and coherent partition at $t_0 = 0$ and $t_1 = 10.25$.

stant density on the 2-torus $\Omega = [0, 2\pi]^2$.

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} &= \nu \Delta \mathbf{v} - (\mathbf{v} \cdot \nabla)\mathbf{v} - \nabla p \\ \nabla \cdot \mathbf{v} &= 0, \end{aligned}$$

where $\mathbf{v}$ denotes the velocity field, p the pressure, and $\nu > 0$ the viscosity. Via introducing the vorticity $w := \nabla \times \mathbf{v}$, the Navier-Stokes equation in 2D can be rewritten as *vorticity equation*

$$\begin{aligned} \frac{Dw}{Dt} = \partial_t w + (\mathbf{v} \cdot \nabla)w &= \nu \Delta w \\ \Delta \psi &= -w. \end{aligned} \tag{4.19}$$

where the pressure p cancels from the equation. We can extract the velocity field $\mathbf{v}$ from the *streamline function* ψ via $\mathbf{v}_1 = \partial_y \psi$ and $\mathbf{v}_2 = -\partial_x \psi$.

The equation can be integrated by standard methods, e.g. a pseudo spectral method as proposed in [Nave, 2008]. For a first experiment, we choose an initial condition inducing three vortices, two with positive and one with negative spin, as initial condition:

$$w(0, x, y) = e^{-5\|(x,y)-(\pi,\frac{\pi}{4})\|_2^2} + e^{-5\|(x,y)-(\pi,-\frac{\pi}{4})\|_2^2} - \frac{1}{2} e^{-\frac{5}{2}\|(x,y)-(\frac{\pi}{4},\frac{\pi}{4})\|_2^2}.$$

We solve (4.19) on a grid with 64 collocation points in both coordinate direction. For the computation of coherent sets we choose $n = 16$ basis functions and $N = 32$ collocation points in both directions, as well as $t_0 = 0$ and $t_1 = 40$. We hence use only every second collocation point of the computed vector field. Because the induced coherent structures being way bigger in comparison to the grid size, this does not affect the result. The underresolution of the vector field can be interpreted as additional diffusion. We use $\epsilon = 10^{-2}$, which is of the same order as the grid resolution. In Figure 4.4, we show the vector field at time $t_0 = 0$ (left) as well as the second right singular vector (center) in the first row, and the vector field at $t_1 = 20$ (left) as well as the second left singular vector (center) in the second row. The computation took 35 seconds.

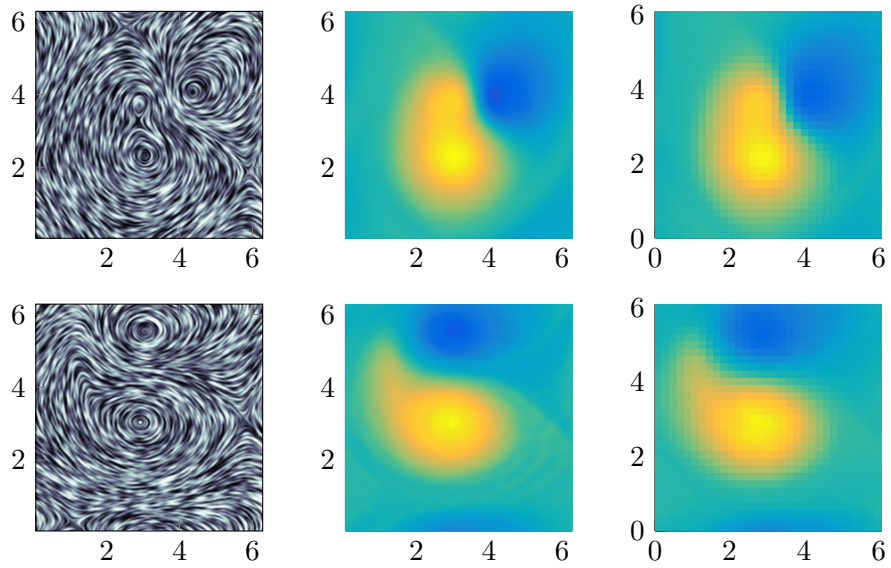

Figure 4.4: Top row: vector field at $t_0 = 0$ (left), second right singular vector computed via Fokker-Planck (center) and via Ulam (right). Bottom row: vector field at $t_1 = 20$ (left), second left singular vector computed via Fokker-Planck (center) and via Ulam (right).

For comparison, we show the same singular vectors computed via Ulam's method (right) on a 32×32 box grid using 100 sample points per box which were integrated by Matlab's `ode45`. Here, we need to interpolate the vector field between the grid points using splines (i.e. using `interp2` in Matlab). This computation also took 35 seconds.

In Figure 4.5, we show the spectrum of the discretized transfer operator. We extract three coherent sets and show them at initial and final time also in Figure 4.5.

In the second experiment, we use a turbulent initial condition by choosing a real number randomly in $[-1, 1]$ from a uniform distribution at each collocation point. For the coherent set computation, we restrict the time domain to $[t_0, t_1] = [20, 40]$ since then the initial vector field has smoothed somewhat, cf. Figure 4.7. Here, we choose $M = 32$, $N = 16$ and $\epsilon = 10^{-2}$. In Figure 4.7, we show the vector field, the computed second singular vectors via Fokker-Planck and via Ulam's method. For comparison, we show the same singular vectors computed via Ulam's method on a 32×32 box grid using 100 sample points per box. Here, we need to interpolate the vector field between the grid points using splines (i.e. using `interp2` in Matlab). Since the vector field is turbulent, using Matlab's `ode45` for the vectorized system is infeasible. We therefore choose a fixed time-step of $h = 0.01$, such that the result does not seem to change when further decreasing h. This computation took roughly 100 seconds.

In Figure 4.8, we show the singular values of the discretized transfer operator. We

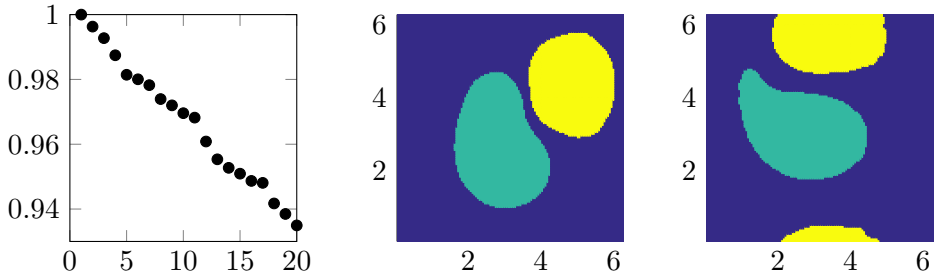

Figure 4.5: Singular values of the discretized transfer operator and a coherent partition at time $t_0 = 0$ and $t_1 = 40$.

decide to extract five coherent sets, and show the coherent partition at initial and final time also in Figure 4.7. In order to further visualize the dynamics, we also added the evolution of an arbitrary chosen (incoherent) set to the figure. The fifth singular vector indicates the (light blue) coherent set in the lower center. Hence, if we compute a 4-partition of the domain, we would get the same figure as in 4.7 without the light blue set.

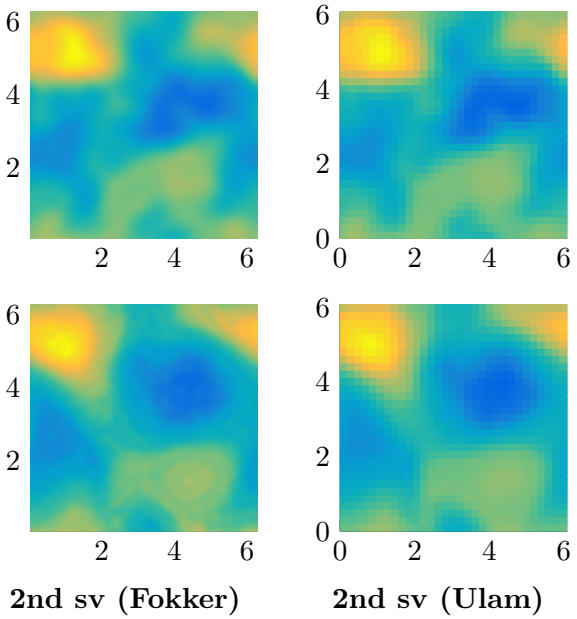

Figure 4.6: First row: second singular vector at time $t_0 = 20$ computed via Fokker-Planck and via Ulam's method, Second row: second singular vector at time $t_1 = 40$ computed via Fokker-Planck and via Ulam's method.

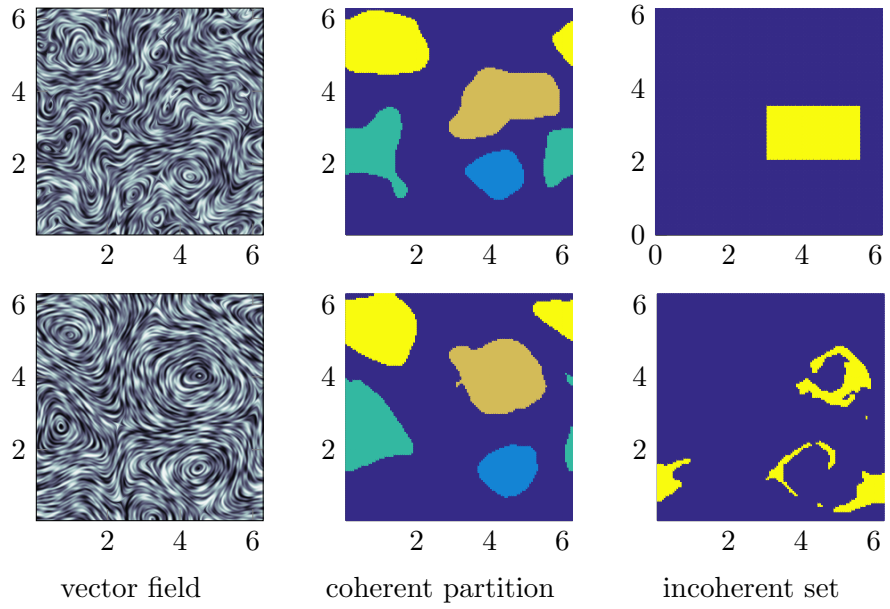

Figure 4.7: First row: Vector field, computed coherent five-partition, and an arbitrary chosen set at initial time $t_0 = 20$. Second row: Vector field, the evolution of the computed coherent five-partition and of the arbitrary chosen set at final time $t_1 = 40$.

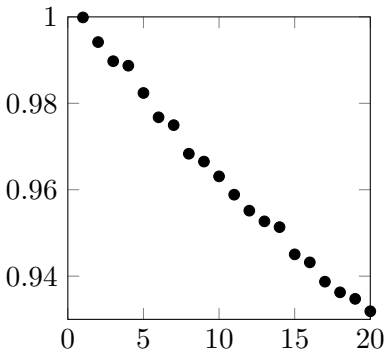

Figure 4.8: Singular values of the discretized transfer operator for the second fluid flow example.

Chapter 5

Transfer operator families

The methods for the computation of coherent sets introduced in Chapter 3 and 4 work well, but have in common that the used transfer operator $\mathcal{P}_\epsilon = \mathcal{P}_\epsilon^{t_0,t_1}$ only "sees" initial and final times $t = t_0$ and $t = t_1$. As a result, we do not directly know what happens to coherent sets during the time evolution. Especially Ulam's method (Section 3.4) does not take into consideration what happens during the evolution, as the diffusion added by the discretization works only at initial and final time. For the computation of the transfer operator, the positions of test particles (Ulam's method) or the evolution of densities (Fokker-Planck) are used only at final time — although all trajectories and densities in the entire time interval $[t_0, t_1]$ are computed. This has an additional drawback: All information obtained during the process is not used. Especially, if only limited data on the flow are available, this can be a problem.
In this chapter, we want to further generalize the concept of coherence to consider intermediate times, too. In our approach, we consider a family of transfer operators, parameterized by time. We use the spectral analysis of this family to extend the methods derived in Chapter 3 and 4. We furthermore discuss different discretizations, one of which leads to a recently developed, purely data-driven algorithm for determining coherent sets. Then we apply the method to several examples, including the Double gyre, the Bickley jet, and an oceanographic flow determined by data from surface drifters. This chapter is based on [Denner et al., 2017], which is in preparation.

In this chapter we again consider a time-dependent ordinary differential equation $\dot{x} = b(t, x)$, where $x(t)$ lies in some bounded domain $\Omega \subset \mathbb{R}^d$, and $b : \mathbb{R} \times \Omega \to \mathbb{R}^d$ is Lipschitz-continuous in x and continuous in t. We fix the initial time $t_0 = 0$ and some final time t_1, and assume that for every time $t \in [0, t_1]$ the associated flow map $\mathcal{T}^{0,t} : \Omega \longrightarrow \Omega$ given by

$$\mathcal{T}^{0,t}(x_0) = \mathcal{T}(0, t, x_0) = x(t; x_0)$$

is defined. We again consider a perturbed stochastic flow map $\mathcal{T}_\epsilon^{t_0,t_1}$ and a reference measure μ, which we assume to be invariant under $\mathcal{T}_\epsilon^{t_0,t_1}$. This holds for example, if the vector field b is divergence-free, and a Brownian motion is used as perturbation of the flow map.
We now introduce a generalization of the notion of coherence to a continuously parameterized family of sets $A_{[0,t_1]} = (A_t)_{t\in[0,t_1]} \subset \Omega$. Intuitively, for such a family of sets, we want to obtain an expression like

$$\frac{1}{t_1}\int_0^{t_1} \frac{\mu(A_0 \cap \mathcal{T}^{t,0}A_t)}{\mu(A_0)} dt \approx 1, \tag{5.1}$$

which is simply the time average of the notion of coherence made in (3.1). This translates to

$$\rho\left((A_t)_{t\in[0,t_1]}\right) := \frac{1}{t_1}\int_0^{t_1} \frac{\mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A,\cdot),\cdot)\big] \cap A\Big)}{\mu(A)}\, dt, \tag{5.2}$$

which is the generalization of the definition (4.5) used in Chapter 4. We use definition (5.2) for the remainder of this chapter.

5.1 Transfer operator families

The question now is, whether there is a canonical generalization of $\mathcal{P}_\epsilon$ and $\mathcal{K}_\epsilon$ which can be connected to (5.2). An intuitive approach is to interpret the stochastic flow map as

$$\begin{aligned}\mathcal{T}_\epsilon^{[0,t_1]} &: [0,t_1]\times\Omega \longrightarrow \Omega\\ &\mathcal{T}_\epsilon^{[0,t_1]}(t,x) := \mathcal{T}_\epsilon^{0,t}(x).\end{aligned}$$

A generalized Koopman operator can then be defined as

$$\begin{aligned}\mathcal{K}_\epsilon &: \mathbb{L}^2(\Omega) \longrightarrow \mathbb{L}^2([0,t_1],\mathbb{L}^2(\Omega))\\ &\mathcal{K}_\epsilon f = f\circ\mathcal{T}_\epsilon^{[0,t_1]}\\ &(\mathcal{K}_\epsilon f)(t) = f\circ\mathcal{T}_\epsilon^{0,t} = \mathcal{K}_\epsilon^t f,\end{aligned}$$

where $\mathbb{L}^2([0,t_1],\mathbb{L}^2(\Omega))$ denotes the space of square-integrable functions on the interval $[0,t_1]$ that take on values in $\mathbb{L}^2(\Omega)$. $\mathbb{L}^2([0,t_1],\mathbb{L}^2(\Omega))$ is called *Bochner space*, see Section 2.6. For a given density $f \in \mathbb{L}^2(\Omega)$, and some time instant $t \in [0,t_1]$, $\mathcal{K}_\epsilon^t$ hands back the initial density f_0, such that f is the image of f_0 under $\mathcal{T}_\epsilon^{0,t}$. In other words, as f_0 is the image of f under $\mathcal{T}^{t,0}$, $\mathcal{K}_\epsilon$ hands back the evolution of f at all times $t \in [0,t_1]$ under the time-inverted dynamics, see Figure 5.1. Hence, our Koopman operator can be interpreted as a family of Koopman operators $\mathcal{K}_\epsilon^t$, as defined in the sense of (4.2). In the following, we use the shorthand notation

$$\begin{aligned}\mathcal{F}_0 &:= \mathbb{L}^2(\Omega,\mu)\\ \mathcal{F}_{01} &:= \mathbb{L}^2([0,t_1],\mathbb{L}^2(\Omega,\mu)).\end{aligned}$$

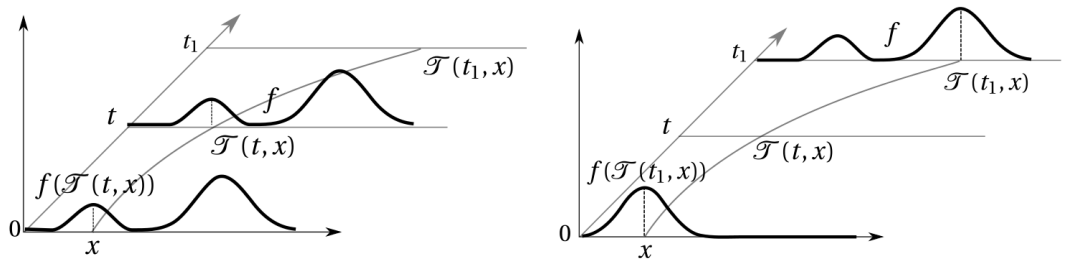

Figure 5.1: The Koopman operator pulls back f from times t and t_1.

