MSc Thesis Applied Mathematics

A unifying framework for estimation of the Koopman operator between Reproducing Kernel Hilbert Spaces

T.P. Middelkamp

Supervisor: Dr. rer. nat. J. A. Iglesias Martínez
Supervisor: dr. F.L. Schwenninger
Chair graduation committe: prof.dr. C. Brune

December, 2024

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

A unifying framework for estimation of the Koopman operator between Reproducing Kernel Hilbert Spaces

Thomas Middelkamp

December 16, 2024

Abstract

Dynamical systems, and estimations thereof, play an important role in various disciplines. The Koopman operator encapsulates properties of a dynamical system. This thesis investigates the estimation of the Koopman operator within the context of Reproducing Kernel Hilbert Spaces (RKHSs). We begin by reviewing the relevant background on RKHSs, including the vector-valued case, and Koopman operators in their natural setting of continuous functions. A general framework that links Koopman theory and discrete-time dynamical systems is provided. We then compare two methods for estimating Koopman operators in a unified framework, namely ridge regression in spaces of Hilbert-Schmidt operators on a RKHS and kernel Extended Dynamic Mode Decomposition. The boundedness of the actual Koopman operator between RKHSs is investigated and illustrated through examples. Dynamics for which the Koopman operator is bounded between Gaussian RKHSs on $\mathbb{R}^d$ are characterized.

Acknowledgements

First of all I would like to thank my supervisors, José Iglesias and Felix Schwenninger. Thank you for being great teachers, I learned a lot during this process.
Second, thank you to all of my friends in no particular order. Bozho, for all the support throughout the years we have been studying together. I couldn't have wished for a better time, thanks for proofreading and listening to my presentation. Daan, thanks for always being there and working together in Langezijds and the Vrijhof, I'm gonna miss these times. All of my (former) roommates, you made my time as a student incredible. Giorgio, Giacomo and Tjeerd-Jan, thanks for listening and giving me feedback halfway through my thesis. Nathan, it was amazing to work (read: struggle) with you the last months of our theses in the Vrijhof. Thanks to Hugo, Lars and Yang-Li for always letting me use their office.
Thanks to my parents and my brothers, for their unconditional support.
Finally, I need to thank Bella. Thank you for all the dinners, for always being there for me and helping me write this thesis. Your support has been incredible, not only in writing but also putting the process in perspective! And thank for you for making the only figure in this report :)

Contents

Chapter 1

Introduction

Dynamical systems play an important role in various disciplines including chemistry, physics, and engineering. Numerous relevant dynamical systems in these respective fields are nonlinear, which may lead to bifurcations and chaos [SH98]. Think of the Lorentz system or a double pendulum in classical mechanics. Linear dynamical systems, however, are completely determined by their spectral decomposition. Koopman operator theory transforms a nonlinear system into a linear one, by acting on an infinite-dimensional function space [Koo31]; [Bru+21]. The benefit of this approach is that the Koopman operator is linear and hence allows for analyzing complex nonlinear systems using linear methods. Koopman theory has gained momentum due to advances in computational techniques and the increasing availability of large datasets, which make it feasible to develop data-driven, operator-theoretic models of complex dynamics [GE22]. However, one of the primary challenges in this field remains the development of finite-dimensional approximations that retain the essential structure of these infinite-dimensional representations.

An approach that tackles this issue is the Koopman mode decomposition, introduced by Mezić [Mez05]. This method allows us to represent the dynamics of a nonlinear system in terms of a sequence of triples $(\lambda_j, \varphi_j, \mathbf{v}_j)$, where λ_j are the Koopman eigenvalues, φ_j are the associated Koopman eigenfunctions and $\mathbf{v}_j$ are the so-called Koopman modes. The nonlinear dynamics are converted through this decomposition into a linear form in the space of these eigenfunctions. We can thus understand complex nonlinear behavior through the spectral properties of the Koopman operator. This Koopman mode expansion can be approximated by a finite sum of only the most significant modes and eigenvalues, making it computationally feasible for practical applications.

One method to approximate the Koopman mode decomposition is the Dynamic Mode Decomposition (DMD), first introduced by Schmid [Sch10]. Initially developed in the fluid dynamics community as a tool to identify coherent structures within flows [Sch10], later to be connected to Koopman theory by Rowley [Row+09]. The DMD algorithm is successful partly because it is a fully data-driven method, it is fast due to its connections to the SVD, and the fact that it is easy to implement. In DMD, the eigenvalues computed from data snapshots serve as approximations of the Koopman eigenvalues λ_j, the DMD modes correspond to the Koopman modes $\mathbf{v}_j$, and the DMD mode amplitudes approximate the values of the Koopman eigenfunctions evaluated at the initial state of the observed dynamical system $\varphi_j(\mathbf{x}_0)$.

Kernel methods have become increasingly popular in data-driven modeling of dynamical systems, especially in combination with Koopman theory[Bat+24]; [Kos+22]. Kernels allow us to consider the non-linear relationships between the data points, as well as the non-linear relation to the system behaviors in high-dimensional spaces [HSS08]. All of the developed kernel methods find their mathematical foundation in Reproducing Kernel Hilbert Space (RKHS) theory, introduced by Aronszajn [Aro50]. A benefit of RKHSs is the presence of so the so-called representer theorem, introduced by Kimeldorf and Wahba [KW70]; [KW71] and later generalized by Schölkopf [SHS01].

This theorem allows us to write the general solutions of certain regularization functional in high-dimensional or infinite-dimensional space as a linear combination of the elements of a data-defined finite-dimensional subspace. This enables us to convert nonconvex optimization problems in high or infinite dimensions, which often are intractable, into convex optimization problems over scalar coefficients.

There is a fundamental conceptual difference concerning the types of approximation offered by DMD versus kernel methods. DMD remains confined to the subspace spanned by the vectors that comprise the observed data snapshots and thus can only capture the dynamics present in these data [Bru+21]. This makes DMD a localized method that cannot extrapolate well outside the data. On the other hand, kernel methods embed the data into a higher-dimensional space shaped by the kernel function, which is an a priori belief about how points are related [HSS08]. This makes it possible for kernel methods to go beyond the limits of the data provided, as the assumed relationships introduced by the kernel allow extrapolation. There is, however, a downside due to this additional flexibility. There is a strong assumption that the kernel accurately characterizes the true structure of the system. If this is not the case, the extrapolation will be inaccurate.

1.1 Our contribution

Most papers that combine kernel methods and the Koopman operator are brief on the foundation of the underlying topics. We do not assume knowledge of RKHSs or Koopman theory and build up an extensive theoretical framework. Furthermore, we contrast the more practical part with a rather theoretical view of the Koopman operator between spaces of continuous functions. Within this framework, we compare two methods that estimate the Koopman operator. We contrast a more recent data-driven kernel ridge regression method in the space of Hilbert-Schmidt operators and kernel Extended DMD. We state several examples in the positive and the negative for the boundedness of the Koopman operator between RKHS. Lastly, we characterize dynamics on Gaussian RKHSs for which the Koopman is bounded, which combines work from [Gon+24]; [Köh+24].

1.2 Outline

In Chapter 2, we start by introducing the mathematical foundation of scalar-valued RKHSs. We discuss and characterize the boundedness of composition operators on RKHSs. We end by introducing vector-valued RKHSs and stating important results for learning theory. In Chapter 3, we view the continuous functions on a compact set K as a C^*-algebra. We show that the Koopman operator between these spaces is uniquely determined by the underlying dynamics. In Chapter 4, we introduce the Koopman paradigm for dynamical systems from snapshot data. Then, we explain DMD algorithms and make a comparison with a recent data-driven kernel-based method. We end this chapter by commenting on the boundedness of the actual Koopman operator between RKHSs with the use of examples. We end with a summary, limitations, and outlook in Chapter 5.

Chapter 2

Reproducing kernel Hilbert spaces

First, we will give the abstract definition of a RKHS and provide several (non)examples. Then we will define what (reproducing) kernels are and show how they are related to RKHSs. We will provide an explicit construction of a RKHS from a kernel. We conclude the theory of scalar-valued RKHSs with characterizations of functions in RKHSs and characterize the boundedness of the composition operator. We extend to vector-valued RKHS, hereafter we present representer theorems and an isomorphism between a vector-valued RKHS and the space of Hilbert-Schmidt operators.

2.1 Basics of RKHS

2.1.1 Definition and non-example

Definition 2.1.1 (Evaluation functional)**.** Let H be a Hilbert space of functions on a non-empty set $\mathcal{X}$. Let $f \in H$ the *evaluation functional* at $x \in \mathcal{X}$ is the linear map defined as

$$\begin{aligned} \delta_x : H &\to \mathbb{R}, \\ f &\mapsto \delta_x(f) := f(x). \end{aligned}$$

Definition 2.1.2 (Reproducing kernel Hilbert space)**.** A *reproducing kernel Hilbert space (RKHS)* is a Hilbert space of complex-valued functions on a non-empty set $\mathcal{X}$ where the evaluation functionals are bounded for every $x \in \mathcal{X}$.

We observe that a Hilbert space of functions can only be a RKHS if norm convergence implies pointwise convergence.

Observation 2.1.3 (Norm convergence implies pointwise convergence in a RKHS)**.** Let H be a RKHS and $(f_n)_{n\in\mathbb{N}}$ in H such that $f_n \to f$, we have

$$|f_n(x) - f(x)| = |\delta_x(f_n) - \delta_x(f)| = |\delta_x(f_n - f)| \leq \|\delta_x\|\|f_n - f\|_H \to 0 \quad (n \to \infty).$$

To get an intuition what function spaces are RKHSs, it can help to look at a non-example. To do so, we generalize a RKHS to non-complete spaces.

Definition 2.1.4 (Reproducing kernel inner product space)**.** A *reproducing kernel inner product space (RKIS)* is an inner product space of complex-valued functions on a non-empty set $\mathcal{X}$ where the evaluation functionals are bounded for every $x \in \mathcal{X}$.

Proposition 2.1.5. A Hilbert space is a RKHS on a set $\mathcal{X}$ if and only every dense subspace $D \subseteq H$ is a RKIS on $\mathcal{X}$.

Proof. Let H be a RKHS, it is clear that any dense subspace D is a RKIS. Conversely, let D be a RKIS that is a dense subspace of a Hilbert space H. Let $f \in H$ and $x \in \mathcal{X}$, there exists a sequence $(h_n)_{n\in\mathbb{N}}$ in D such that $h_n \to f$. We see that H is a RKHS, as point evaluations are bounded

$$|\delta_x(f)| = |\delta_x(\lim_{n\to\infty} h_n)| = \lim_{n\to\infty} |\delta_x(h_n)| \leq \lim_{n\to\infty} \|\delta_x\|\|h_n\| = \|\delta_x\|\|f\|.$$

□

Using Proposition 2.1.5, we can show that $L^2(0,1)$ is not a RKHS. We do so by showing that a dense subspace, $C[0,1]$ equipped with the standard inner product, is not a RKIS.

Example 2.1.6 ($L^2(0,1)$ is not a RKHS)**.** Equip $C[0,1]$ with the standard inner product, so that $C[0,1]$ is dense in $L^2(0,1)$. Define a sequence of functions in $C[0,1]$

$$f_n(t) = \begin{cases} (\frac{t}{x})^n & 0 \leq t \leq x, \\ (\frac{1-t}{1-x})^n & x < t \leq 1. \end{cases}$$

Let $x \in (0,1)$, $M > 0$ and define $g_n = \frac{f_n}{\|f_n\|_{L^2}}$. Because $f_n(x) = 1$ for all $n \in \mathbb{N}$ and $\|f_n\|_{L^2} \to 0$ as $n \to \infty$, there exists $n(M) \in \mathbb{N}$ such that

$$g_{n(M)}(x) = \frac{f_{n(M)}(x)}{\|f_{n(M)}\|_{L^2}} = \frac{1}{\|f_{n(M)}\|_{L^2}} > M/2.$$

By continuity of $g_{n(M)}$, there exists an interval $I(M)$, that includes x, such that $g_{n(M)}(t) > M/2$ for all $t \in I(M)$. Since $\|g_n\| = 1$ for all $n \in \mathbb{N}$ we have that $\|\delta_y\| > M/2$ for $y \in I(M)$. Observe that for $N > M$ we have that the interval $I(N)$ is contained in $I(M)$. Letting $N \to \infty$, we conclude that the evaluation functionals are unbounded on the interval $I(M)$, in the L^2 sense. Such an interval can be constructed for any $0 < x < 1$, for $x = 0$ consider the function

$$f_n(t) = \begin{cases} (1 - nt) & 0 \leq t \leq \frac{1}{n}, \\ 0 & \frac{1}{n} < t \leq 1. \end{cases}$$

Then, $f_n(0) = 1$ for all $n \in \mathbb{N}$ and $\|f_n\|_{L^2} \to 0$ as $n \to \infty$ and a similar argument can be made, also for $x = 1$. Since $(C[0,1], \langle\cdot,\cdot\rangle_{L^2})$ is dense in $(L^2(0,1), \langle\cdot,\cdot\rangle_{L^2})$ and not a RKIS, we conclude by Proposition 2.1.5 that $(L^2(0,1), \langle\cdot,\cdot\rangle_{L^2})$ cannot be a RKHS.

Remark. Consider the space $(L^2(0,1), \langle\cdot,\cdot\rangle_{L^2})$ and the function $f_n(t) = \mathbf{1}_{[0,1/n]}$. Then, f_n converges to zero in norm and from Observation 2.1.3 we have $|f_n(0)| \leq \|f_n\|_{L^2} \to 0$ as $n \to \infty$. This appears to be a contradiction, since $|f_n(0)| = 1 \nrightarrow 0$. However, recall that point evaluations are meaningless, as we can modify f_n at $t = 0$ and remain in the same equivalence class of functions. To conclude, since the above argument depends on the representative of $f_n \in L^2(0,1)$, it is not sufficient to conclude that $L^2(0,1)$ is *not* a RKHS.