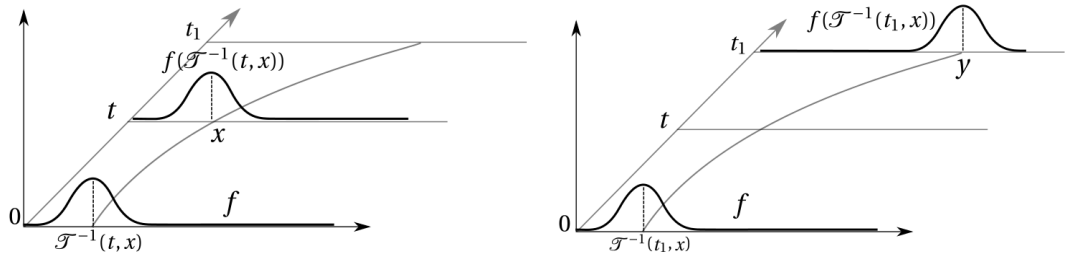

Figure 5.2: The Frobenius-Perron operator pushes forward f to times t and t_1.

The Bochner space $\mathcal{F}_{01}$ can be equipped with an inner product

$$\langle f, g\rangle_{\mathcal{F}_{01}} = \frac{1}{t_1}\int_0^{t_1} \langle f(t), g(t)\rangle_{\mathcal{F}_0} dt,$$

leading to a norm

$$\|f\|_{\mathcal{F}_{01}} = \frac{1}{\sqrt{t_1}}\left(\int_0^{t_1} \|f(t)\|^2_{\mathcal{F}_0}\ dt\right)^{1/2}.$$

It, hence, is simply isomorphic to $\mathbb{L}^2([0,t_1]\times\Omega)$. Note that $\|1_{[0,t_1]\times\Omega}\|_{\mathcal{F}_{01}} = \sqrt{\mu(\Omega)} = \|1_\Omega\|_{\mathcal{F}_0}$. We further denote by μ the Lebesgue measure on $[0,t_1]\times\Omega$.

A similar generalization applies to the Frobenius-Perron operator: we define $\mathcal{P}_\epsilon$ by

$$\begin{aligned} \mathcal{P}_\epsilon &: \mathcal{F}_0 \longrightarrow \mathcal{F}_{01} \\ \mathcal{P}_\epsilon(f_0)(t) &= \mathcal{P}_\epsilon^t f_0, \end{aligned} \tag{5.3}$$

which may again be regarded as a family of Frobenius-Perron operators defined in the traditional sense. Here, the interpretation is straightforward: for a given $f_0 \in \mathcal{F}_0$ and every time instant t, $\mathcal{P}_\epsilon$ hands back the image of f_0 under $\mathcal{T}_\epsilon^{0,t}$, i.e., $\mathcal{P}_\epsilon$ hands back the whole evolution of f_0 at all times, see Figure 5.2.

To summarize, so far we have defined generalizations of the Frobenius-Perron and Koopman operators as families of pointwise (in time) operators. Note that in this setting $\mathcal{P}_\epsilon^t$ and $\mathcal{K}_\epsilon^t$ are dual at each fixed t, but $\mathcal{P}_\epsilon$ and $\mathcal{K}_\epsilon$ are no dual operators.

However, we can formally compute the dual of $\mathcal{P}_\epsilon$:

$$\begin{aligned} t_1\langle \mathcal{P}_\epsilon f, g\rangle_{\mathcal{F}_{01}} &= \int_0^{t_1} \langle \mathcal{P}_\epsilon^t f, g(t)\rangle_{\mathcal{F}_0} dt \\ &= \int_0^{t_1} \langle f, \mathcal{K}_\epsilon^t g(t)\rangle_{\mathcal{F}_0} dt \\ &= \int_0^{t_1} \int_\Omega f(x)(\mathcal{K}_\epsilon^t g(t))(x) dx\, dt \\ &= \int_\Omega f(x) \int_0^{t_1} (\mathcal{K}_\epsilon^t g(t))(x) dt\, dx \\ &= \left\langle f, \int_0^{t_1} \mathcal{K}_\epsilon^t g(t)\, dt \right\rangle_{\mathcal{F}_0}. \end{aligned}$$

Hence, the formal dual $\hat{\mathcal{K}}_\epsilon := \mathcal{P}_\epsilon^*$ is

$$\begin{aligned} \hat{\mathcal{K}}_\epsilon &: \mathcal{F}_{01} \longrightarrow \mathcal{F}_0 \\ \hat{\mathcal{K}}_\epsilon g &= \frac{1}{t_1} \int_0^{t_1} \mathcal{K}_\epsilon^t g(t)\, dt, \end{aligned} \tag{5.4}$$

i.e., the time-average over all images of $\mathcal{K}_\epsilon^t$.

Analogously, we can derive $\hat{\mathcal{P}}_\epsilon := \mathcal{K}_\epsilon^*$ as

$$\begin{aligned} \hat{\mathcal{P}}_\epsilon &: \mathcal{F}_{01} \longrightarrow \mathcal{F}_0 \\ \hat{\mathcal{P}}_\epsilon g &= \frac{1}{t_1} \int_0^{t_1} \mathcal{P}_\epsilon^t g(t)\, dt. \end{aligned}$$

Note that in this case $\hat{\mathcal{P}}_\epsilon$ is also defined via

$$\langle \hat{\mathcal{P}}_\epsilon g, f\rangle_{\mathcal{F}_0} = \langle g, f \circ \mathcal{T}_\epsilon^{[0,t_1]}\rangle_{\mathcal{F}_{01}} \quad \forall f \in \mathcal{F}_0.$$

For the transfer operators $\mathcal{P}_\epsilon$ and $\mathcal{K}_\epsilon$ holds that $\mathcal{P}1_\Omega = \mathcal{K}_\epsilon 1_\Omega = 1_{[0,t_1]\times\Omega}$ and $\|\mathcal{P}_\epsilon\|_{\mathcal{F}_0\to\mathcal{F}_{01}} = \|\mathcal{K}_\epsilon\|_{\mathcal{F}_0\to\mathcal{F}_{01}} = \|\hat{\mathcal{K}}_\epsilon\|_{\mathcal{F}_{01}\to\mathcal{F}_0} = \|\hat{\mathcal{P}}_\epsilon\|_{\mathcal{F}_{01}\to\mathcal{F}_0} = 1$.

5.1.1 Computing coherent sets using transfer operator families

In this section we want to use the transfer operator families in order to compute coherent sets as stated in section 5.1. We start by establishing a connection between (5.2) and the Frobenius-Perron operator (5.3):

$$\begin{aligned} \left\langle \hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon 1_{A_0}, 1_{A_0} \right\rangle_{\mathcal{F}_0} &= \frac{1}{t_1} \int_0^{t_1} \left\langle \mathcal{K}_\epsilon^t \mathcal{P}_\epsilon^t 1_{A_0}, 1_{A_0} \right\rangle_{\mathcal{F}_0} dt \\ &= \frac{1}{t_1} \int_0^{t_1} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A_0,\cdot),\cdot)\big] \cap A_0 \Big)\, dt \end{aligned} \tag{5.5}$$

where we used (4.4) for the second step.

The following works analogous to the previous chapters: We want to compute a *coherent partition* of the domain $\Omega \subset \mathbb{R}^d$ into n *coherent sets*. The number n of those sets however is again not known a priori. So we try to find a formulation, which shows the partition into n coherent sets for several $n \geq 1$, and choose a meaningful number from that. To this end, we formulate the optimization problem for $A_1, \dots, A_n \in \mathcal{A}$

$$\max_{\Omega = \dot{\bigcup} A_k} \sum_{k=1}^{n} \frac{1}{t_1} \int_0^{t_1} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A_k,\cdot),\cdot) \Big] \;\Big|\; A_k \Big)\, dt \tag{5.6}$$

which we interpret in $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$, demanding that the coherent sets A_k form a partition of the domain and that they keep their mass under the evolution. We can reformulate (5.6) as

$$\begin{aligned}
&\max_{\Omega = \dot{\bigcup} A_k} \sum_{k=1}^{n} \frac{1}{t_1} \int_0^{t_1} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A_k,\cdot),\cdot) \Big] \;\Big|\; A_k \Big)\, dt \\
=& \max_{\Omega = \dot{\bigcup} A_k} \sum_{k=1}^{n} \frac{\left\langle \hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon 1_{A_0}, 1_{A_0} \right\rangle_{\mathcal{F}_0}}{\|1_{A_k}\|_{\mathcal{F}_0}^2} \\
\leq& \max_{f_1 \perp \dots \perp f_n \in \mathcal{F}_0} \sum_{k=1}^{n} \frac{\left\langle \hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon f_k, f_k \right\rangle_{\mathcal{F}_0}}{\|f_k\|_{\mathcal{F}_0}^2} \\
=& \max_{f_1 \perp \dots \perp f_n \in \mathcal{F}_0} \sum_{k=1}^{n} \left\langle \hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon \left(\frac{f_k}{\|f_k\|_{\mathcal{F}_0}} \right), \frac{f_k}{\|f_k\|_{\mathcal{F}_0}} \right\rangle_{\mathcal{F}_0} \\
=& \max_{f_1,\dots,f_n \text{ ONB of } V_n \leq \mathbb{L}^2(\Omega,\mu)} \sum_{k=1}^{n} \langle \hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon f_k, f_k \rangle_{\mathcal{F}_0} \\
=& \sum_{k=1}^{n} \lambda_k(\hat{\mathcal{K}}_\epsilon \mathcal{P}_\epsilon) \\
=& \sum_{k=1}^{n} \sigma_k(\mathcal{P}_\epsilon)^2
\end{aligned}$$

due to the Courant-Fischer Theorem 8. In order to apply the theorem, again we have to assume that the diffusion in the system makes the transfer operator $\mathcal{P}_\epsilon$ compact (or at least, that it has n singular values below the bottom of the essential spectrum). The maximizing functions are the corresponding first n singular functions of $\mathcal{P}_\epsilon$.

Remark. *Note that if we consider the probability space $([0,t_1] \times \Omega, \mathcal{B} \times \mathcal{A}, \pi)$ with the probability measure $\pi(C^{[0,t_1]}) = 1/t_1 \int_0^{t_1} \mu(C^t)\, dt$ for all $C \in \mathcal{B} \times \mathcal{A}$, we can interpret our notion of coherence again as conditional probability*

$$\frac{1}{t_1} \int_0^{t_1} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A_k,\cdot),\cdot) \Big] \;\Big|\; A_k \Big)\, dt = \pi\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t,0}(\mathcal{T}_\epsilon^{0,t}(A_k,\cdot),\cdot) \Big] \;\Big|\; A_k \Big).$$

5.1.2 Compactness of $\mathcal{P}_\epsilon$

In the previous section we discussed how to compute indicator functions of coherent sets via singular value decomposition, by adding some small random perturbations to the system, such that the transfer operator $\mathcal{P}_\epsilon$, $\mathcal{P}_\epsilon(t) := \mathcal{P}_\epsilon^t$ of the such perturbed system is compact. The same problem occurs in the classic setting for $\mathcal{P}^{t_1}$, and there are well-known perturbations like *Ulam's method* (Chapter 3.4) or *Brownian motion* (Chapter 4) which have the desired properties. The question is the following: if we incorporate such a perturbation that renders $\mathcal{P}_\epsilon^t$ compact for every $t \in (0, t_1]$, will then $\mathcal{P}_\epsilon$ also be compact? In other words, is a family of pointwise (in time) compact operators compact? The answer is yes, as established by the following theorem.

Theorem 11. *If we apply a perturbation to the system, such that $\mathcal{P}_\epsilon^t$ is compact for any $t \in (0, t_1]$, $\mathcal{P}_\epsilon^0$ is bounded and im$(\mathcal{P}_\epsilon)$ is equicontinuous on $(0, t_1]$, then $\mathcal{P}_\epsilon : \mathcal{F}_0 \to \mathcal{F}_{01}$ is compact.*

Proof. We want to use the Arzelà Ascoli Theorem, but have the problem that $\mathcal{P}^{\epsilon,0}$ (which is the identity, if no perturbation is applied at time $t = 0$) is not necessarily compact. As a first step, we fix $\delta > 0$, and define $\mathcal{F}_{\delta 1} := \mathbb{L}^2([\delta, t_1], \mathbb{L}^2(\Omega))$ and $\mathcal{P}_\epsilon^\delta : \mathcal{F}_0 \to \mathcal{F}_{\delta 1}$ as the restriction of $\mathcal{P}_\epsilon$ to $[\delta, t_1]$.

The Arzelà Ascoli Theorem states, that a subspace $F \subset C([\delta, t_1], \mathbb{L}^2(\Omega))$ is relatively compact with respect to the topology of uniform convergence, if F is equicontinuous and for every t in $[\delta, t_1]$ holds that $\{f(t) : f \in F\}$ is relatively compact in $\mathbb{L}^2(\Omega)$, see e.g. [Hirzebruch and Scharlau, 1971].

We assumed, that im$(\mathcal{P}_\epsilon)$ is equicontinuous on $(0, t_1]$ and, hence, on $[\delta, t_1]$. Therefore $\mathcal{P}_\epsilon^\delta$ maps in particular to $C([\delta, t_1], \mathbb{L}^2(\Omega)) \cap \mathcal{F}_{01}$. Consider a bounded set A in $\mathbb{L}^2(\Omega)$. As $\mathcal{P}_\epsilon^t$ is compact for all $t \in [\delta, t_1]$, $\mathcal{P}_\epsilon^t A$ is relatively compact in $\mathbb{L}^2(\Omega)$. Hence, due to the Arzelà Ascoli Theorem, $\mathcal{P}_\epsilon^\delta$ is compact as a mapping $\mathbb{L}^2(\Omega) \to C([\delta, t_1], \mathbb{L}^2(\Omega)) \cap \mathcal{F}_{01}$ with respect to $\|\cdot\|_\infty$.

We now show, as a second step, that $\mathcal{P}_\epsilon^\delta$ is also compact as a mapping from $\mathbb{L}^2(\Omega) \to \mathbb{L}^2([\delta, t_1], \mathbb{L}^2(\Omega))$. To this end, let $(u_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{L}^2(\Omega)$. As $\mathcal{P}_\epsilon^\delta$ is compact, $\mathcal{P}_\delta^\epsilon u_n$ has a converging subsequence $\mathcal{P}_\epsilon^\delta(u_n)_k$ with respect to $\|\cdot\|_\infty$. As $\mu([\delta, t_1] \times \Omega)$ is finite, $\mathcal{P}_\epsilon^\delta(u_n)_k$ also converges with respect to $\|\cdot\|_2$. Consequently, $\mathcal{P}_\epsilon^\delta$ is sequentially compact and, hence, compact on $\mathbb{L}^2([\delta, t_1], \mathbb{L}^2(\Omega))$.