2.1.2 Reproducing kernels

We have not defined what a kernel precisely means, yet we have defined a RKHS which contains the word "kernel". To bridge this gap, we will define a RKHS from a given "kernel". Then, we show that these two definitions are equivalent. To do so, we have to define what a kernel is first.

Definition 2.1.7 (Positive semidefinite kernel)**.** Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a bivariate function. The map k is said to be a *positive semidefinite (PSD) kernel* if it is conjugate symmetric and if for any collection $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{C}^n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_j)\overline{\mathbf{a}_j} \geq 0. \tag{2.1}$$

We allow functions in a RKHS and a kernel to take on complex values. In most examples that follow we will restrict to the real case and will return to complex-valued RKHS at a later time.

Remark. Given $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{C}^n$, the matrix $\mathbf{K}_{ij} := k(\mathbf{x}_i, \mathbf{x}_j)$ is a positive semidefinite matrix, since

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_j)\overline{\mathbf{a}}_j = \mathbf{a}^T \mathbf{K} \overline{\mathbf{a}} \geq 0.$$

We will also call this matrix $\mathbf{K}$, that is induced by the kernel k once $\mathbf{x} \in \mathcal{X}^n$ is chosen.

For brevity, we will refer to a positive semidefinite kernel as a kernel from now on. We make a simple yet important observation about the pointwise sum and product of two kernels.

Observation 2.1.8 (Sums and products of kernels are kernels)**.** Let k_1, k_2 be kernels on a set $\mathcal{X}$. Then $k_s(x,y) := k_1(x,y) + k_2(x,y)$ and $k_p(x,y) := k_1(x,y) \cdot k_2(x,y)$ are kernels, too. Choose $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{a} \in \mathbb{C}^n$ and denote by K_1 and K_2 be the matrices induced by k_1 and k_2, respectively. The sum $K_s := K_1 + K_2$. is PSD, as it is the sum of PSD matrices.
Let K_p be the Hadamard product of K_1 and K_2. By the Schur product Theorem[1], it follows that K_p is PSD.

Example 2.1.9 (Linear kernel)**.** An example is the *linear kernel* on $\mathbb{R}^d$, defined as $k(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, \mathbf{y} \rangle$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathbb{R}^d}$ is the standard inner product on $\mathbb{R}^d$. It is clear that the standard inner product (on $\mathbb{R}^d$) is symmetric, and also PSD for any subset $\mathbf{x} \in \mathbb{R}^d$ because for any $\mathbf{a} \in \mathbb{R}^d$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i \langle \mathbf{x}_i, \mathbf{x}_j \rangle \mathbf{a}_j \overset{\text{lin.}}{=} \langle \sum_{i=1}^{n} \mathbf{a}_i \mathbf{x}_i, \sum_{j=1}^{n} \mathbf{a}_j \mathbf{x}_j \rangle = \| \sum_{i=1}^{n} \mathbf{a}_i \mathbf{x}_i \|^2 \geq 0.$$

To illustrate what the interplay is between kernels and feature maps, we give some examples.

Definition 2.1.10 (Feature map)**.** Let $\mathcal{X}$ be a non-empty set. A map $\Phi : \mathcal{X} \to \ell^2(\mathbb{N})$, or a (finite dimensional) subspace thereof, is called a *feature map*.

Example 2.1.11 (Feature map for quadratic kernel I)**.** We will define the feature map $\Phi : \mathbb{R}^d \to \mathbb{R}^D$, where $D = d + \binom{d}{2}$.

$$\Phi(\mathbf{x}) = \begin{cases} \mathbf{x}_j^2 & \text{for } j = 1, 2, \ldots, d, \\ \sqrt{2}\mathbf{x}_i \mathbf{x}_j & \text{for } i < j. \end{cases}$$

Take $d = 3$ and let $\mathbf{x} \in \mathbb{R}^3$, then

$$\Phi(\mathbf{x}) = (\mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_3^2, \sqrt{2}\mathbf{x}_1\mathbf{x}_2, \sqrt{2}\mathbf{x}_1\mathbf{x}_3, \sqrt{2}\mathbf{x}_2\mathbf{x}_3).$$

Example 2.1.12 (Homogeneous polynomial kernel)**.** Take the *homogeneous polynomial kernel* on $\mathbb{R}^d$ defined as $k(\mathbf{x}, \mathbf{y}) := (\langle \mathbf{x}, \mathbf{y} \rangle)^m$ for some $m \geq 2$. Recall Observation 2.1.8 to note that $k(\mathbf{x}, \mathbf{y})$ is a kernel, since it is a power of the linear kernel. Take $m = 2$, we get the *quadratic kernel*

$$k(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle)^2 = \sum_{j=1}^{d} \mathbf{x}_j^2 \mathbf{y}_j^2 + 2\sum_{i<j} \mathbf{x}_i \mathbf{y}_j.$$

Observation 2.1.13 (Feature map for quadratic kernel II)**.** Observe that the feature map from Example 2.1.11 is closely related to the quadratic kernel $k(x,y) = (\langle x, y \rangle)^2$. As before, choose $d = 3$ so that $D = 3 + \binom{3}{2} = 6$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, we rewrite the inner product of the feature maps in $\mathbb{R}^6$.

$$\begin{aligned} \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathbb{R}^6} &= \langle (\mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_3^2, \sqrt{2}\mathbf{x}_1\mathbf{x}_2, \sqrt{2}\mathbf{x}_1\mathbf{x}_3, \sqrt{2}\mathbf{x}_2\mathbf{x}_3), (\mathbf{y}_1^2, \mathbf{y}_2^2, \mathbf{y}_3^2, \sqrt{2}\mathbf{y}_1\mathbf{y}_2, \sqrt{2}\mathbf{y}_1\mathbf{y}_3, \sqrt{2}\mathbf{y}_2\mathbf{y}_3) \rangle_{\mathbb{R}^6}, \\ &= \mathbf{x}_1^2\mathbf{y}_1^2 + \mathbf{x}_2^2\mathbf{y}_2^2 + \mathbf{x}_3\mathbf{y}_3^2 + 2\mathbf{x}_1\mathbf{x}_2\mathbf{y}_1\mathbf{y}_2 + 2\mathbf{x}_1\mathbf{x}_3\mathbf{y}_1\mathbf{y}_3 + 2\mathbf{x}_2\mathbf{x}_3\mathbf{y}_2\mathbf{y}_3, \\ &= (\mathbf{x}_1\mathbf{y}_1 + \mathbf{x}_2\mathbf{y}_2 + \mathbf{x}_3\mathbf{y}_3)^2, \\ &= (\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^3})^2. \end{aligned}$$

[1] Appendix ?

We see that the inner product between the features can be computed as an inner product between the points itself. For large values of d, there is a substantial computational difference in evaluating $\langle \Phi(x), \Phi(y) \rangle_{\mathbb{R}^D}$ and $(\langle x, y \rangle_{\mathbb{R}^3})^2$. Solely computing $\Phi(x)$, for some $x \in \mathbb{R}^d$, takes $O(d + \binom{d}{2}) = O(d + d^2)$, whereas the inner product $(\langle x, y \rangle_{\mathbb{R}^d})^2$ takes $O(d)$. For large values of d, this difference is substantial.

Example 2.1.14 (Inhomogeneous polynomial kernel)**.** In Example 2.1.12, we have seen the homogeneous polynomial kernel. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $c \in \mathbb{R}$, the inhomogeneous polynomial kernel is defined as $k(\mathbf{x}, \mathbf{y}) = (c + \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^d})^m$. It can be shown that the matrix with all entries equal to one is PSD, hence $k(\mathbf{x}, \mathbf{y}) = c$ is a kernel. The power of a linear kernel is a kernel too, hence the sum $k(\mathbf{x}, \mathbf{y}) = (c + \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^d})^m$ is a kernel.
Next, we will try to derive a feature map $\Phi : \mathbb{R}^d \to \mathbb{R}^D$, where $D = \binom{d+m}{m}$, such that $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathbb{R}^D}$. Using the multinomial theorem, we get

$$
\begin{aligned}
k(\mathbf{x}, \mathbf{y}) = (c + \langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{R}^d})^m = (c + \sum_{j=1}^{d} \mathbf{x}_j \mathbf{y}_j)^m &= (c + \mathbf{x}_1 \mathbf{y}_1 + \ldots + \mathbf{x}_d \mathbf{y}_d)^m \\
&= \sum_{\substack{k_0 + k_1 + \ldots + k_d = n \\ k_0, k_1, \ldots, k_d \geq 0}} \binom{n}{k_0, k_1, \ldots, k_d} c^{k_0} \prod_{i=1}^{d} (\mathbf{x}_j \mathbf{y}_j)^{k_i}, \\
&= \sum_{\substack{k_0 + k_1 + \ldots + k_d = n \\ k_0, k_1, \ldots, k_d \geq 0}} \binom{n}{k_0, k_1, \ldots, k_d} c^{k_0} \prod_{i=1}^{d} (\mathbf{x}_i \mathbf{y}_i)^{k_i}, \\
&= \sum_{\substack{k_0 + k_1 + \ldots + k_d = n \\ k_0, k_1, \ldots, k_d \geq 0}} \underbrace{\binom{n}{k_0, k_1, \ldots, k_d} c^{k_0}}_{:= b_n} (\mathbf{x}_1 \mathbf{y}_1)^{k_1} (\mathbf{x}_2 \mathbf{y}_2)^{k_2} \ldots (\mathbf{x}_d \mathbf{y}_d)^{k_d}, \\
&= \sum_{\substack{k_0 + k_1 + \ldots + k_d = n \\ k_0, k_1, \ldots, k_d \geq 0}} \left(\sqrt{b_n} \mathbf{x}_1^{k_1} \mathbf{x}_2^{k_2} \ldots \mathbf{x}_d^{k_d} \right) \left(\sqrt{b_n} \mathbf{y}_1^{k_1} \mathbf{y}_2^{k_2} \ldots \mathbf{y}_d^{k_d} \right), \\
&= \sum_{\substack{k_0 + k_1 + \ldots + k_d = n \\ k_0, k_1, \ldots, k_d \geq 0}} \left(\sqrt{b_n} \prod_{i=1}^{d} \mathbf{x}_i^{k_i} \right) \left(\sqrt{b_n} \prod_{i=1}^{d} \mathbf{y}_i^{k_i} \right), \\
&= \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathbb{R}^D}.
\end{aligned}
$$

The feature map $\Phi : \mathbb{R}^d \to \mathbb{R}^D$ contains all components such that $k_0 + k_1 + \ldots + k_d = m$ and that $k_i \in \mathbb{N}_0$ for $1 \leq i \leq d$. A component of $\Phi(x)$ looks like

$$
\Phi(x) = (\ldots, \sqrt{\binom{n}{\tilde{k}_0, \tilde{k}_1, \ldots, \tilde{k}_d}} c^{\tilde{k}_0} \prod_{i=1}^{d} \mathbf{x}_i^{\tilde{k}_i}, \ldots),
$$

where $(\tilde{k}_i)_{i=1}^d \in \mathbb{N}_0$ are one of the $\binom{d+m}{m}$ possibilities such that $\sum_i \tilde{k}_i = m$. Directly evaluating $(c + \langle \mathbf{x}, \mathbf{y} \rangle)^m$ can be done in $O(d + m)$, whereas computing $\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathbb{R}^D}$ is done in $O(\binom{d+m}{m})$, a significant difference for large d and m.

Definition 2.1.15 (Reproducing kernel)**.** Let H be a Hilbert space of real-valued functions on a non-empty set $\mathcal{X}$. A bivariate map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is a *reproducing kernel* of H if

1. $k(x, \cdot) \in H$ for all $x \in \mathcal{X}$,
2. $\langle f, k(x, \cdot) \rangle = f(x)$ for all $f \in H$ and for all $x \in \mathcal{X}$.

The second requirement is called the reproducing property.

If requirements one and two are satisfied for a Hilbert space H, we say that H *admits* the reproducing kernel k. The name "reproducing kernel" is justified, since a reproducing kernel is a kernel, too.

Proposition 2.1.16 (Reproducing kernel is a kernel)**.** We need to show that k is conjugate symmetric and positive semidefinite. The conjugate symmetry follows from the reproducing property and the conjugate symmetry of the inner product

$$\overline{k(y,x)} = \overline{\langle k(y,\cdot), k(x,\cdot)\rangle} = \langle k(x,\cdot), k(y,\cdot)\rangle = k(x,y) \quad \forall x,y \in \mathcal{X}.$$

Take any collection $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{C}^n$ nonzero, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_j)\overline{\mathbf{a}_j} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i \langle k(\mathbf{x}_i,\cdot), k(\mathbf{x}_j,\cdot\rangle\overline{\mathbf{a}_j} = \langle \sum_{i=1}^{n} \mathbf{a}_i \mathbf{x}_i, \sum_{j=1}^{n} \mathbf{a}_j \mathbf{x}_j\rangle \geq 0.$$

At this point it would, at least semantically, make sense to define a RKHS as a Hilbert space of functions on a non-empty set $\mathcal{X}$ that contains a reproducing kernel. It turns out that this definitions is sensible and equivalent to point evaluations being bounded, as explained in the following theorem.

Theorem 2.1.17. A Hilbert space of functions on $\mathcal{X}$ is a RKHS if an only if it admits a reproducing kernel.

Proof. Suppose that H is a RKHS, by Riesz-Fréchet there exists a unique $g_x \in H$ such that $\delta_x(\cdot) = \langle \cdot, g_x\rangle$. Define $k(x,\cdot) := g_x(\cdot) \in H$ for all $x \in \mathcal{X}$. To show the reproducing property, take $f \in H$, we see that $\langle f, k(x,\cdot)\rangle = \langle f, g_x\rangle = \delta_x(f) = f(x)$. Conversely, suppose that H is a Hilbert space of functions that contains a reproducing kernel $k(x,\cdot)$ for any $x \in \mathcal{X}$. Let $f \in G$, by Cauchy-Schwarz wee see that all point evaluations are bounded

$$|\delta_x(f)| = |\langle f, k(x,\cdot)\rangle| \leq \|f\|\|k(x,\cdot)\| \quad \forall x \in \mathcal{X}.$$

□

Proposition 2.1.18 (Reproducing kernel is unique)**.** Suppose that H is a RKHS that admits two reproducing kernels, k_1 and k_2. Then $k_1 = k_2$.