This argument holds for all $\delta > 0$, but not for $\delta = 0$, as $\mathcal{P}_\epsilon^0 = \mathrm{Id}$ is not compact and, hence, the Arzelà Ascoli Theorem can not be applied. As a third step, we formulate $\mathcal{P}_\epsilon$ as limit of the compact operators $\mathcal{P}_\epsilon^\delta$ for $\delta \to 0$. To this end, we embed $\mathcal{P}_\epsilon^\delta$ into $\mathcal{F}_{01}$ via

$$\tilde{\mathcal{P}_\epsilon^\delta} : \mathcal{F}_0 \to \mathcal{F}_{01}$$
$$(\tilde{\mathcal{P}_\epsilon^\delta} f)(t) = \begin{cases} \mathcal{P}_\epsilon^t & t \in [\delta, t_1] \\ 0 & t \in [0, \delta) \end{cases}.$$

It is easy to show that $\tilde{\mathcal{P}}_\epsilon^\delta$ is still compact. Furthermore

$$\begin{aligned}
&\|\tilde{\mathcal{P}}_\epsilon^\delta - \mathcal{P}_\epsilon\|^2_{\mathcal{F}_0 \to \mathcal{F}_{01}} = \frac{1}{t_1} \sup_{\|f\|_{\mathcal{F}_{01}}=1} \|(\tilde{\mathcal{P}}_\epsilon^\delta - \mathcal{P}_\epsilon) f\|^2_{\mathcal{F}_{01}} \\
=&\frac{1}{t_1} \sup_{\|f\|_{\mathcal{F}_{01}}=1} \int_0^{t_1} \|(\tilde{\mathcal{P}}_\epsilon^\delta - \mathcal{P}_\epsilon) f(t)\|_{\mathcal{F}_0}\, dt \\
=&\frac{1}{t_1} \sup_{\|f\|_{\mathcal{F}_{01}}=1} \int_0^{\delta} \|\mathcal{P}_\epsilon^t f\|_{\mathcal{F}_0}\, dt \\
\leq&\frac{1}{t_1} \sup_{\|f\|_{\mathcal{F}_{01}}=1} \int_0^{\delta} \|\mathcal{P}\|_{\mathcal{F}_0 \to \mathcal{F}_{01}} \|f\|_{\mathcal{F}_0}\, dt \\
\leq&C\delta \to 0 \ (\delta \to 0),
\end{aligned}$$

where $C = 1/t_1 \sup_{t\in[0,\delta]} \|\mathcal{P}\|_{\mathcal{F}_0 \to \mathcal{F}_{01}} < \infty$, as $\mathcal{P}_\epsilon^t$ is bounded for all $t \in [0, t_1]$. Hence, $\mathcal{P}_\epsilon^t$ is compact as limit of compact operators. □

5.2 Time-continuous diffusion

In this section, we consider a hyperrectangle $\Omega \subset \mathbb{R}^d$, $b : [0, t_1] \times \Omega \to \Omega$ Lipschitz and periodic in x. We add a continuous-time diffusion via Brownian motion ξ_t

$$\dot{x} = b(t, x) + \epsilon \xi_t$$

and obtain the associated Fokker-Planck equation

$$\begin{aligned}
\partial_t u &= \frac{\epsilon^2}{2} \triangle u - \operatorname{div}(bu) \\
u(0) &= u_0 \in \mathbb{L}^2(\Omega)
\end{aligned} \tag{5.7}$$

with periodic boundary conditions as presented in Chapter 4, where $\triangle$ denotes the Laplacian. The transfer operator $\mathcal{P}_\epsilon : \mathcal{F}_0 \to \mathcal{F}_{01}$ is the family of solution operators $\mathcal{P}_\epsilon^t,\ t \in [0, t_1]$, of (5.7). In Lemma 6, we showed that $\mathcal{P}_\epsilon^t$ is compact for every $t > 0$.

Lemma 8. *For the family of solution operators $\mathcal{P}_\epsilon$ of (5.7) holds that $im(\mathcal{P}_\epsilon)$ is equicontinuous on $(0, t_1]$.*

Proof. $\mathcal{P}_\epsilon$ is the family of solution operators of (5.7) and, hence, $\mathcal{P}^{t,\epsilon} u_0$ is continuously differentiable for every $u_0 \in \mathcal{F}_0$ and every $t \in (0, t_1]$ with a common upper bound $\|b\|_\infty$. Therefore $\mathcal{P}_\epsilon$ is equicontinuous in $(0, t_1]$. □

We can now apply Theorem 11 and obtain that $\mathcal{P}_\epsilon$ is compact.

5.2.1 Spectral discretization

In order to get a finite-dimensional approximation of $\mathcal{P}_\epsilon$, we use spectral collocation presented in 4.5. In time we simply use constant ansatz functions leading to basis functions φ and ψ of the approximation spaces $V_0^n \subset \mathcal{F}_0$ and $V_{01}^N \subset \mathcal{F}_{01}$:

$$\begin{aligned} \varphi_k(x) &= e^{i\langle k,x\rangle}, && |k| \leq (n-1)/2,\ n \text{ odd} \\ \psi_{mk}(t,x) &= \sqrt{\frac{t_1-0}{M}} e^{i\langle k,x\rangle} 1_{(s_{m-1},s_m]}, && |k| \leq (N-1)/2,\ N \text{ odd},\ m = 1,\dots M \\ & && 0 = s_0 < s_1 < \cdots < s_M = t_1. \end{aligned}$$

where we choose the normalization such that $\|\psi_{mk}\|_{\mathcal{F}_{01}} = \|\varphi_k\|_{\mathcal{F}_0} = 1$. The matrix P_ϵ takes the form

$$\begin{aligned} P_\epsilon &: V_0^n \longrightarrow V_{01}^N \\ P_\epsilon &= \sqrt{\frac{t_1}{M}} \begin{pmatrix} P_{\epsilon,s_1} \\ \vdots \\ \mathcal{P}_{\epsilon,s_M} \end{pmatrix} \in \mathbb{R}^{MN\times n} \end{aligned}$$

where $P^{\epsilon,s_k} : V_0^n \to V_0^N$ is the approximation of $\mathcal{P}^{\epsilon,s_k}$ as given in Chapter 4.5.

5.2.2 A fluid example

As numerical example, we reconsider the incompressible Navier-Stokes equation (see also Section 4.6.2) with constant density on the 2-torus $X = [0, 2\pi]^2$,

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} &= -(\mathbf{v}\cdot\nabla)\mathbf{v} - \nabla p + \nu\triangle\mathbf{v} \\ \nabla\cdot\mathbf{v} &= 0, \end{aligned}$$

where $\mathbf{v}$ denotes the velocity field, p the pressure, and $\nu > 0$ the kinematic viscosity Defining the vorticity $\omega := \nabla\times\mathbf{v}$ and writing the (two-dimensional) velocity in terms of a streamfunction ψ as $\mathbf{v} = (\partial_y\psi, -\partial_x\psi)$, the equation can be rewritten as the *vorticity equation*

$$\begin{aligned} \frac{\partial \omega}{\partial t} &= -(\mathbf{v}\cdot\nabla)\omega + \nu\triangle\omega \\ \triangle\psi &= -\omega, \end{aligned}$$

in which the pressure no longer appears. The equation can be integrated by standard methods, e.g. a pseudo spectral method as proposed in [Nave, 2008]. We set $\nu = 10^{-3}$ and choose an initial condition inducing three vortices, two with positive and one with negative spin:

$$\omega(0,x,y) = e^{-5\|(x,y)-(\pi,\frac{\pi}{4})\|_2^2} + e^{-5\|(x,y)-(\pi,-\frac{\pi}{4})\|_2^2} - \frac{1}{2}e^{-\frac{5}{2}\|(x,y)-(\frac{\pi}{4},\frac{\pi}{4})\|_2^2}.$$

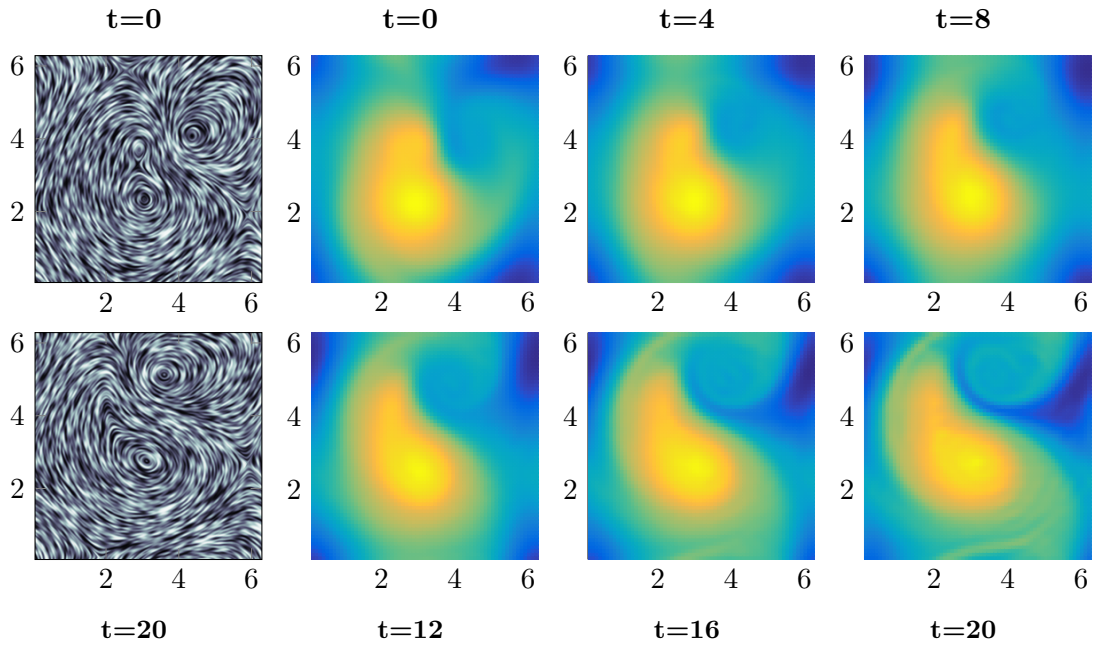

Figure 5.3: First column: vector field at $t_0 = 0$ and $t_1 = 20$. Other columns: Time-slices of the second left singular vector. Yellow (blue) indicates 1 (-1).

We solve (5.8) on a grid with 64 collocation points in both coordinate directions. For the computation of coherent sets, we choose $n = 16$ basis functions for V_0 and $N = 32$ collocation points in both directions, as well as $t_0 = 0$ and $t_1 = 20$ and $\delta t = 1$ for the time discretization. We use $\epsilon = 10^{-2}$, which is of the same order as the grid resolution. Assembling P_ϵ does not take longer than assembling $P_\epsilon^{t_1}$, as we have to evolve the basis functions φ of V_0^n anyway. The singular value decomposition takes longer, because the matrix P_ϵ is larger. Here, the computation took 120 seconds. In Figure 5.3 we show the vector field at the initial time 0 and the final time t_1, and snapshots of the second left singular vector $v_2 \in V_{01}^N$ at times $t = 0, 4, 8, 12, 16, 20$.

5.3 Data-driven discretizations

One motivation for the use of families of transfer operators is to be able to use data at all intermediate times. This is beneficial, if the transfer operator at time t_1 equals the identity, or in applications where only very limited information in space but at several times are available, as in the experimental dataset we consider in Section 5.4.4.

We consequently assume, that we do not know the underling flow map $\mathcal{T}_\epsilon^{[t_0,t_1]}$, but only N observed discrete trajectories of particles $T_i^k,\ i = 1,\dots,N,\ k = 1\dots M$ starting at initial points $x_i,\ i = 1,\dots,N$ and having their positions measured at times $s_k = (k-1)\Delta t,\ k = 1,\dots,M,\ \Delta t = t_1/(M-1)$, of a slightly perturbed flow $\mathcal{T}_\epsilon^{[0,t_1]}$. The

flow $\mathcal{T}_\epsilon^{[0,t_1]}$ is perturbed either by measurement errors, numerical dissipation, see e.g. [Froyland, 2013] or by force, see e.g. [Williams et al., 2014a]. In other words, we have access to only one realization $\mathcal{T}_\epsilon^{[t_0,t_1]}(x_i,\tilde{\xi})$ of the stochastic flow map, and that also only for the initial points $x_i,\ i=1,\dots,N$. Hence we can compute only a rough approximation to the Koopman operator of the stochastic flow:

$$\begin{aligned}\mathcal{K}_\epsilon^t f(x_i) &= \mathbb{E}_{\mathbb{P}}[f\circ\mathcal{T}_\epsilon^{0,t_1}(x_i,\cdot)] = \int_\Sigma f\circ\mathcal{T}_\epsilon^{0,t_1}(x_i,\xi)\ d\mathbb{P}(\xi)\\ &\approx \mathbb{P}(\Sigma)\left(f\circ\mathcal{T}_\epsilon^{0,t_1}(x_i,\tilde{\xi})\right) = f\circ\mathcal{T}_\epsilon^{0,t_1}(x_i,\tilde{\xi})\end{aligned} \tag{5.8}$$

which is the deterministic Koopman operator $\mathcal{K}^t$ for the realization of the flow.

We compute an approximation of the Koopman operator $\hat{\mathcal{K}}_\epsilon$ as given in (5.4) of the form

$$\begin{aligned}&\hat{K}_\epsilon : V_{01} \longrightarrow V_0\\ &\hat{K}_\epsilon = \begin{pmatrix} | & | & & | \\ R\hat{\mathcal{K}}_\epsilon\psi_1 & R\hat{\mathcal{K}}_\epsilon\psi_2 & \dots & R\hat{\mathcal{K}}_\epsilon\psi_N \\ | & | & & | \end{pmatrix},\end{aligned}$$

where

$$R : \mathbb{L}^2(\Omega) \to V_0$$

is the $\mathbb{L}^2-$projection on the approximation space $V_0 \subset \mathcal{F}_0$. As ansatz functions for V_0, we use the indicator functions 1_{Ω_i} of *Voronoi cells* Ω_i containing all points in Ω closest to x_i in Euclidean distance. Note that the constant function 1_Ω is contained in V_0. $\operatorname{span}(\psi_1,\dots,\psi_N) = V_{01} \subset \mathcal{F}_{01}$ is the approximation space for $\mathcal{F}_{01}$ with basis functions $\psi_i,\ i=1,\dots,N$. The $\mathbb{L}^2-$projection $R : \mathcal{F}_0 \to V_0$ is then given by

$$g \to \sum_{i=1}^N \frac{1}{|\Omega_i|}\left(\int_{\Omega_i} g\right) 1_{\Omega_i}.$$

As V_0 is finite dimensional, the projection R is compact and as $\hat{\mathcal{K}}_\epsilon$ is bounded, $R\hat{\mathcal{K}}_\epsilon$ is compact, too. Similar to Ulam's method, R can be interpreted as uniform distribution on the chosen Voronoi cells making the discretized operator $\hat{K}_\epsilon$ compact. We compute

$$\begin{aligned}\hat{\mathcal{K}}_\epsilon\psi_j &= \frac{1}{t_1}\int_0^{t_1} \mathcal{K}_\epsilon^t\psi_j(t)dt \overset{(5.8)}{\approx} \frac{1}{t_1}\int_0^{t_1} \psi_j(t)\circ\mathcal{T}_\epsilon^{0,t}(.,\tilde{\xi})dt\\ &= \frac{1}{t_1}\int_0^{t_1} \psi_j(t,\mathcal{T}_\epsilon^{0,t}(\cdot,\tilde{\xi}))dt\end{aligned} \tag{5.9}$$

and furthermore

$$\begin{aligned} R\hat{\mathcal{K}}_\epsilon \psi_j(x) &= R\left(\frac{1}{t_1}\int_0^{t_1} \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x,\tilde{\xi}))\ dt\right) \\ &= \sum_{i=1}^{N}\left(\frac{1}{t_1|\Omega_i|}\int_{\Omega_i}\int_0^{t_1} \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x,\tilde{\xi}))\ dt\ dx\right) 1_{\Omega_i} \\ &= \sum_{i=1}^{N}\left(\frac{1}{t_1|\Omega_i|}\int_0^{t_1}\int_{\Omega_i} \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x,\tilde{\xi}))\ dx\ dt\right) 1_{\Omega_i}. \end{aligned} \tag{5.10}$$

As we do not have any information on Ω_i at time 0 except the trajectory originating from x_i, we apply a *one-point rule*:

$$\approx \sum_{i=1}^{N}\left(\frac{1}{t_1}\int_0^{t_1}\frac{1}{|\Omega_i|}|\Omega_i|\ \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi}))\ dt\right) 1_{\Omega_i}, \tag{5.11}$$

and as we only have information at times $s_k,\ k = 1\ldots, M$, we approximate further

$$\approx \sum_{i=1}^{N}\frac{1}{t_1}\left(\sum_{k=1}^{M}\Delta t \psi_j(s_k, \mathcal{T}_\epsilon^{0,s_k}(x,\tilde{\xi}))\right) 1_{\Omega_i}.$$

Hence, the ij-th entry of $\hat{K}_\epsilon$ is

$$\hat{K}_{\epsilon,ij} = \frac{1}{t_1}\sum_{k=1}^{M}\Delta t \psi_j(s_k, \mathcal{T}_\epsilon^{0,s_k}(x_i,\tilde{\xi})). \tag{5.12}$$

The approximations we made are $\mathcal{O}(\Delta t)$ and $\mathcal{O}(h)$, where $h = \max_i \min_j(\|x_i - x_j\|_{\mathbb{R}^d})$ is the fill distance of the initial points.
We now have to choose basis functions ψ_j of V_{01}. As we only have several data trajectories given, a natural choice are radial basis functions $\phi : [0,\infty] \to [0,\infty]$ of a meshfree discretization, see [Wendland, 2004] [Fasshauer, 2007] , e.g.