Proof. For any $f \in H$ and $x \in \mathcal{X}$, $\langle f, k_1(x,\cdot)\rangle = f(x) = \langle f, k_2(x,\cdot)\rangle$ by the reproducing property. Choose $f = k_1(x,\cdot) - k_2(x,\cdot)$, definiteness of the inner product implies that $k_1 = k_2$. □

We will denote a RKHS H that admits the kernel k as H_k or $H(k)$ or simply H, depending on the context.

Proposition 2.1.19. Let H be a RKHS on $\mathcal{X}$ that admits the kernel k, then $\operatorname{span}\{k(x,\cdot) : x \in \mathcal{X}\}$ is dense in H.

Proof. Take $f \in \operatorname{span}\{k(x,\cdot) : x \in \mathcal{X}\}^{\perp}$, then $0 = \langle f, k(x,\cdot)\rangle = f(x)$ for all $x \in \mathcal{X}$, which implies that $f = 0$. □

We conclude this subchapter with two examples of RKHSs and determine their kernels.

Example 2.1.20 (Sobolev kernel)**.** Consider the space $\mathrm{H}_0^1(0,1) := \{f \in \mathrm{H}^1(0,1) : f(0) = 0\}$ equipped with the inner product $\langle f, g\rangle_{\mathrm{H}_0^1} := \langle f', g'\rangle_{\mathrm{L}^2}$. It is a standard result that $(\mathrm{H}_0^1(0,1), \langle\cdot,\cdot\rangle_{\mathrm{H}_0^1})$ is a Hilbert space. To show that $(\mathrm{H}_0^1(0,1), \langle\cdot,\cdot\rangle_{\mathrm{H}_0^1})$ is a RKHS, we show that point evaluations are bounded. Let $f \in \mathrm{H}_0^1(0,1)$ and $x \in [0,1]$, we have

$$|f(x)| = \left|\int_0^x f'(t)\mathrm{d}t\right| \leq \int_0^1 |f'(t)\mathbf{1}_{[0,x]}(t)|\mathrm{d}t \leq \left(\int_0^1 |f'(t)|^2\mathrm{d}t\right)^{1/2}\left(\int_0^1 \mathbf{1}_{[0,x]}(t)\mathrm{d}t\right)^{1/2} = \|f\|\sqrt{x}.$$

We can read off the derivative of the kernel,

$$\langle f, k(x,\cdot)\rangle_{\mathrm{H}^1} = f(x) = \int_0^x f'(t)\mathrm{d}t = \int_0^1 f'(t)\mathbf{1}_{[0,x]}(t)\mathrm{d}t.$$

We see that $k'(x,\cdot) = \mathbf{1}_{[0,x]}(\cdot)$, we evaluate

$$k(x,y) = \langle k(x,\cdot), k(y,\cdot)\rangle_{\mathrm{H}^1} = \int_0^1 \mathbf{1}_{[0,x]}(t)\mathbf{1}_{[0,y]}(t)\mathrm{d}t = \int_0^{\min(x,y)} \mathrm{d}t = \min(x,y)$$

Since we can write $k(x,y) = \langle \mathbf{1}_{[0,x]}, \mathbf{1}_{[0,y]}\rangle_{\mathrm{L}^2}$, we see that k is conjugate symmetric and PSD.

Another interesting example of a RKHS is the Hardy space on the unit disc.

Definition 2.1.21 (Hardy space)**.** Analytic functions on the unit disk $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ where

$$\sup_{0\le r<1}\left(\frac{1}{2\pi}\int_0^{2\pi} |f(r\mathrm{e}^{i\theta})|^2 \mathrm{d}\theta\right)^{1/2} < \infty \tag{2.2}$$

are called the *Hardy space*, denoted $H^2(\mathbb{D})$. The expression in Equation 2.2 is called the Hardy space norm, denoted $\|f\|_{H^2(\mathbb{D})}$.

We will show that $H^2(\mathbb{D})$ is a RKHS in two steps, first that $H^2(\mathbb{D})$ can be identified with a Hilbert space and secondly that point evaluations are bounded.

Example 2.1.22 ($H^2(\mathbb{D})$ is a RKHS)**.** Let $f \in H^2(\mathbb{D})$, which can be written as $f = \sum_{n=0}^\infty \alpha_n z^n$ for $z \in \mathbb{D}$ and $(\alpha_n)_{n=0}^\infty$ in $\mathbb{C}$ as f is analytic. From the Weierstrass M test and the ratio test it follows that the series $f(z) = \sum_{n=0}^\infty \alpha_n z^n$ converges absolutely and uniformly on $\mathbb{D}$. Any $z \in \mathbb{D}$ can be expressed as $r\mathrm{e}^{i\theta}$ for some $\theta \in [0, 2\pi)$ and $0 \le r < 1$. Substituting results in $f(r\mathrm{e}^{i\theta}) = \sum_{n=0}^\infty \alpha_n r^n \mathrm{e}^{in\theta}$, a Fourier series with only non-negative coefficients $\alpha_n r^n$. By Parseval's identity, we have

$$\frac{1}{2\pi}\int_0^{2\pi} |f(r\mathrm{e}^{i\theta})|^2\mathrm{d}\theta = \sum_{n=0}^\infty |\alpha_n|^2 r^{2n}.$$

Consider the sequence of functions $g_r(n) = |\alpha_n|^2 r^{2n}$ on $\mathbb{N}_0 := \mathbb{N}\cup\{0\}$. Observe that $\sum_{n=0}^\infty |\alpha_n|^2 r^{2n}$ is the same as integrating $g_r(n)$ on $\mathbb{N}_0$ with respect to the counting measure. Since $|\alpha_n|^2 r^{2n} \uparrow |\alpha_n|^2$ as $r \uparrow 1$, it follows by the Monotone Convergence Theorem that

$$\infty > \sup_{0\le r<1}\left(\frac{1}{2\pi}\int_0^{2\pi} |f(r\mathrm{e}^{i\theta})|^2\mathrm{d}\theta\right)^{1/2} = \sup_{0\le r<1}\sum_{n=0}^\infty |\alpha_n|^2 r^{2n} = \lim_{r\uparrow 1}\sum_{n=0}^\infty |\alpha_n|^2 r^{2n} = \sum_{n=0}^\infty |\alpha_n|^2.$$

Define the map $\phi : H^2(\mathbb{D}) \to \ell^2(\mathbb{N}_0)$ where $\phi(f) = (\alpha_n)_{n\in\mathbb{N}_0}$ for $f(z) = \sum_{n=0}^\infty \alpha_n z^n$ with $z \in \mathbb{D}$. Any $\ell^2(\mathbb{N}_0)$ sequences defines an $H^2(\mathbb{D})$ function, and vice versa. This shows that $H^2(\mathbb{D})$ is a Hilbert space with inner product

$$\langle f, g\rangle_{H^2(\mathbb{D})} := \langle \phi(f), \phi(g)\rangle_{\ell^2(\mathbb{N}_0)}$$

for $f, g \in H^2(\mathbb{D})$ and that ϕ is an isometric isomorphism.
Evaluating f at $d \in \mathbb{D}$ can be bounded as

$$|\delta_d(f)| = |\sum_{n=0}^\infty \alpha_n d^n| \le \sum_{n=0}^\infty |\alpha_n|d||^n \le \left(\sum_{n=0}^\infty |\alpha_n|^2\right)^{1/2}\left(\sum_{n=0}^\infty |d|^{2n}\right)^{1/2} = \|f\|_{H^2(\mathbb{D})}\frac{1}{\sqrt{1-|d|^2}}.$$