- distance functions $\psi(r) = \|r\|_2$,
- Gaussians, $\phi(r) = \exp(-r^2)$,
- Wendland functions, e.g.
  $\phi(r) = \max(1-r,0)^6\ (35r^2 + 18r + 3)/3$ for the Wendland 3,2 function, which is positive definite on $\mathbb{R}^3$ and has compact support $[0,1]$.

Our basis functions then read $\psi_j(t,x) = \phi(\alpha\|\mathcal{T}(t,x_k) - x\|_2)$, where α is the scaling parameter, determining how fast the basis functions decay. For Wendland functions, the support of the basis function is then $1/\alpha$.

If we use simple distance functions, the discretization leads to the matrix $\hat{K}_\epsilon$, $\hat{K}_{\epsilon,ij} = \frac{1}{t_1}\sum_{k=1}^{M} \Delta t \|T_j^k - T_i^k\|_2$ occuring in [Froyland and Padberg-Gehle, 2015]. There, $\hat{K}_\epsilon$ is interpreted as adjacency matrix of a graph whose nodes are the trajectories T_j and whose edge weights are the distances in space-time of those trajectories. If Gaussians are used, the similarity matrix $\hat{K}_\epsilon$, $\hat{K}_{\epsilon,ij} = \frac{1}{t_1}\sum_{k=1}^{M} \Delta t \ \exp(-\epsilon^2 \|T_j^k - T_i^k\|_2^2)$ from [Hadjighasem et al., 2016] is derived.

For all those choices of basis functions the constant density $1_{[0,t_1]\times\Omega}$ is not contained in V_{01} and the approximated operator $\hat{K}_\epsilon$ is not stochastic. To overcome this we consider the normalized operator family $\tilde{\mathcal{P}}_\epsilon$, $\tilde{\mathcal{P}}_\epsilon f = \mathcal{P}_\epsilon(f1_\Omega)/\mathcal{P}_\epsilon 1_\Omega$, which is the natural transfer operator family when analyzing systems induced by vector fields with non-zero divergence (see chapter 3.4.1). For divergence free vector fields and reference measure $\mu = \lambda$, the Lebesgue measure, $\tilde{\mathcal{P}}_\epsilon$ equals $\mathcal{P}_\epsilon$. In order to obtain an approximation to $\tilde{\mathcal{P}}_\epsilon$, we first compute $P_\epsilon = \hat{K}'_\epsilon$ (first discretize, then adjungate) and normalize the matrix P_ϵ to $\tilde{P}_\epsilon = P_\epsilon / P_\epsilon 1_\Omega$, as $1_\Omega = \sum_{i=1}^{N} 1_{\Omega_i}$.

5.4 Numerical experiments

5.4.1 Double gyre

We now look at the benchmark example of the Double gyre, see also Section 3.5.1. The model describes a flow with stream function

$$\Psi(t,x,y) := A \sin(\pi f(t,x)) \sin(\pi y)$$

where $f(t,x) = \delta \sin(\xi t)x^2 + (1 - 2\delta \sin(\xi t))x$, in a domain $\Omega = [0,2]\times[0,1]$. The velocity field is given by

$$b(t,x,y) = \begin{pmatrix} -\frac{\partial \Psi}{\partial y} \\ \frac{\partial \Psi}{\partial x} \end{pmatrix} = \begin{pmatrix} -\pi A \sin(\pi f(t,x)) \cos(\pi y) \\ \pi A \cos(\pi f(t,x)) \sin(\pi y) \frac{\partial f}{\partial x}(t,x) \end{pmatrix}. \tag{5.13}$$

This model is not intended to describe a real fluid flow but a simplified Double gyre pattern seen in geophysical flows [Shadden et al., 2005].

The example describes two counter-rotating vortices separated by a periodically moving leaky transport barrier between them. For $\delta = 0$ the flow is autonomous. For $\delta > 0$, the flow is non-autonomous, and the gyres conversely expand and contract periodically in the x-direction such that the domain Ω enclosing the gyres remains invariant. A determines the magnitude of the vector field b, $\xi/2\pi$ is the frequency of oscillation, and δ is approximately how far the line separating the gyres moves to the left and to the right, see [Shadden et al., 2005]. We fix the parameter values $A = 0.25$, $\delta = 0.25$ and $\xi = 2\pi$.

In our configuration the Double gyre is 1-periodic. In Figure 3.4, we show the velocity field of the Double gyre at times 0, 0.25 and 0.75.

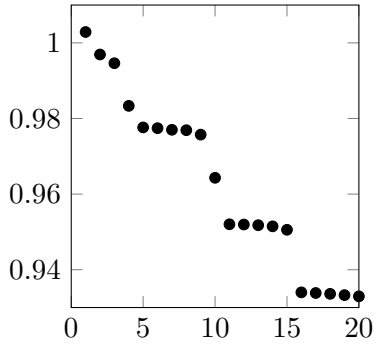

Figure 5.4: Singular spectrum of $\tilde{P}_\epsilon$ for the Double gyre example with a gap occurring after the third singular value.

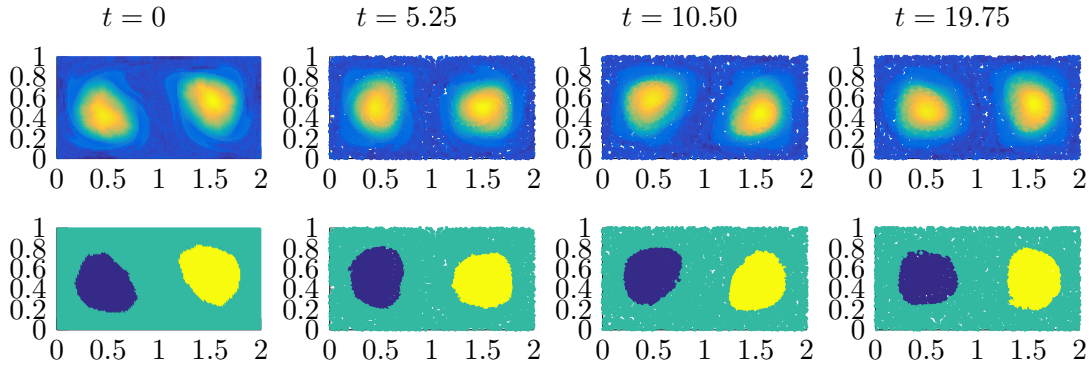

Figure 5.5: First row: Time slices of the third left singular vector and partition into three coherent sets at times $t = 0, 5.25, 10.50, 19.75$. Yellow (blue) indicates 1 (-1). Second row: Coherent 3-partition. See also Figure 3.7 for comparison.

In order to extract the coherent sets, we release $2^7 \times 2^6 = 8192$ trajectories starting on an equispaced grid, and look at 301 samples in time. This is several magnitudes less than the $2^8 \times 2^7 \times 100 = 3276800$ test points we used for Ulam's method for the Double gyre in Section 3.5.1. As basis functions we use Wendland 3,2 functions, and choose the shape parameter such that the support of one basis function intersects the support of 50 other basis functions on average. In Figure 5.4, we show the singular spectrum of the computed transfer operator $\tilde{P}$. We spot a gap after the third singular value, and consider the first three singular vectors to extract three coherent sets. In order to extract coherent sets from these, we use the *kmeans++* algorithm, see Section 2.8. In Figure 5.5, we show time slices of the third left singular vector. We see that the singular vector is indeed a linear combination of the indicator functions of three (connected) sets. Furthermore, we show the extracted coherent 3-partition.

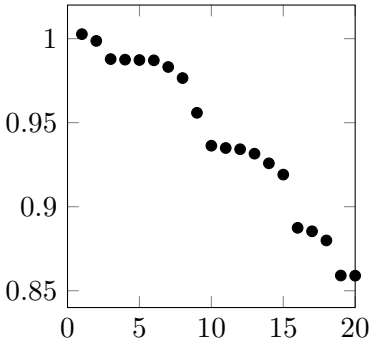

Figure 5.6: Singular Values of $\tilde{P}$ for the Bickley jet, including several spectral gaps.

5.4.2 Bickley jet

Next we consider the Bickley jet, an idealized model of a meandering zonal jet, see Section 3.5.2. A visualization of the vector field at time $t_0 = 0$ is given in Figure 3.8. We release $2^7 \times 2^6 = 8192$ trajectories starting on a grid, and consider 346 time samples. We again use Wendland 3,2 functions as basis functions, and choose the shape parameter such that the support of one basis function intersects the support of 50 other basis functions on average. In Figure 5.6, we show the singular spectrum of the computed transfer operator $\tilde{P}_\epsilon$. As more structures exists in the Bickley jet, it is less obvious to spot the spectral gap than in the Double gyre example. We found three gaps after the second, eighth and ninth singular value inducing partitions into two, eight and nine coherent sets. In Figure 5.7, we show the resulting partitions at initial and final time, and the corresponding last singular vector contributing to the partition. For comparison, also consider the results obtained via Ulam's method in Chapter 3.5.2. The partitions shown in Figure 3.11 are the same obtained here. One main difference is, that for Ulam's method we used $2^7 \cdot 2^5 \cdot 100 = 409600$ trajectories compared to 8192 for the data focused method. Of course, in fact $8192 \cdot 346 = 2834432$ data points are used here, as the family uses the intermediate times.

5.4.3 Incomplete data

In practice most often only an incomplete set of data is available, i.e. the trajectories T_j are known only on a sub interval $\mathcal{S}_j \subset [0, t_1]$. We can deal very naturally with this case, if we simply change our basis slightly to

$$\tilde{\psi}_j(t,x) = \begin{cases} \frac{t_1}{\sqrt{\mu(\mathcal{S}_j)}}\psi_j(t,x), & \text{if } t \in \mathcal{S}_j \\ 0 & \text{else.} \end{cases}, \qquad i = 1, \ldots, N.$$

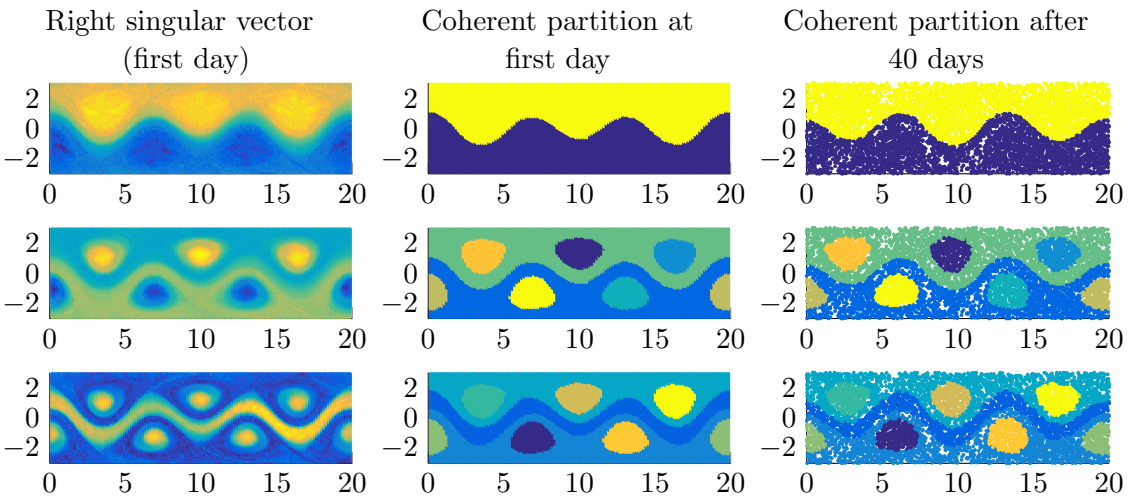

Figure 5.7: First row: Second right singular vector and the coherent 2-partition at initial and final time. Second row: Eighth right singular vector and the coherent 8-partition at initial and final time. Third row: Ninth right singular vector and the coherent 9-partition at initial and final time. Yellow (blue) indicates 1 (-1).

The factor $1/\sqrt{\mu(\mathcal{S}_j)}$ ensures all $\tilde{\psi}_j$ having the same norm. We can do the same calculations as in the full data case, and obtain analogously to (5.11)

$$\begin{aligned}\hat{K}_{\epsilon,ij} &= \frac{1}{t_1}\int_{\mathcal{S}_j} \tilde{\psi}_j(t, \mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi}))dt \\ &= \frac{1}{\sqrt{\mu(\mathcal{S}_j)}}\int_{\mathcal{S}_j} \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi}))dt,\end{aligned}$$

which is a mean value of $\psi_j(t, \mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi}))$ over time. The problem here is, that we may not have information of $\mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi})$ at times $t \in \mathcal{S}_j$ but only at times $t \in \mathcal{S}_i$. However, our best guess for the mean is simply the mean over the time interval, where we have information of T_i available, namely $\mathcal{S}_i \cap \mathcal{S}_j$. So

$$\begin{aligned}\hat{K}_{\epsilon,ij} &\approx \frac{1}{\sqrt{\mu(\mathcal{S}_i \cap \mathcal{S}_j)}}\int_{\mathcal{S}_i\cap\mathcal{S}_j} \psi_j(t, \mathcal{T}_\epsilon^{0,t}(x_i,\tilde{\xi}))dt \\ &\approx \frac{\Delta t}{\sqrt{\#(s_k \in \mathcal{S}_i \cap \mathcal{S}_j)\Delta t}}\sum_{s_k\in\mathcal{S}_i\cap\mathcal{S}_j} \psi_j(s_k, T_i^k) \\ &= \frac{\sqrt{\Delta t}}{\sqrt{\#(s_k \in \mathcal{S}_i \cap \mathcal{S}_j)}}\sum_{s_k\in\mathcal{S}_i\cap\mathcal{S}_j} \psi_j(s_k, T_i^k).\end{aligned}$$

If two trajectories live on disjoint time intervals $\mathcal{S}_i \cap \mathcal{S}_j = \{\}$, the corresponding entry $K_{ij} = 0$.

5.4.4 Martha's Vineyard

In this last example we look at 72 buoys released in August 2011 into Martha's Vineyard bay by Irina Rypina and her team, [Rypina et al., 2014]. The buoys sent their positions 100 times or roughly every 45 minutes from August 1st, 14:30 until August 4th, 20:00. The buoys are not all released at the same time and only three of them are still alive at the final time August 4th, 20:00. Hence the data set is very sparse and incomplete and therefore unsuited for traditional transfer operator methods. Furthermore the underlying flow is not really conservative, as up and downwelling can occur. However, our transfer operator $\tilde{P}_\epsilon$ is designed to deal with this setting. As basis for V_{01} we use the Gaussians

$$\psi_j(t, x) = \exp(-\alpha^2 \|\mathcal{T}(t, x_j) - x\|_V^2)$$

with global support as basis. Here, $\|.\|_V$ is the 2-norm on the sphere. The buoys are all released south of Martha's vineyard and live on average for 64 time steps, see Figure 5.8. In Figure 5.9, we show a basis function ψ_i at fixed time as well as the spectrum of the computed transfer operator $\tilde{P}_\epsilon$. We see that in this data set there are no truely apparent coherent structures as the singular values decay rapidly. However we use the gap after the third singular value and identify three sets. One of those sets is a very separated group of drifters diverting to the north-east through the narrow between Martha's Vineyard and Tuckernuch Island. The other groups are first heading west and then north and south, respectively. Of course this is a very rough approximation of the transfer operator and only in some sub-domains of Ω. However the data foundation is very sparse in this example and at least some coarse separated regions can be identified. Note also that in the beginning, as all buoys are released together they are all very close and hence the first time measurements do not contribute much to the separation.

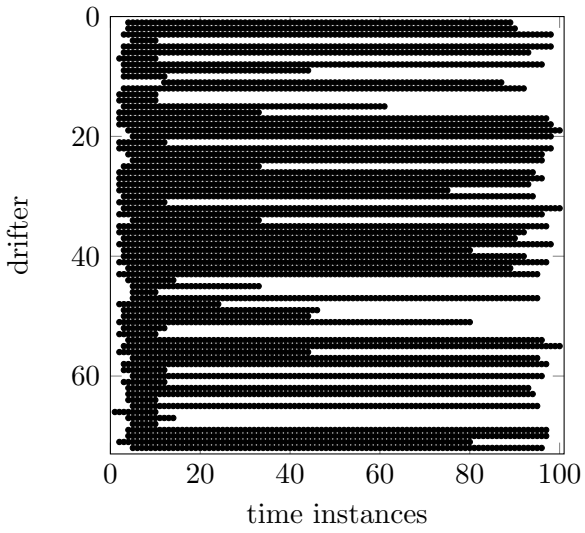

Figure 5.8: Lifetime of the drifters.

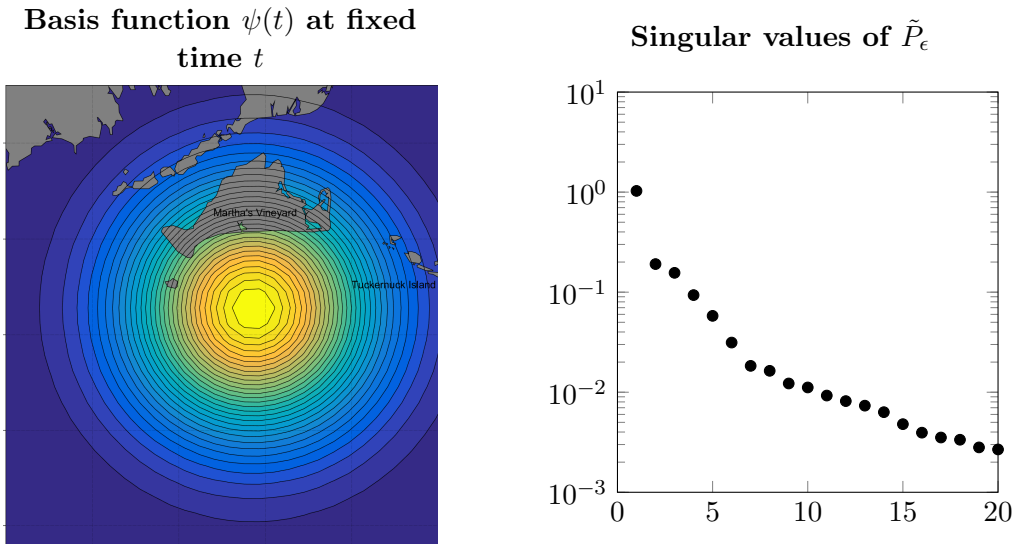

Figure 5.9: Martha's vineyard: A basis function ψ_i at fixed time and the singular values of $\tilde{P}_\epsilon$.

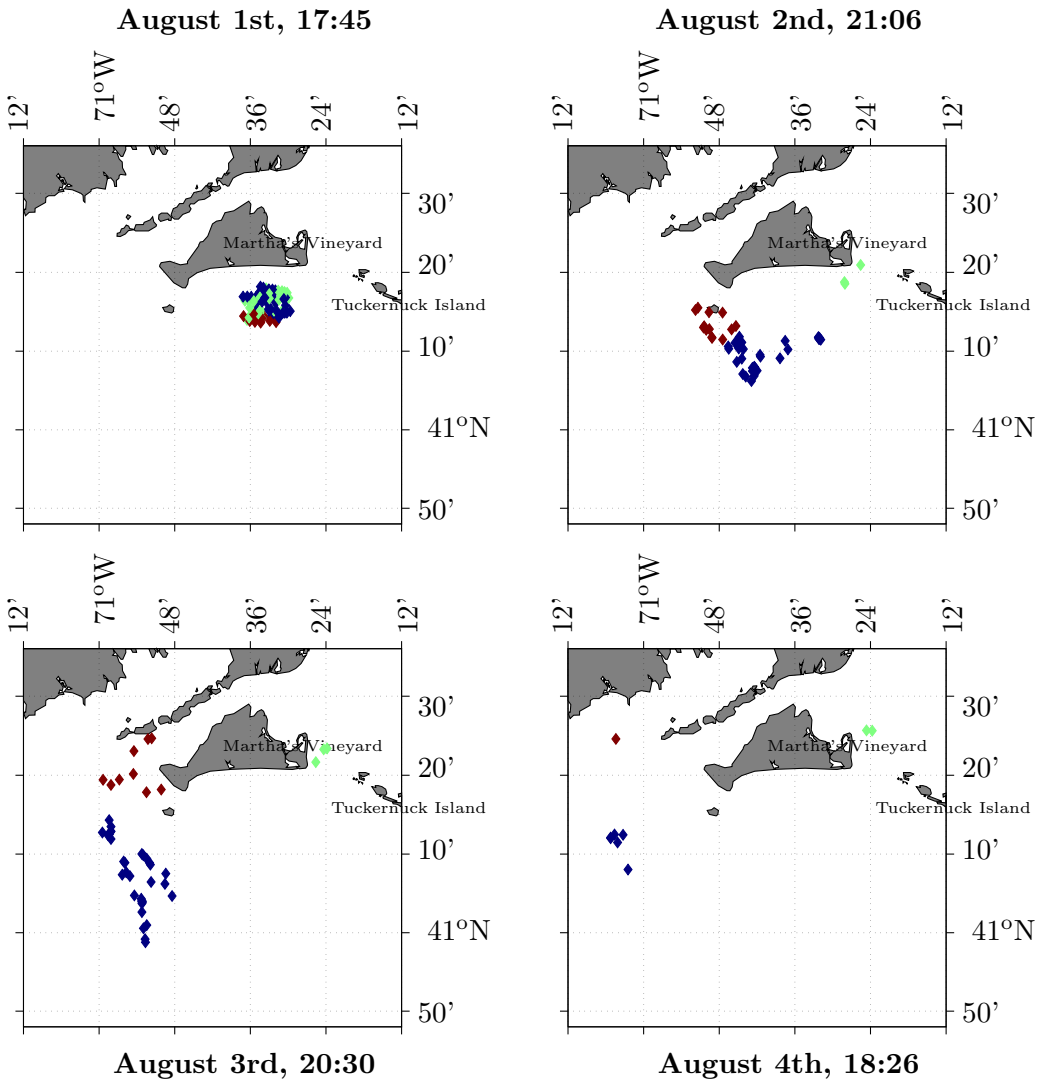

Figure 5.10: 3-partition of the drifters at Martha’s vineyard.

Chapter 6

Coherent sets in plasma physics

In this chapter, we want to apply the concepts and algorithms developed in the previous chapters to problems in plasma physics. We discuss several famous instabilities occurring in the kinetic description of a plasma, such as *two-stream instability*, *Bump-on-tail instability* and *streaming Weibel instability*. In difference to classical fluids, particles in a plasma are described via their position *and* their velocity. Hence the state space consists of up to six dimensions, three position, and three velocity dimensions. This is one reason, why the numerics of a plasma are exceedingly challenging. Especially in high dimensional state spaces, particle based algorithms such as particle in cell (PIC) and particle in Fourier (PIF) codes are widespread in plasma numerics. This is beneficial for the trajectory based discretization of the transfer operator family (see Chapter 5.3). Like that we can directly access the data of the theoretical particles, and use those to compute a coarse approximation of the transfer operator family even in high dimensions. Hence, we will use this approach for most of the examples. All computations of the particles are carried out by Jakob Ameres[1] [Ameres and Sonnendrücker, 2015, Ameres and Sonnendrücker, 2016].

6.1 Two-stream instability

In our first example, we consider the Vlasov-Poisson equation (2.25),

$$f_t + f_x \mathbf{v} + \frac{q}{m}\mathbf{E} f_v = 0$$
$$\Delta\Phi = \frac{\rho}{\epsilon_0}.$$

[1] Max-Planck-Institut für Plasmaphysik (IPP), Boltzmannstraße 2, D-85748 Garching and TUM, Center for Mathematics, Boltzmannstraße 3, D-85748 Garching.

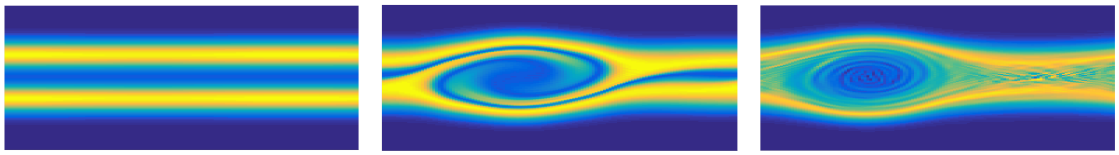

Figure 6.1: Particle density of the two-stream instability example at times $t = 0,\ 30,\ 130$.

with initial condition

$$f_0(x,v) = \frac{1}{2\sqrt{2\pi}}(1+\epsilon\cos(kx))(\exp(-\frac{1}{2}(v-v_0)^2))(\exp(-\frac{1}{2}(v+v_0)^2)),$$

where $v_0 = 2.4,\ k = 0.2,\ \epsilon = 0.01$ on a domain $\Omega = [0, 2\pi/0.16] \times [-8, 8]$ describing the position-velocity space (x-v space). We assume Ω to be periodic in x. This describes two particle beams shot at each other in a one dimensional tube. Depending on k and v_0, this configuration can be stable or unstable in the sense that the two beams are perturbed and a vortex forms out. With our choice, the system is indeed unstable. We normalize the equation, such that $q/m = 1$.

There are many methods to solve this standard example, e.g. particle methods, [Hockney and Eastwood, 1988], Semi-Lagrangian methods [Sonnendrücker et al., 1999], or splitting schemes [Klimas and Farrell, 1994]. An overview of computational plasma physics in general is given in [Birdsall and Langdon, 2004]. An overview over the various numerical methods for solving the Vlasov-Poisson equation is given in [Filbet and Sonnendrücker, 2003]. Here, we compute the electric field induced by the particles via a standard semi-Lagrangian method. In Figure 6.1, we show the evolution of the particle density. As time evolves, some particles get trapped in state space, and a vortex in position-velocity space forms out. This phenomena is called *two-stream instability*. We extract the electric field $\mathbf{E}$ from this computation.
In this example, we aim at computing the coherent sets in the plasma via Ulam's method. To this end we have to compute the trajectories of single particles, in this case electrons. As we have now obtained the electric field $\mathbf{E}$, we can compute these trajectories in the following way:
As we assume the magnetic field $\mathbf{B} = 0$, the force applied to an electron in the plasma is $\mathbf{F}(t,x) = q\mathbf{E}(t,x)$. With $\mathbf{F} = m\mathbf{v}$ follows

$$\dot{\mathbf{v}}(t,x) = \frac{q}{m}\mathbf{E}(t,x).$$

In reality, for the electron holds

$$q = -1,602 \cdot 10^{-19} C$$
$$m = 9,109 \cdot 10^{-31} kg.$$

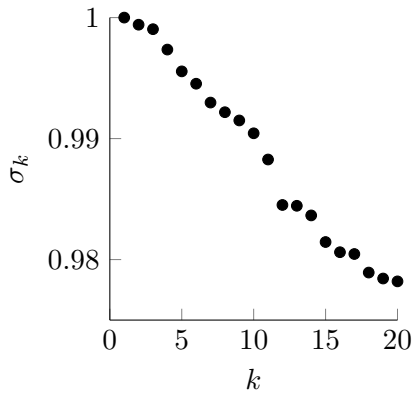

Figure 6.2: Singular values of the transfer operator for the two-stream instability

and, therefore, as $\mathbf{E} = \frac{F}{q}$ and $N = \frac{kg \cdot m}{s^2}$

$$[\dot{\mathbf{v}}] = \frac{m}{s^2}$$

and

$$\dot{\mathbf{v}}(t,x) = \frac{q_e}{m_e}\mathbf{E}(t,x) = -0.1759 \cdot 10^{12} \cdot \mathbf{E}(t,x).$$

As we consider the normalized Vlasov-Poisson equation, we get

$$\dot{\mathbf{v}}(t,x) = -\mathbf{E}(t,x).$$

which we rewrite as a two dimensional system

$$\begin{pmatrix} \dot{x} \\ \dot{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ -\mathbf{E}(t,x) \end{pmatrix}$$

For the computation of the coherent partition, we start at time $t_0 = 100$, when the vortex already has formed out, and compute up to time $t_1 = 130$. We release 100 particles per box on a grid containing $2^6 \times 2^6$ boxes. We use the Lebesgue measure as reference measure. In Figure 6.2, we show the singular spectrum of the transfer operator. We spot possible gaps after the third and fourth singular values. In Figure 6.3, we show the second to fourth singular vectors, and in Figure 6.4, we show two partitions into three and four coherent sets, respectively.

As we compute a coherent 3-partition, we can distinguish the northern and southern hemisphere from the center stream. In the coherent 4-partition also the vortex is identified.

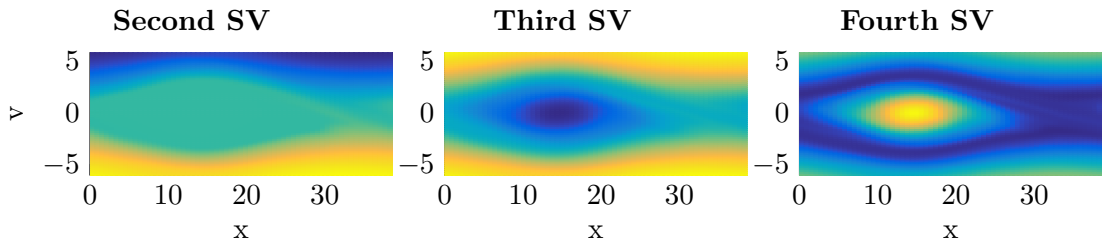

Figure 6.3: Second, third and fourth singular vectors at initial time for the two-stream instability.

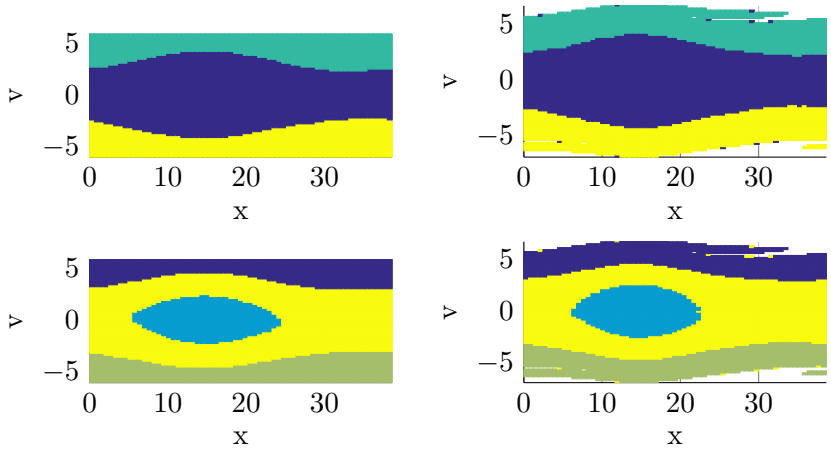

Figure 6.4: Coherent partitions at initial (first row) and final time (second row).

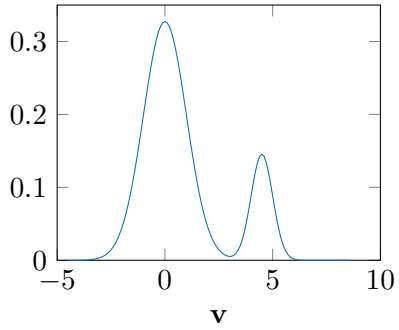

Figure 6.5: Spatially averaged initial distribution function plotted against velocity for the one dimensional Bump-on-tail instability.

6.2 Bump-on-tail instability

Next we consider the Vlasov-Poisson equation with initial condition

$$f(t=0,x,\mathbf{v}) = \left(1+\epsilon\left(\sum_{i=1}^{d} k_i x_i\right)\right)\frac{1}{\sqrt{2\pi}}\frac{1}{(1+a)}e^{-\frac{\mathbf{v}^2}{2}} + \frac{a}{\sigma}e^{\frac{-\|\mathbf{v}-\mathbf{v}_0\|^2}{2\sigma^2}},$$

describing a diffuse electron beam drifting with a velocity v_0 relative to a stationary hot plasma. It models the situation of a fast particle injection, similar to the neutral beam injection (NBI) in a tokamak. Here, σ^2 is the thermal spread, and the ions are assumed to form an immobile background. We set the parameters to $\epsilon = 0.001$, $\sigma = 0.5$, $a = 2/9$, $\mathbf{v}_0 = 4.5$, $q/m = -1$. The domain considered is $\Omega = \Omega_x \times \Omega_v$, $\Omega_x = [0, 2\pi/k]^d$, $\Omega_v = [-5, 8.5]$. We assume the position space Ω_x to be periodic.

6.2.1 1D × 1D Bump-on-tail

We first consider the system in one position dimension and one velocity dimension. Therefore, $d = 1$. We set the wave number $k = k_1 = 0.3/3$. In Figure 6.5, we show the spatially averaged initial distribution function f showing the Bump-on-tail perturbation.