Since $H^2(\mathbb{D})$ is a vector space of functions, all requirements of Definition 2.1.2, we conclude that $H^2(\mathbb{D})$ is a RKHS. Theorem 2.1.17 guarentees the existence of a reproducing kernel in $H^2(\mathbb{D})$. By working backwards, it is not difficult to read off the kernel. Let $w \in \mathbb{D}$, the kernel $k(w,\cdot)$ can be written as $k(w,z) = \sum_{n=0}^\infty b_n z^n$. Evaluate f at w, we get

$$f(w) = \sum_{n=0}^\infty \alpha_n w^n = \langle f, k(w,\cdot)\rangle_{H^2(\mathbb{D})} = \langle \phi(f), \phi(k(w,\cdot))\rangle_{\ell^2(\mathbb{N}_0)} = \sum_{n=0}^\infty \alpha_n \overline{b_n}.$$

Hence, $b_n = \overline{w^n}$ for $n \in \mathbb{N}_0$ and $k(w,z) = \sum_{n=0}^\infty \overline{w^n} z^n$. The kernel function for the Hardy space $H^2(\mathbb{D})$ is called the *Szëgo kernel*.

2.1.3 Construction of a RKHS using a kernel

Suppose we are given a RKHS H, the existence of a reproducing kernel that is admitted by H is guaranteed, as we have seen in Theorem 2.1.17. The converse path holds as well, as first shown by Aronszajn [Aro50]. Given a kernel k, there exists a Hilbert space of functions where k satisfies the conditions of a reproducing kernel. In the proof of the next theorem, the existence is not only guaranteed; the construction is made explicit. A priori, it is not clear how such a RKHS can be defined; Proposition 2.1.19 is, however, a helpful starting point.

Theorem 2.1.23 (RKHS from a kernel)**.** Let k be a kernel, then there exists a RKHS with reproducing kernel k.

Proof. Define $H_0 := \operatorname{span}\{k(x,\cdot)\,|x \in \mathcal{X}\}$ and the bivariate map

$$\langle\cdot,\cdot\rangle_{H_0} : H_0 \times H_0 \to \mathbb{C},$$
$$(f,g) \mapsto \langle f,g\rangle_{H_0} := \sum_{j=1}^{n}\sum_{i=1}^{k} \mathbf{a}_j\overline{\mathbf{b}}_i k(\mathbf{x}_j,\mathbf{y}_i)$$

for $f(\cdot) = \sum_{j=1}^{n} \mathbf{a}_j k(\mathbf{x}_j,\cdot)$, $g(\cdot) = \sum_{i=1}^{k} \mathbf{b}_i k(\mathbf{y}_i,\cdot) \in H_0$. Then, the pair $(H_0, \langle\cdot,\cdot\rangle_{H_0})$ is an inner product space. Let $f \in H_0$ and $y \in \mathcal{X}$. By construction we have

$$f(y) = \sum_{j=1}^{n} \mathbf{a}_j k(\mathbf{x}_j, y) = \langle f, k(y,\cdot)\rangle_{H_0}.$$

Under the standard pointwise addition and scalar multiplication, H_0 is a vector space. The linearity of $\langle\cdot,\cdot\rangle_{H_0}$ is straightforward, the symmetry follows from the commutativity of scalar multiplication and the fact that k is symmetric. Since k is a kernel, we have $\langle f,f\rangle_{H_0} \geq 0$. Suppose that $\langle f,f\rangle_{H_0} = 0$ and $x \in \mathcal{X}$. From Cauchy-Schwarz for linear symmetric nonnegative maps $|f(x)|^2 = |\langle f, k(x,\cdot)\rangle_{H_0}|^2 \leq \langle f,f\rangle_{H_0}\langle k(x,\cdot), k(x,\cdot)\rangle_{H_0} = \langle f,f\rangle_{H_0}\|k(x,\cdot)\|^2_{H_0}$, which implies that f is zero on K. We conclude that the map $\langle\cdot,\cdot\rangle_{H_0}$ is an inner product and therefore $(H_0, \langle\cdot,\cdot\rangle_{H_0})$ is an inner product space.

What is left to do is to complete the space. Before doing so, we will show that any Cauchy sequence in H_0 such that $f_n(x) \to 0$ satisfies $\|f_n\|_{H_0} \to 0$, as $n \to \infty$. Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in H_0 bounded from above by $M > 0$. Let $\varepsilon > 0$ and $N \in \mathbb{N}$ be such that for $n > N$ we have $\|f_n - f_N\|_{H_0} < \varepsilon/2M$. Since $f_N \in H_0$, one can write $f_N = \sum_{j=1}^{p} \mathbf{c}_j k(\mathbf{x}_j,\cdot)$ for some $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^n$. For $n > N$, we get

$$\|f_n\|^2_{H_0} = |\langle f_n - f_N, f_n\rangle_{H_0}| + |\langle f_N, f_n\rangle_{H_0}|,$$
$$\leq \varepsilon/2 + |\sum_{j=1}^{p} \mathbf{c}_j f_n(\mathbf{x}_j)|.$$

Since $f_n(\mathbf{x}_j)$ converges to zero for each $\mathbf{x}_j$, there exists $L \in \mathbb{N}$ such that for all $n > L$ we have $|\sum_{j=1}^{p} \mathbf{c}_j f_n(\mathbf{x}_j)| < \varepsilon/2$. For $n > \max\{N, L\}$ we get that $\|f_n\|^2_{H_0} \leq \varepsilon$ and we are done.

Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in H_0. For each $x \in \mathcal{X}$, the sequence $(f_n(x))_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{C}$, and therefore has a pointwise limit $f(x) := \lim_{n\to\infty} f_n(x)$. Let H be the set of functions which are pointwise limits of Cauchy sequences in H_0. Note that $H_0 \subset H$, since for each $f \in H_0$ the sequence $(f_n)_{n\in\mathbb{N}} = (f, f, \ldots)$ converges to f pointwise. Next, we will define an inner product on H so that the reproducing property is satisfied and that each Cauchy sequence converges.

Let $f, g \in H$, which are pointwise limits of Cauchy sequences $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$, respectively. Define

$$\langle f,g\rangle_H := \lim_{n\to\infty} \langle f_n, g_n\rangle_{H_0}.$$

To show that this definition is reasonable, we need to first of all show that $(\langle f_n, g_n\rangle_{H_0})_{n\in\mathbb{N}}$ converges and that $\langle f,g\rangle_H$ is independent on the Cauchy sequences that converge to f and g.