This initial distribution turns out to be unstable due to the perturbation. This leads to a growth of this initial perturbation followed by saturation and oscillation of the particles trapped in the potential through the wave [Shoucri, 1979].

In this example we have access to 10000 computed particles computed via a particle-in-Fourier (PIF) solver [Ameres et al., 2016]. We apply the data-driven discretization developed in section 5.3. As basis functions we use Wendland 3-2 functions, and choose the support of a basis function such that it intersects with the support of approximately 100 other basis functions. We choose the initial time as $t_0 = 75$ and the final time as $t_1 = 110$. In Figure 6.6, we show the singular spectrum of the such computed transfer

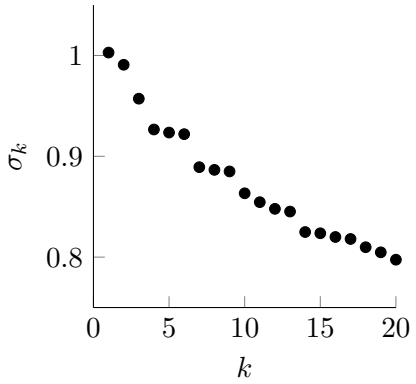

Figure 6.6: Singular values of the transfer operator for the 1×1 dimensional Bump-on-tail instability.

operator and spot several gaps after the second, third and the sixth singular value. In Figure 6.7, we show the coherent partitions into two and six sets at initial and final time. The partition into two sets separates the southern part, a partition into three sets would also separate the northern part from the center. The partition into six sets additionally identifies three vortices that move to the east over time.

6.2.2 2D $\times$ 2D Bump-on-tail

Next we consider the Bump-on-tail instability in two spatial and two velocity dimensions. We set $k = 0.3$. Hence, the periodic position space $\Omega_x = [0, 2\pi/0.3]^2$. The velocity space is $\Omega_v = [-9, 9]^2$. In principle, the velocity space is unbounded. However, the are very few particles in the gaussian tail of the initial distribution and therefore they can be neglected. We compute 5000 particles via a particle-in-Fourier (PIF) solver [Ameres et al., 2016]. We consider the time interval $[t_0, t_f] = [50, 80]$, and use time resolution of $dt = 0.2$, resulting in 150 time slices for the computation of the transfer operator family. We use Wendland 4-2 functions $\psi_{4,2}$

$$\begin{aligned} \psi_{4,2} : [0, \infty) &\longrightarrow [0, \infty) \\ \psi_{4,2}(r) &= \max(1 - r, 0)^6 (32r^2 + 18r + 3), \end{aligned}$$

which are strictly positive definite and radial on $\mathbb{R}^4$, for the construction of basis functions [Wendland, 2004]. The basis functions then read $\psi_j(t, x) = \phi(\alpha \| \mathcal{T}(x_j(t)) - x \|_2)$, where $x_j(t)$ denotes the j-th particle at time t, see Chapter 5. The support of ψ_j is $1/\alpha$. We choose α, such that the support of one basis function intersects with the support of 50 other basis functions on average. In Figure 6.8, we show the spectrum of the transfer operator family. The singular values decay quite fast. We spot a spectral gap after the third singular value and decide to extract a coherent 3-partition. In Figure 6.9 we show several projections of this partition at time $t = 50$. We observe in the

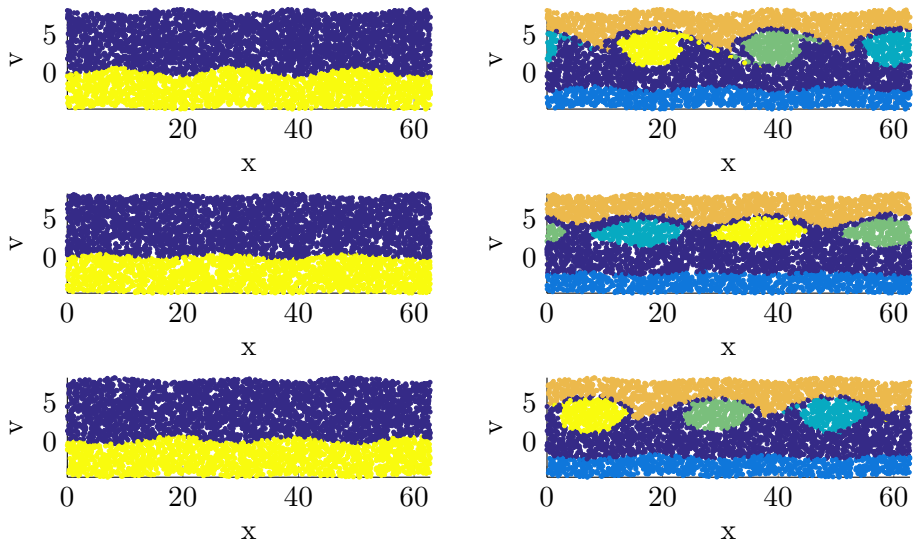

Figure 6.7: Parition into two and six coherent sets at initial time $t_0 = 75$ (first row), time $t = 100$ (second row) and final time $t_1 = 110$ (third row).

projection of the coherent partition to velocity space, that the velocity of the particles is very homogeneous. On the other hand, the projection to position space seems totally chaotic. With the other two projections, we see the reason for that: In position space the coherent sets are "stacked" onto each other. In this example, it is, hence, not sufficient to consider the position of the particles only, in order to detect the coherent structures. The full dimensions are needed.
Nevertheless, the question remains what can be found in position (velocity) space. This encourages us to look at the position and velocity separately. We redo the computation for reduced state spaces Ω_x and Ω_v, instead of $\Omega = \Omega_x \times \Omega_v$. The results are shown in Figure 6.10. The spectrum for the position space reveals, that there are indeed no really coherent sets there, with the second singular value $\sigma_2 \approx 0.3$. However if we compute the 2-partition we actually rediscover the evolution of the initial condition as a traveling wave in position space. The spectrum for the velocity space reveals the sets there indeed being very coherent.

6.3 Diocotron instability

Next, we consider the Diocotron instability in a Guiding Center Model. The evolution of a density $f(x, y)$ coupled to the Poisson equation is given as

$$\begin{aligned} \partial_t f + \nabla f \times \mathbf{E} = 0, \quad t \in [0, T] \Leftrightarrow \\ \partial_t f + (\nabla \Phi)_2 \partial_x f - (\nabla \Phi)_1 \partial_y f = 0, \quad t \in [0, T], \end{aligned}$$

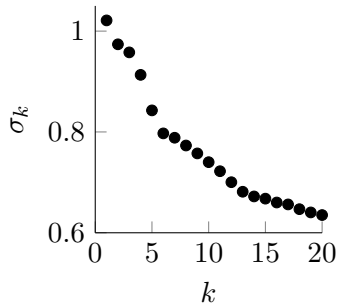

Figure 6.8: Spectrum of the transfer operator family for the 2×2 dimensional Bump-on-tail instability.

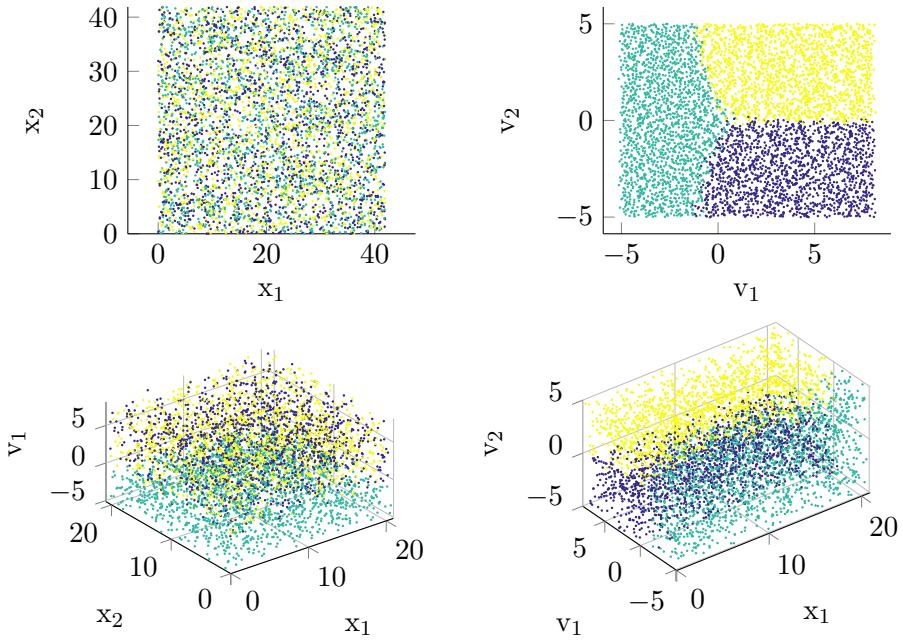

Figure 6.9: Projections of the three-partition onto several subspaces of Ω at time $t = 50$.

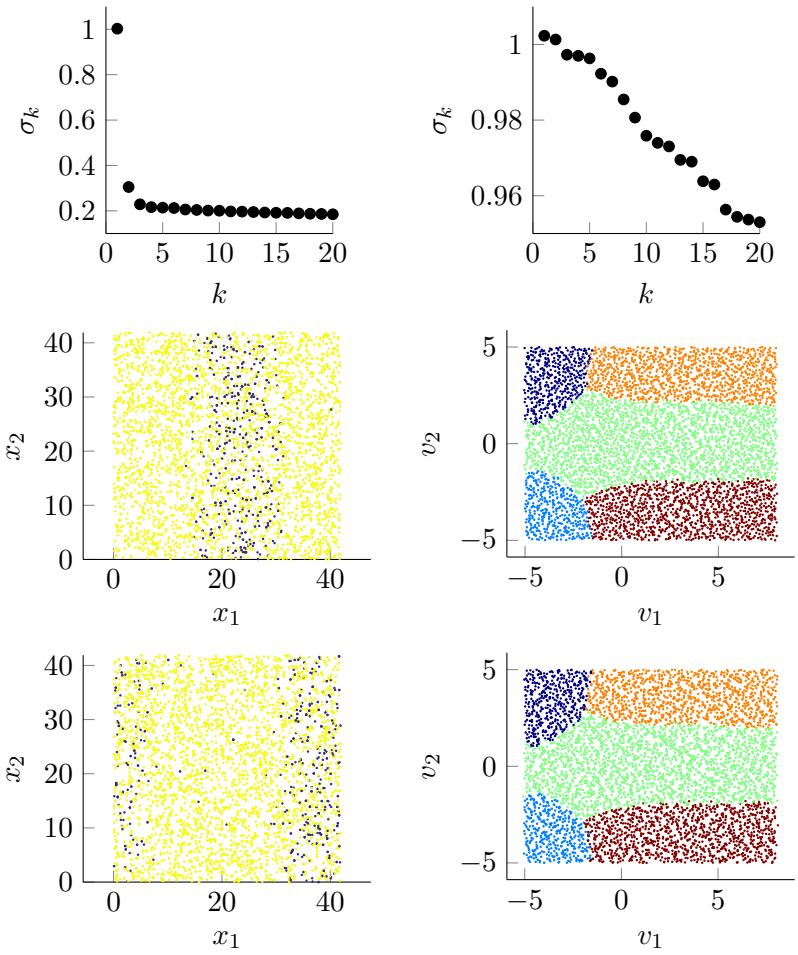

Figure 6.10: First column: Spectrum of the transfer operator family for the four dimensional Bump-on-tail instability reduced to position space, and partition into two coherent sets at initial and final time. We identify the particle wave traveling through position space. Second row: Spectrum of the transfer operator family for the four dimensional Bump-on-tail instability reduced to velocity-space and partition into five coherent sets at initial and final time. The partitions are very coherent and do not move much in time at all.

where $(\nabla\Phi)_i$ denotes the i-th component of $\nabla\Phi$, where the electric field is induced by the potential Φ via

$$\mathbf{E} = -\nabla\Phi,$$

and the Poisson equation is

$$-\Delta\Phi(x, y) = f(x, y).$$

The characteristics for every particle are given as

$$\frac{d}{dt}x(t) = -E_y(t, x(t), y(t)), \quad \frac{d}{dt}y(t) = -E_x(t, x(t), y(t)).$$

For more details about the equations and their observation in nature, see [Levy, 1965]. We used Finite Elements coupled with a particle-in-cell code provided by [Ameres, 2016]. In a cylinder, the most unstable θ-mode is five, therefore we excite this one as initial condition, see Figure 6.11. This method again has the advantage, that it directly computes the trajectories of particles we can use for the approximation of our transfer operator family. In this case, we used 10000 particles equally distributed on a ring around the origin. Their evolution is quite complex, and the five stable vortices form quite leaky clusters, see Figure 6.11. We consider the time interval $[t_0, t_f] = [0, 500]$, as we want to detect where the vortices origin from. We compute the approximation to the transfer operator family with our data driven approach, and set $\alpha = 0.025$ such that the resulting matrix has approximately 2.8% non-zero entries. We obtain the spectrum shown in Figure 6.11. We spot two dominant gaps after five and ten singular values. In Figure 6.12, we show the extracted coherent five and ten-partitions. The five-partition cuts the initial ring into five parts which evolve into five stable but leaky modes including their tails. The ten-partition additionally separates those tails from the vortices.

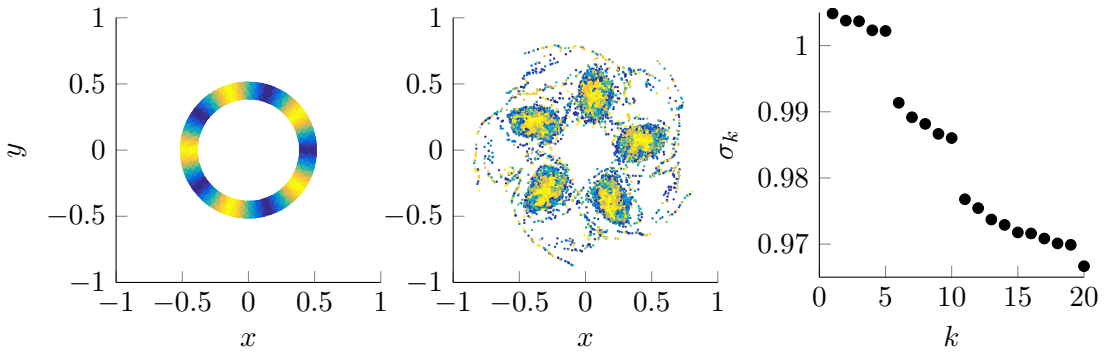

Figure 6.11: Particle density in the Diocotron instability at initial and final time. Right: Singular values of the associated transfer operator family.

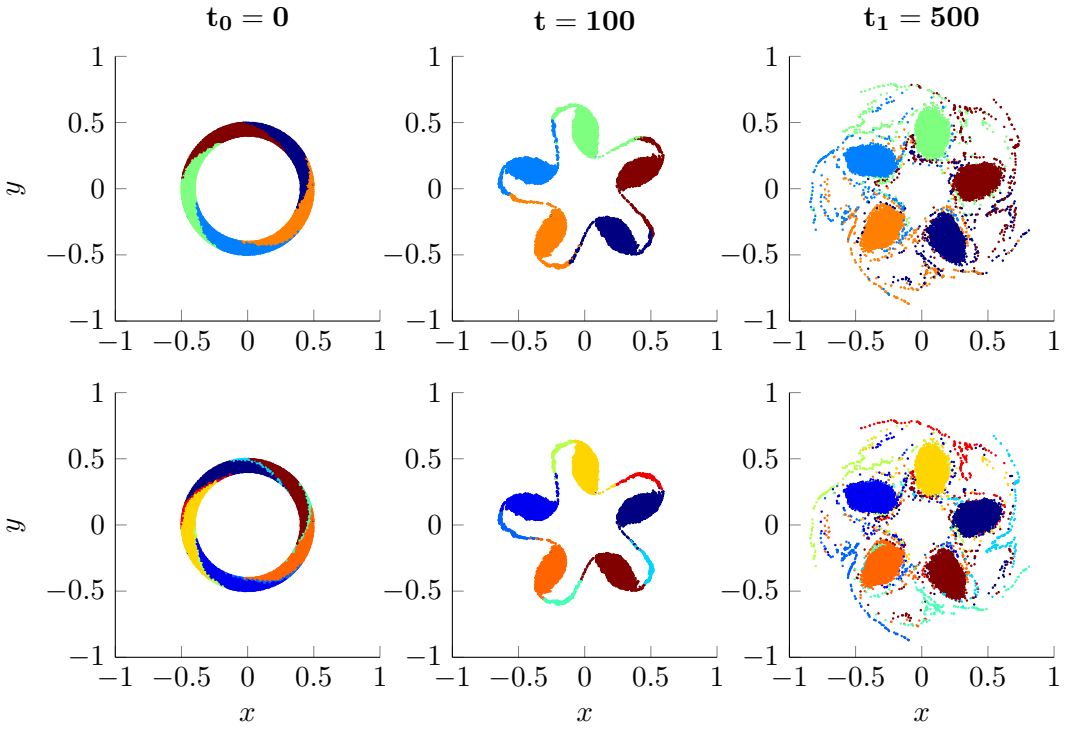

Figure 6.12: First row: Coherent 5-partition identifying the five stable modes in the Diocotron instability example at initial time $t_0 = 0$, time $t = 100$ and final time $t_1 = 500$. Second row: Coherent 10-partition additionally separating the tails.

6.4 Streaming Weibel instability (Vlasov-Maxwell)

In this section, we consider the Vlasov-Maxwell equation (2.24) in three dimensions. We take the equations, the initial condition, and the Finite Element Particle-in-Cell code for their solution from [Kraus et al., 2016]. The Vlasov-Maxwell equations in one dimensional position space and two dimensional velocity space read

$$\begin{aligned} f_t + f_x \mathbf{v} + \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_v &= 0 \\ \partial_t B &= -\partial_x E_2, \\ \partial_t E_1 &= -j_1, \\ \partial_t E_2 &= -\partial_x B - j_2, \\ \partial_x E_1 &= \rho \end{aligned}$$

The currents $j_1(x,t), j_2(x,t)$ and the charge density $\rho(x,t)$ are defined as

$$\begin{aligned} j_1(x,t) &:= \int_{\mathbb{R}^2} v_1 f(x, v_1, v_2, t) dv_1 dv_2, \\ j_2(x,t) &:= \int_{\mathbb{R}^2} v_2 f(x, v_1, v_2, t) dv_1 dv_2, \\ \rho(x,t) &:= \int_{\mathbb{R}^2} f(x, v_1, v_2, t) dv_1 dv_2. \end{aligned}$$

The initial condition is set to

$$\begin{aligned} f(t=0,x,v) &= \frac{1}{\pi\sigma} e^{-\frac{v_1^2}{2\sigma^2}} \left(\delta e^{-\frac{(v_2 - v_{0,1})^2}{2\sigma^2}} + (1-\delta) e^{-\frac{(v_2 - v_{0,2})^2}{2\sigma^2}} \right) \\ B_3(x, t=0) &= \beta \sin(kx) \\ E_2(x, t=0) &= 0, \end{aligned}$$

and $E_1(x, t=0)$ is computed from the Poisson equation. We set the parameters to the following values $\sigma = 0.1/\sqrt{2}$, $k = 0.2$, $\beta = 10^3$, $v_{0,1} = 0.5$, $v_{0,2} = -0.1$, $\delta = 1/6$. The domain size $\Omega_x = [0, 10\pi]$. The parameters were chosen like this originally in [Cheng et al., 2014]. We release 10000 particles for the computation of the field and for the approximation of our transfer operator family. The time step we use for the computation of the evolution of the particles is 0.01. As time interval we choose $[t_0, t_1] = [0, 300]$. The domain for the velocity is in principle $\mathbb{R}^2$, but only finite values are taken by the particles, of course. We show the evolution of the particle density in Figure 6.13. We observe that some particles are separated from the center main stream. For the computation of the approximation to the transfer operator family, we use Wendland 3-2 functions as basis functions. One challenge in this example is, that position and velocity live on different scales. We observe that the maximum absolute velocity is approximately 0.2, which is around 20 times less than the interval length of the position space. Hence,

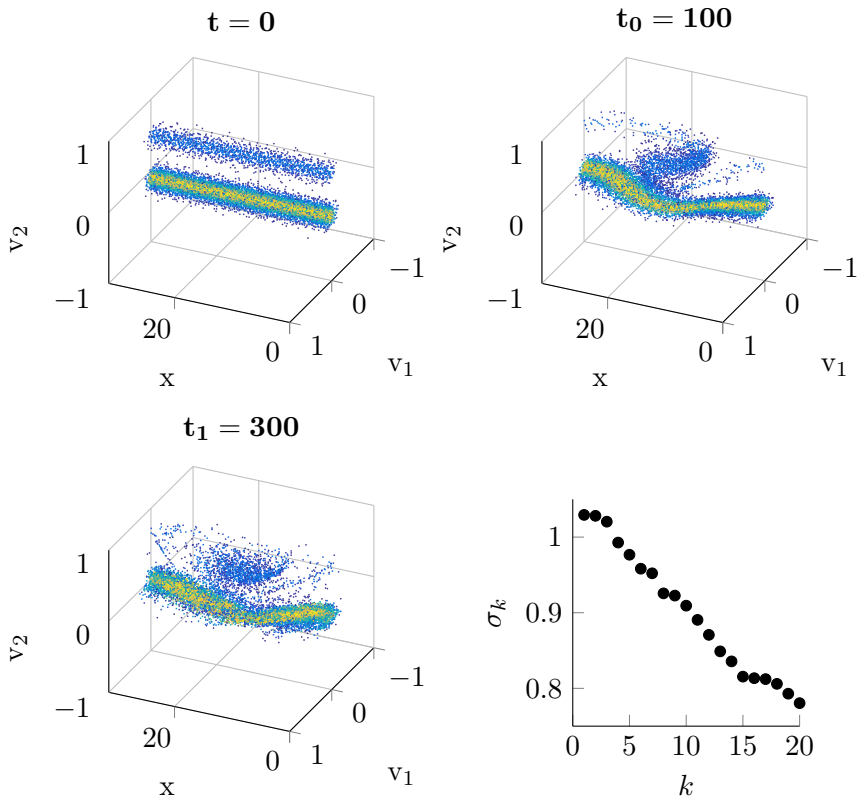

Figure 6.13: Particle density in the streaming Weibel instability at time $t = 0$, initial time $t_0 = 100$ and final time $t_1 = 300$ and singular values of the associated transfer operator family.

if we do not choose the support of a basis function carefully each basis function lives on the whole velocity domain or only on a very small part of the position domain. To overcome this problem, we set the radius of the basis functions to 0.2 in position space and 0.02 in velocity space (the support is an ellipse). For the computation of the transfer operator family, we use the coarser step size 1 in time, resulting in 301 time slices for its computation. In Figure 6.13, we show the spectrum of the so computed transfer operator family. We spot a gap after the third singular value and decide to extract three coherent sets. In Figure 6.14, we show the coherent three partition of the particles. We can identify clearly a vortex forming out above the center main stream. Additionally, the vortex is equipped with a ring of particles orbiting it.

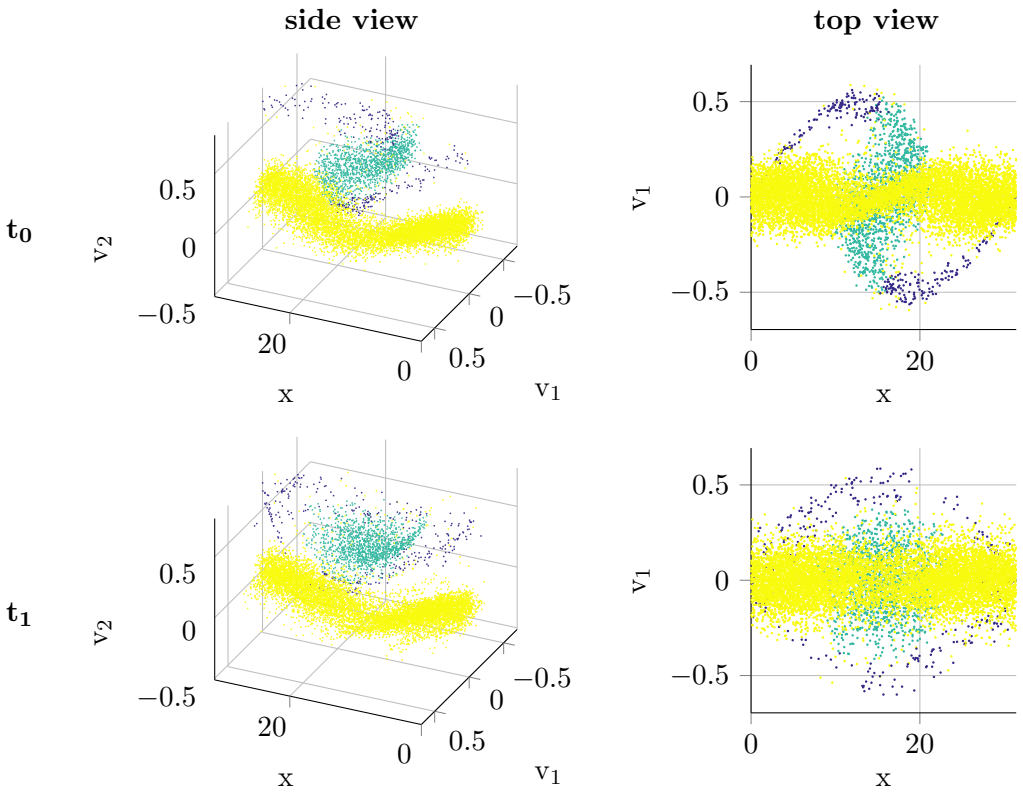

Figure 6.14: First row: Coherent three-partition at initial time $t_0 = 100$, side and top view. Second row: Coherent three-partition at initial time $t_0 = 300$, side and top view.

Chapter 7

Conclusion and outlook

Starting point of this thesis was [Froyland, 2013] using transfer operators for the computation of coherent sets, and identifying these sets as coherent pairs via

$$\frac{\mu(A_0 \cap \mathcal{T}^{t_1,t_0}(A_1))}{\mu(A_0)} \approx 1,$$

subject to some small (numerical) diffusion. We developed this mathematical notion of coherence in several steps and identified connections to related mathematical fields. First, we recaptured well-known techniques from *spectral clustering* theory to extend Ulam's method for the identification of an arbitrary meaningful number of n coherent sets in chapter 3 via considering

$$\frac{1}{n}\sum_{k=1}^{n} \frac{\mu(A_0^k \cap \mathcal{T}^{t_1,t_0} A_1^k)}{\mu(A_0^k)} \approx 1$$

subject to some small diffusion.
In chapter 4, we used *stochastic flow maps* to establish a more rigorous definition directly including the necessary diffusion:

$$\mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t_1,t_0}(\mathcal{T}_\epsilon^{t_0,t_1}(A,\cdot),\cdot)\Big] \;\Big|\; A\Big) \approx 1.$$

Like that we could establish a connection to the originating theory of *almost invariant sets*. We used Brownian motion to connect this notion of coherence to the Fokker-Planck equation allowing a spectral discretization.
We extended this definition in chapter 5 to consider the state of the system at all time instances of the time interval. We called a set A coherent, if

$$\frac{1}{t_1}\int_{t_0}^{t_1} \mu\Big(\mathbb{E}_{\mathbb{P}\times\mathbb{P}}\Big[\mathcal{T}_\epsilon^{t,t_0}(\mathcal{T}_\epsilon^{t_0,t}(A,\cdot),\cdot)\Big] \;\Big|\; A\Big)\, dt \approx 1,$$

and established a connection to families of transfer operators, which are also known in *optimization*. We derived a data based discretization of this family using *meshfree methods*, and connected it to recently established *data-driven algorithms*.
We applied the methods derived to various instabilities in *plasma physics* in chapter 6. We thoroughly identified the previously hard to catch instabilities in plasma physics models, also in high dimensions.

An important question not touched by this thesis is the zero diffusion limit, i.e. what happens if the diffusion added to the system approaches zero. This question was investigated recently in [Karrasch and Keller, 2016] (see also [Froyland, 2015]). There, a Laplace-Beltrami operator is obtained and used for the computation of (geometrically) coherent sets.

As the the concept of coherence is now sufficiently well understood, there open up several opportunities for future research. In many applications, it is interesting not only to identify coherent structures in a given time interval, but to influence the coherence ratio and the areas coherent structures cover.

In this sense a natural next step is to try to control and optimize the coherent structures. For ocean fluids that could be to optimize the flow rate for a fluid via suppressing coherence or trapping oil films inside coherent structures.

Concerning plasma physics a next step should be a further investigation of plasma physics models including thorough parameter studies. In the long run, the control of coherence in plasma physics is of great interest. Especially in fusion power plant construction, the confinement of particles inside the reactor and the mixing therein is important for a profitable operation.

List of Figures

Appendix

Motivation for the exponential time-differencing scheme

We motivate an exponential time-differencing scheme for stiff systems as introduced in [Cox and Matthews, 2002]. To begin with, we consider the one-dimensional time dependent ODE

$$u_t = cu + F(u,t), \tag{.1}$$

where $c \in \mathbb{R}$ is constant and $F(u,t)$ represents nonlinear or forcing terms. We derive an alternative formulation of .1 via first multiplying with the integrating factor e^{-ct} and then integrating from t_0 to $t_1 = t_0 + h$:

$$\begin{aligned}
e^{-ct}u_t =& e^{-ct}(cu + F(u,t)) \\
\int_{t_0}^{t_1} e^{-c\tau}u_t d\tau =& \int_{t_0}^{t_1} e^{-c\tau}(cu + F(u,\tau))d\tau \\
ue^{-c\tau}|_{t_0}^{t_1} + \int_{t_0}^{t_1} uce^{-c\tau}d\tau =& \int_{t_0}^{t_1} e^{-c\tau}(cu + F(u,t))d\tau \\
e^{-ct_1}u(t_1) - e^{-ct_0}u(t_0) + \int_{t_0}^{t_1} uce^{-c\tau}d\tau =& \int_{t_0}^{t_1} cue^{-c\tau}d\tau + \int_{t_0}^{t_1} e^{-c\tau}F(u,t)d\tau \\
e^{-ct_1}u(t_1) =& e^{-ct_0}u(t_0) + \int_{t_0}^{t_1} e^{-c\tau}F(u,t)d\tau \\
u(t_1) =& e^{ct_1}e^{-ct_0}u(t_0) + \\
& e^{ct_1}\int_0^h e^{-c(\tau+t_0)}F(u(t_0+\tau), t_0+\tau)d\tau \\
u(t_1) =& e^{ch}u(t_0) + e^{ch}\int_0^h e^{-c\tau}F(u(t_0+\tau), t_0+\tau)d\tau.
\end{aligned}$$

This formula is *exact* and the integrating scheme depends on the approximation of the integral. The simplest approximation is assuming $F(u,t)$ to be constant on $[t_0, t_1]$. We hence can compute

$$\begin{aligned}
e^{ch}\int_0^h e^{-c\tau}F(u(t_0+\tau), t_0+\tau)d\tau &= e^{ch}F\int_0^h e^{-c\tau}d\tau \\
&= e^{ch}F(-\frac{1}{c}e^{-c\tau})|_0^h \\
&= \frac{F}{c}(e^{ch} - 1),
\end{aligned}$$

and the *etd1* scheme then reads

$$u(t_1) \approx e^{ch}u(t_0) + \frac{F}{c}(e^{ch} - 1).$$

In the same manner higher order schemes can be derived, e.g. the fourth order *etdrk4* scheme, see [Cox and Matthews, 2002]. Furthermore this result can be directly generalized for systems of ODEs, whose linear part is a diagonal matrix $\mathcal{L}$

$$u_t = \mathcal{L}u + \mathcal{F}(u,t),$$

$\mathcal{L} \in \mathbb{R}^{n\times n}$ is diagonal, e.g. a discretized Laplacian. The *etd* formulation then becomes

$$u(t_1) = e^{\mathcal{L}h}u(t_0) + e^{\mathcal{L}h}\int_0^h e^{-\mathcal{L}\tau}\mathcal{F}(u(t_0+\tau), t_0+\tau)d\tau.$$

In Chapter 4 we use an version of an *etdrk4* scheme from [Kassam and Trefethen, 2005], which elegantly treats a cancellation problem occurring in a naive formulation of *etdrk4* by means of a contour integral approximated by the trapezoidal rule. The corresponding Matlab function is shown in the next section.