To address the first issue, using Cauchy-Schwarz and the fact that Cauchy sequences are bounded,

$$\begin{aligned}|\langle f_n, g_n\rangle_{H_0} - \langle f_m, g_m\rangle_{H_0}| &= |\langle f_n, g_n - g_m\rangle_{H_0} + \langle f_n - f_m, g_n\rangle_{H_0}|,\\ &\leq \|f_n\|_{H_0}\|g_n - g_m\| + \|f_n - f_m\|_{H_0}\|g_n\|_{H_0} \to 0 \quad (n\to\infty).\end{aligned}$$

This shows that $(\langle f_n, g_n\rangle_{H_0})_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{C}$ and therefore has a limit. To address the second issue, let $(f'_n)_{n\in\mathbb{N}}$ and $(g'_n)_{n\in\mathbb{N}}$ be Cauchy sequences, different from $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$, that converge to f and g, respectively. The sequences $(f_n - f'_n)_{n\in\mathbb{N}}$ and $(g_n - g'_n)_{n\in\mathbb{N}}$ converge pointwise to zero, hence $\|f_n - f'_n\|_{H_0} \to 0$ and $\|g_n - g'_n\|_{H_0} \to 0$, as shown above. Taking the absolute difference, we get

$$\begin{aligned}|\langle f_n, g_n\rangle_{H_0} - \langle f'_n, g'_n\rangle_{H_0}| &= |\langle f_n, g_n - g'_n\rangle_{H_0} + \langle f_n - f'_n, g'_n\rangle_{H_0}|,\\ &\leq \|f_n\|_{H_0}\|g_n - g'_n\| + \|f_n - f'_n\|_{H_0}\|g'_n\|_{H_0} \to 0 \quad (n\to\infty).\end{aligned}$$

Next, we will show that $\langle\cdot,\cdot\rangle_H$ is an inner product. The symmetry, linearity and nonnegativity follows readily from $\langle\cdot,\cdot\rangle_{H_0}$. Let $f \in H$ such that $\langle f, f\rangle_H = 0 = \|f\|_H^2 = \lim_{n\to\infty}\|f_n\|_{H_0}$, for a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ in H_0 converging to f. Then, for any $x \in \mathcal{X}$

$$|f(x)| = \lim_{n\to\infty}|f_n(x)| = \lim_{n\to\infty}|\langle f_n, k(x,\cdot)\rangle_{H_0}| \leq \lim_{n\to\infty}\|f_n\|_{H_0}\|k(x,\cdot)\|_{H_0} = 0,$$

which shows that f is the zero function. We conclude that $(H, \langle\cdot,\cdot\rangle_H)$ is an inner product space. From Proposition 2.1.19, it follows that H_0 is dense in H. Let $(h_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in H, then for each $k \in \mathbb{N}$, there exists a sequence $(g_{km})_{m\in\mathbb{N}}$ such that $\lim_{m\to\infty}\|h_k - g_{km}\|_H = 0$. In other words, for any $k \in \mathbb{N}$, there exists an $N(k) \in \mathbb{N}$ such that for all $m \geq N(k)$ we have $\|h_k - g_{km}\| < 1/k$. Define $(h'_k)_{k\in\mathbb{N}} := (g_{kN(k)})_{k\in\mathbb{N}}$, then $\lim_{k\to\infty}\|h_k - h'_k\|_H = 0$. Let $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that for $k, l > N$ we have $\|h_k - h_l\| < \varepsilon/3$ and $\|h_l - h'_l\| < \varepsilon/3$. Since $\|\cdot\|_{H_0}$ coincides with $\|\cdot\|_H$ on H_0, we get that

$$\|h'_k - h'_l\|_{H_0} = \|h'_k - h'_l\|_H \leq \|h'_k - h_k\|_H + \|h_k - h_l\|_H + \|h_l - h'_l\|_H < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.$$

We have shown that $(h'_k)_{k\in\mathbb{N}}$ is Cauchy in H_0 and therefore $\lim_{k\to\infty} h'_k := h$, the pointwise limit, exists and is an element of H. Since $h \in H$ and $(h'_k)_{k\in\mathbb{N}}$ is Cauchy in H_0 converging pointwise to h, we get

$$\lim_{k\to\infty}\|h - h'_k\|_H = \lim_{n\to\infty}\lim_{k\to\infty}\|h'_n - h'_k\|_{H_0} = 0.$$

Therefore

$$\|h - h_k\|_H \leq \|h - h'_k\|_H + \|h'_k - h_k\|_H \to 0.$$

We have shown that any Cauchy sequence $(h_n)_{n\in\mathbb{N}}$ in H has a limit, therefore $(H, \langle\cdot,\cdot\rangle_H)$ is complete. A complete Hilbert space of functions that contains a reproducing kernel is a RKHS, by Theorem 2.1.17. □

In case we have multiple kernels, k_1, k_2, we will denote the corresponding RKHSs with H_{k_i} or H_i and their norms as $\|\cdot\|_{H_{k_i}}$, or $\|\cdot\|_i$, for $i = 1, 2$.

Theorem 2.1.24 (Uniqueness of a RKHS)**.** The kernel determines a RKHS uniquely.

Proof. Suppose that G and H are RKHSs that both admit a kernel k. We have that $k(x,\cdot) \in H$ for all $x \in \mathcal{X}$. As H is a Hilbert space, it is complete and closed under addition and scalar multiplication. From Proposition 2.1.19, we conclude that H is a closed linear subspace of G. By the decomposition theorem, we can write $G = H \oplus H^\perp$. Take $g \in H^\perp$, $x \in \mathcal{X}$ and since $g(x,\cdot) \in H$ we get $0 = \langle g, k(x,\cdot)\rangle_H = g(x)$. Since x was arbitrary, it follows that $g \equiv 0$. We conclude that $H^\perp = \{0\}$ and that $G = H$. □

Combining Theorems 2.1.23 and 2.1.24 leads to the following Corollary.

Corollary 2.1.25. Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a kernel, then there exists a unique RKHS with reproducing kernel k.

We denote H_k as the RKHS with reproducing kernel k. If $\mathcal{X}$ is a topological space, we naturally equip the product $\mathcal{X} \times \mathcal{X}$ with the product topology. From Proposition 2.1.19, it is not surprising that H_k is contained in the continuous functions if k is continuous on $\mathcal{X} \times \mathcal{X}$, denoted $\mathrm{C}(\mathcal{X})$.

Theorem 2.1.26. Let $\mathcal{X}$ be a topological space and k a continuous kernel function. Then $H(k)$ is contained in the continuous functions on $\mathcal{X}$.