MATLAB code for the quadruple gyre example

```
function u = etdrk4(t0, m, h, L, v, u)

% compare kursiv.m in AK Kassam and LN Trefethen
% "FOURTH-ORDER TIME-STEPPING FOR STIFF PDEs"
% SIAM Journal on Scientific Computing, 2005

L = L(:); eL = exp(h*L); eL2 = exp(h*L/2);  t = t0;
M = 16; r = exp(1i*pi*((1:M)-0.5)/M);
R  = h*L*ones(1,M) + r(ones(size(u,1),1),:);
Q  = h*real(mean((exp(R/2)-1)./R,2));
f1 = h*real(mean((-4-R+exp(R).*(4-3*R+R.^2))./R.^3,2));
f2 = h*real(mean((   4+2*R+exp(R).*(-4+2*R))./R.^3,2));
f3 = h*real(mean((-4-3*R-R.^2+exp(R).*(4-R))./R.^3,2));

for j = 1:m
    Nv = v(t,u);         a = eL2.*u + Q.*Nv;
    Na = v(t+h/2,a);     b = eL2.*u + Q.*Na;
    Nb = v(t+h/2,b);     c = eL2.*a + Q.*(2*Nb-Nv);
    Nc = v(t+h,c);
    u = eL.*u + Nv.*f1 + (Na+Nb).*f2 + Nc.*f3;
    t = t+h;
end
```

```
% computation of coherent sets for a quadruple gyre system
%
%  Andreas Denner and Oliver Junge, TUM, 2015

M = 15; x = 2/M*(0:M-1)'; [X,Y] = meshgrid(x);     % collocation points
D = 1i*ifftshift(-(M-1)/2:(M-1)/2)';    % derivative in frequency space
Dy = D*ones(1,M); Dx = -ones(M,1)*D';
ep = 0.02; L = ep^2/2*(Dy.^2+Dx.^2);                % Laplace operator

%% vector field
dl = 0.25; om = 2*pi;
f = @(t,x) dl*sin(om*t).*x.^2 + (1-2*dl*sin(om*t)).*x;
df = @(t,x) 2*dl*sin(om*t).*x + 1-2*dl*sin(om*t);
g = @(t,x,y) sin(pi*f(t,x)).*cos(pi*f(t,y)).*df(t,y);
w = @(t,v) -1/2*(Dx.*fft2(-g(t,X,Y).*ifft2(v))+Dy.*fft2(g(t,Y,X).*ifft2(v)) +
    fft2(-g(t,X,Y).*ifft2(Dx.*v))+fft2(g(t,Y,X).*ifft2(Dy.*v)));
v = @(t,x) reshape(w(t,reshape(x,M,M)),M^2,1);

%% intial values
N = 5;                                      % number of basis functions
U0 = zeros(M,M,M,M);
for k = 1:M, for l = 1:M, U0(k,l,k,l) = 1; end, end
Y0 = U0(:,:,[1:(N-1)/2+1,(M-1)-((N-1)/2-2):M], ...
            [1:(N-1)/2+1,(M-1)-((N-1)/2-2):M]);

%% time integration / construction of transfer operator
t0 = 0; t1 = 10.25; m = 50; h = (t1-t0)/m;
P = zeros(M^2,N^2);
for l = 1:N
    for k = 1:N
        y0 = reshape(Y0(:,:,k,l),M^2,1);
        P(:,N*(l-1)+k) = etdrk4(t0,m,h,L,v,y0);
    end
end
[UU,S,VV] = svd(P); diag(S);      % compute singular values and vectors

%% plot singular vectors
figure(1); clf; p = 128; [Xp,Yp] = meshgrid(2*(0:p-1)/p);
for j = 2:5
   subplot(2,2,j-1)
   v = reshape(VV(:,j),N,N);
   vp = zeros(p);
   vp([1:(N+1)/2,p-(N-1)/2+1:p],[1:(N+1)/2,p-(N-1)/2+1:p]) = v*p^2/N^2;
   vp = ifft2(vp,'symmetric');
   vp=vp/max(max(abs(vp))); %normed for same colorbar
   surf(Xp,Yp,vp), shading flat, view(0,90), caxis([-1 1]), axis equal, hold on,
        axis off, axis tight
end
```

Bibliography

[Ameres, 2016] Ameres, J. (2016). Particle in cell 2d for the guiding center model. http://home.in.tum.de/~ameres/matlab_examples/matlab_fem/html/Diocotron_particles.html. Accessed: 2016-12-28.

[Ameres et al., 2016] Ameres, J., Kormann, K., and Sonnendrücker, E. (2016). Particle in fourier discretization of kinetic equations. In *PASC: Proceedings of the Platform for Advanced Scientific Computing Conference*, New York, NY, USA. ACM.

[Ameres and Sonnendrücker, 2015] Ameres, J. and Sonnendrücker, E. (2015). A particle in fourier method for field aligned gyrokinetic models. In *Verhandlungen DPG (VI) 50, 1, 2015*.

[Ameres and Sonnendrücker, 2016] Ameres, J. and Sonnendrücker, E. (2016). Particle in fourier discretization of kinetic equations. In *Verhandlungen DPG (VI) 51, 1, 2016*.

[Arthur and Vassilvitskii, 2007] Arthur, D. and Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035.

[Banisch and Koltai, 2016] Banisch, R. and Koltai, P. (2016). Understanding the geometry of transport: diffusion maps for lagrangian trajectory data unravel coherent sets. *arXiv preprint arXiv:1603.04709*.

[Becker et al., 2007] Becker, R., Meidner, D., and Vexler, B. (2007). Efficient numerical solution of parabolic optimization problems by finite element methods. *Optimisation Methods and Software*, 22(5):813–833.

[Birdsall and Langdon, 2004] Birdsall, C. K. and Langdon, A. B. (2004). *Plasma physics via computer simulation*. CRC Press.

[Bittencourt, 2013] Bittencourt, J. A. (2013). *Fundamentals of plasma physics*. Springer Science & Business Media.

[Bittracher, 2016] Bittracher, A. (2016). *The generating structure of spatial conformation dynamics*. PhD thesis, Technische Universität München.

[Boyarsky and Gora, 2012] Boyarsky, A. and Gora, P. (2012). *Laws of Chaos: invariant measures and dynamical systems in one dimension*. Springer Science & Business Media.

[Boyd, 2001] Boyd, J. P. (2001). *Chebyshev and Fourier spectral methods*. Courier Dover Publications.

[Cheng et al., 2014] Cheng, Y., Gamba, I. M., Li, F., and Morrison, P. J. (2014). Discontinuous galerkin methods for the vlasov–maxwell equations. *SIAM Journal on Numerical Analysis*, 52(2):1017–1049.

[Cox and Matthews, 2002] Cox, S. and Matthews, P. (2002). Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455.

[del Castillo-Negrete and Morrison, 1993] del Castillo-Negrete, D. and Morrison, P. (1993). Chaotic transport by rossby waves in shear flow. *Physics of Fluids A: Fluid Dynamics (1989-1993)*, 5(4):948–965.

[Dellnitz et al., 2001] Dellnitz, M., Froyland, G., and Junge, O. (2001). The algorithms behind GAIO – Set oriented numerical methods for dynamical systems. In Fiedler, B., editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 145–174. Springer.

[Dellnitz and Junge, 1999] Dellnitz, M. and Junge, O. (1999). On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515.

[Denner et al., 2016] Denner, A., Junge, O., and Matthes, D. (2016). Computing coherent sets using the fokker-planck equation. *Journal of Computational Dynamics*, 3(2):3–3.

[Denner et al., 2017] Denner, A., Rowley, C. W., and Junge, O. (2017). Computing coherent sets using transfer operator families. manuscript in preparation.

[Deuflhard et al., 2000] Deuflhard, P., Huisinga, W., Fischer, A., and Schütte, C. (2000). Identification of almost invariant aggregates in reversible nearly uncoupled markov chains. *Linear Algebra and its Applications*, 315(1):39–59.

[Deuflhard and Schütte, 2004] Deuflhard, P. and Schütte, C. (2004). Molecular conformation dynamics and computational drug design. *Applied Mathemetics Entering the 21st Century. Proceedings ICIAM*, pages 91–119.

[Deuflhard and Weber, 2005] Deuflhard, P. and Weber, M. (2005). Robust perron cluster analysis in conformation dynamics. *Linear algebra and its applications*, 398:161–184.

[Evans, 2010] Evans, L. (2010). *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society.

[Fasshauer, 2007] Fasshauer, G. E. (2007). *Meshfree approximation methods with MATLAB*, volume 6. World Scientific.

[Filbet and Sonnendrücker, 2003] Filbet, F. and Sonnendrücker, E. (2003). Numerical methods for the vlasov equation. In *Numerical mathematics and advanced applications*, pages 459–468. Springer.

[Froyland, 2013] Froyland, G. (2013). An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. *Physica D: Nonlinear Phenomena*.

[Froyland, 2015] Froyland, G. (2015). Dynamic isoperimetry and the geometry of lagrangian coherent structures. *Nonlinearity*, 28(10):3587.

[Froyland et al., 2013] Froyland, G., Junge, O., and Koltai, P. (2013). Estimating long-term behavior of flows without trajectory integration: The infinitesimal generator approach. *SIAM Journal on Numerical Analysis*, 51(1):223–247.

[Froyland and Kwok, 2016] Froyland, G. and Kwok, E. (2016). A dynamic laplacian for identifying lagrangian coherent structures on weighted riemannian manifolds. *arXiv preprint arXiv:1610.01128*.

[Froyland et al., 2010a] Froyland, G., Lloyd, S., and Santitissadeekorn, N. (2010a). Coherent sets for nonautonomous dynamical systems. *Physica D: Nonlinear Phenomena*, 239(16):1527–1541.

[Froyland and Padberg, 2009] Froyland, G. and Padberg, K. (2009). Almost-invariant sets and invariant manifolds—connecting probabilistic and geometric descriptions of coherent structures in flows. *Physica D: Nonlinear Phenomena*, 238(16):1507–1523.

[Froyland and Padberg-Gehle, 2014] Froyland, G. and Padberg-Gehle, K. (2014). Almost-invariant and finite-time coherent sets: directionality, duration, and diffusion. In *Ergodic Theory, Open Dynamics, and Coherent Structures*, pages 171–216. Springer.

[Froyland and Padberg-Gehle, 2015] Froyland, G. and Padberg-Gehle, K. (2015). A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087406.

[Froyland et al., 2010b] Froyland, G., Santitissadeekorn, N., and Monahan, A. (2010b). Transport in time-dependent dynamical systems: Finite-time coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):043116.

[Gibbs, 1902] Gibbs, J. W. (1902). *Elementary principles in statistical mechanics*. C. Scribner's sons.

[Griebel et al., 2007] Griebel, M., Knapek, S., and Zumbusch, G. (2007). *Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications*, volume 5. Springer Science & Business Media.

[Hadjighasem et al., 2016] Hadjighasem, A., Karrasch, D., Teramoto, H., and Haller, G. (2016). Spectral-clustering approach to lagrangian vortex detection. *Phys. Rev. E*, 93:063107.

[Haller, 2000] Haller, G. (2000). Finding finite-time invariant manifolds in two-dimensional velocity fields. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 10(1):99–108.

[Haller, 2001] Haller, G. (2001). Distinguished material surfaces and coherent structures in three-dimensional fluid flows. *Physica D: Nonlinear Phenomena*, 149(4):248–277.

[Haller and Beron-Vera, 2012] Haller, G. and Beron-Vera, F. J. (2012). Geodesic theory of transport barriers in two-dimensional flows. *Physica D: Nonlinear Phenomena*, 241(20):1680–1702.

[Hirzebruch and Scharlau, 1971] Hirzebruch, F. and Scharlau, W. (1971). *Einführung in die Funktionalanalysis*. BI Hochschultaschenbücher. Bibliographisches Institut.

[Hockney and Eastwood, 1988] Hockney, R. W. and Eastwood, J. W. (1988). *Computer simulation using particles*. CRC Press.

[Huisinga and Schmidt, 2006] Huisinga, W. and Schmidt, B. (2006). Metastability and dominant eigenvalues of transfer operators. In *New Algorithms for Macromolecular Simulation*, pages 167–182. Springer.

[Junge, 1999] Junge, O. (1999). *Mengenorientierte Methoden zur numerischen Analyse dynamischer Systeme*. Shaker.

[Karrasch and Keller, 2016] Karrasch, D. and Keller, J. (2016). A geometric heat-flow theory of lagrangian coherent structures. *arXiv preprint arXiv:1608.05598*.

[Kassam and Trefethen, 2005] Kassam, A.-K. and Trefethen, L. N. (2005). Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, 26(4):1214–1233.

[Klimas and Farrell, 1994] Klimas, A. and Farrell, W. (1994). A splitting algorithm for vlasov simulation with filamentation filtration. *Journal of computational physics*, 110(1):150–163.

[Koltai, 2010] Koltai, P. (2010). *Efficient approximation methods for the global long-term behavior of dynamical systems – Theory, algorithms and examples*. PhD thesis, Technische Universität München.

[Kraus et al., 2016] Kraus, M., Kormann, K., Morrison, P. J., and Sonnendrücker, E. (2016). Gempic: Geometric electromagnetic particle-in-cell methods. *arXiv preprint arXiv:1609.03053*.

[Kunita, 1997] Kunita, H. (1997). *Stochastic flows and stochastic differential equations*, volume 24. Cambridge university press.

[Lasota and Mackey, 1993] Lasota, A. and Mackey, M. (1993). *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, volume 97 of *Applied Mathematical Sciences*. Springer.

[Levy, 1965] Levy, R. (1965). Diocotron instability in a cylindrical geometry. *Physics of Fluids (1958-1988)*, 8(7):1288–1295.

[Lloyd, 1982] Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.

[Ma and Bollt, 2013] Ma, T. and Bollt, E. M. (2013). Relatively coherent sets as a hierarchical partition method. *International Journal of Bifurcation and Chaos*, 23(07):1330026.

[Nave, 2008] Nave, J.-C. (2008). Computational science and engineering. http://ocw.mit.edu/courses/mathematics/18-336-numerical-methods-for-partial-differential-equations-spring-2009/matlab-files/. (Massachusetts Institute of Technology: MIT OpenCouseWare), Accessed 2015-07-13. License: Creative Commons BY-NC-SA.

[Oksendal, 2003] Oksendal, B. (2003). Stochastic differential equations: an introduction with applications. *Springer Science & Business Media*.

[Padberg et al., 2007] Padberg, K., Hauff, T., Jenko, F., and Junge, O. (2007). Lagrangian structures and transport in turbulent magnetized plasmas. *New Journal of Physics*, 9(11):400.

[Revuz, 2008] Revuz, D. (2008). *Markov Chains*. North-Holland Mathematical Library. Elsevier Science.

[Rowley et al., 2009] Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P., and Henningson, D. S. (2009). Spectral analysis of nonlinear flows. *Journal of fluid mechanics*, 641:115–127.

[Rypina et al., 2007] Rypina, I. I., Brown, M. G., Beron-Vera, F. J., Koak, H., Olascoaga, M. J., and Udovydchenkov, I. A. (2007). On the lagrangian dynamics of atmospheric zonal jets and the permeability of the stratospheric polar vortex. *Journal of the Atmospheric Sciences*, 64(10):3595–3610.

[Rypina et al., 2014] Rypina, I. I., Kirincich, A. R., Limeburner, R., and Udovydchenkov, I. A. (2014). Eulerian and Lagrangian correspondence of high-frequency radar and surface drifter data: effects of radar resolution and flow components. *Journal of Atmospheric and Oceanic Technology*, 31(4):945–966.

[Shadden et al., 2005] Shadden, S. C., Lekien, F., and Marsden, J. E. (2005). Definition and properties of lagrangian coherent structures from finite-time lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, 212(3):271–304.

[Shoucri, 1979] Shoucri, M. M. (1979). Nonlinear evolution of the bump-on-tail instability. *Physics of Fluids (1958-1988)*, 22(10):2038–2039.

[Sonnendrücker et al., 1999] Sonnendrücker, E., Roche, J., Bertrand, P., and Ghizzo, A. (1999). The semi-lagrangian method for the numerical resolution of the vlasov equation. *Journal of computational physics*, 149(2):201–220.

[Tröltzsch, 2005] Tröltzsch, F. (2005). *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen*. Vieweg Verlag.

[Ulam, 1960] Ulam, S. M. (1960). *Problems in modern mathematics*. Courier Dover Publications.

[Wendland, 2004] Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

[Williams et al., 2014a] Williams, M. O., Kevrekidis, I. G., and Rowley, C. W. (2014a). A data-driven approximation of the koopman operator: extending dynamic mode decomposition. *arXiv preprint arXiv:1408.4408*.

[Williams et al., 2014b] Williams, M. O., Rypina, I. I., and Rowley, C. W. (2014b). Identifying finite-time coherent sets from limited quantities of lagrangian data. *arXiv preprint arXiv:1412.7441*.

[Yosida, 1995] Yosida, K. (1995). *Functional analysis*. Classics in Mathematics. Springer, 6th edition.

[Zang, 1991] Zang, T. A. (1991). On the rotation and skew-symmetric forms for incompressible flow simulations. *Applied Numerical Mathematics*, 7(1):27–40.