Proof. Fix $y \in \mathcal{X}$, $\varepsilon > 0$ and let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ converging to y. By the reproducing property and Cauchy-Schwarz $|f(y_n) - f(y)| = \langle f, k(y_n,\cdot) - k(y,\cdot)\rangle \leq \|f\|\|k(y_n,\cdot) - k(y,\cdot)\|$. Showing that $\lim_{n\to\infty}\|k(y_n,\cdot) - k(y,\cdot)\| = 0$ completes the proof. By continuity of k on $\mathcal{X} \times \mathcal{X}$, we also have that $k(y,\cdot)$ and $k(\cdot,y)$ are continuous on $\mathcal{X}$. By the reproducing property and the triangle inequality

$$\begin{aligned}
\|k(y_n,\cdot) - k(y,\cdot)\|^2 &= \langle k(y_n,\cdot) - k(y,\cdot), k(y_n,\cdot) - k(y,\cdot)\rangle, \\
&= |k(y_n,y_n) - k(y_n,y) - k(y,y_n) + k(y,y)|, \\
&= |(k(y_n,y_n) - k(y,y)) - (k(y_n,y) - k(y,y)) - (k(y,y_n) - k(y,y))|, \\
&\leq |k(y_n,y_n) - k(y,y)| + |k(y_n,y) - k(y,y)| + |k(y,y_n) - k(y,y)| \to 0 \quad (n \to \infty).
\end{aligned}$$

□

2.2 Characterizations

In Example 2.1.14, we constructed a feature map from a given kernel. It turns out that this construction is always possible. In this section we build up the necessary theory to prove Mercer's Theorem [Mer09], which directly implies two results. The first is a spectral definition of a RKHS, the other being the so called *kernel trick*, a backbone in machine learning first discovered by Schölkopf [Sch00].
Thereafter, we will characterize when, a Hilbert space of functions is a RKHS, a function is an element of a RKHS, the difference of kernels is again a kernel and construct the RKHS of a sum of two kernels.
We will state an assumption that is necessary for H_k to be separable. Whenever necessary, we shall freely assume this.

Assumption 1 (Separability)**.** To ensure that the RKHS H_k is separable, we need to make an additional assumption on the underlying space $\mathcal{X}$ and the kernel k. It turns out that H_k is separable if $\mathcal{X}$ is a compact topological space and k is continuous [SC08, Lemma 4.33].

2.2.1 Mercer's theorem and the kernel trick

We assume that $\mathcal{X}$ is a compact metric space and that the kernel function k is continuous. It follows that $C_k = \sup_{x,t\in\mathcal{X}}|k(x,t)| < \infty$. First we will define an integral operator from a space of square integrable functions into the continuous functions. Hereafter, we show that this map is well-defined.

Definition 2.2.1 (Integral operator)**.** Let k be a continuous kernel on a compact metric space $\mathcal{X}$. Let ω be a finite measure on $\mathcal{X}$ and let $\mathrm{L}^2(\mathcal{X},\omega)$ be the space of square integrable functions on $\mathcal{X}$ with respect to the measure ω. Let B_k be the bounded linear map

$$\begin{aligned}
B_k : \mathrm{L}^2(\mathcal{X},\omega) &\to \mathrm{C}(\mathcal{X}), \\
f &\mapsto (B_k f)(\cdot) := \int_{\mathcal{X}} k(x,\cdot) f(x) \mathrm{d}\omega(x).
\end{aligned}$$

The *integral operator* with kernel k is the bounded linear map given by $A_k := I_k \circ B_k$, where $I_k : \mathrm{C}(\mathcal{X}) \hookrightarrow \mathrm{L}^2(\mathcal{X},\omega)$.

Definition 2.2.2 (Hilbert-Schmidt operator)**.** Let H, K be Hilbert spaces. A bounded linear operator $D : H \to K$ is called a *Hilbert-Schmidt operator* if

$$\sum_j \|De_j\|^2 < \infty,$$

for some orthonormal basis $(e_j)_j$ of H.

Proposition 2.2.3 (Hilbert-Schmidt operators are compact)**.** Every abstract Hilbert-Schmidt operator is compact.

Proof. Let $D : H \to K$ be a Hilbert-Schmidt operator acting between Hilbert spaces. As D acts between Hilbert spaces, it is equivalent to show that there exists a sequence of finite rank operators $(D_n)_{n\in\mathbb{N}}$ such that $\|D - D_n\|_{\mathcal{L}(H,K)} \to 0$ as $n \to \infty$. Let $f \in H$, define $D_n f := \sum_{i=1}^{n} \langle f, e_i\rangle De_i$, we get

$$\|Df - D_n f\|^2 = \| \sum_{i=n+1}^{\infty} \langle f, e_i\rangle De_i\|^2 \le \|f\|^2 \sum_{i=n+1}^{\infty} \|De_j\|^2 < \infty,$$

and we have $\|D - D_n\|_{\mathcal{L}(H,K)} \le \sum_{i=n+1}^{\infty} \|De_j\|^2 \to 0$ as $n \to \infty$. □

To show that this is a meaningful definition, we need to show that B_k and I_k are bounded maps into their respective spaces. It turns out that A_k is compact and we can derive an upper bound on its operator norm.

Proposition 2.2.4. Suppose k is continuous, then A_k is well defined and compact self-adjoint Hilbert-Schmidt operator. Furthermore, $\|A_k\| \le \sqrt{\omega(\mathcal{X})} C_k$.

Proof. Let $f \in \mathrm{C}(\mathcal{X})$, then

$$\|I_k(f)\|_{\mathrm{L}^2_\omega} = \left(\int_{\mathcal{X}} f(x)^2 \mathrm{d}\omega(x) \right)^{1/2} \le \|f\|_\infty \sqrt{\omega(\mathcal{X})}.$$

Since k is continuous and $\mathcal{X}$ is compact metric space, it follows that k is uniformly continuous. Let $y_1, y_2 \in \mathcal{X}$ and $f \in \mathrm{L}^2(\mathcal{X}, \omega)$, then

$$\begin{aligned} |B_k f(y_1) - B_k f(y_2)| &= \left| \int_{\mathcal{X}} f(x)(k(x, y_1) - k(x, y_2)) \mathrm{d}\omega(x) \right|, \\ &\overset{\text{C.S.}}{\le} \|f\|_{\mathrm{L}^2_\omega} \|k(\cdot, y_1) - k(\cdot, y_2)\|_{\mathrm{L}^2_\omega}, \\ &\le \|f\|_{\mathrm{L}^2_\omega} \sqrt{\omega(\mathcal{X})} \max_{t\in\mathcal{X}} |k(t, y_1) - k(t, y_2)|. \end{aligned}$$

It follows from the continuity of $k(x, \cdot)$ that $B_k f$ is continuous. Let $\varepsilon > 0$ and choose $\delta_x > 0$ such that $|y_1 - y_2| < \delta_x$ implies that $\max_{x\in\mathcal{X}} |k(x, y_1) - k(x, y_2)| < \varepsilon / \|f\|_{\mathrm{L}^2_\omega} \sqrt{\omega(\mathcal{X})}$. For any $y \in \mathcal{X}$ and $f \in \mathrm{L}^2(\mathcal{X}, \omega)$,

$$|A_k f(y)| \overset{\text{C.S.}}{\le} \|f\| \|k(\cdot, y)\| = \|f\| \left(\int_{\mathcal{X}} k(x, y) \mathrm{d}\omega(x) \right) \le \|f\| \sqrt{\omega(\mathcal{X})} \sup_{x\in\mathcal{X}} |k(x, y)|.$$

By definition of the operator norm, we have the inequality $\|A_k\|_{\mathrm{L}^2_\omega} \le \sqrt{\omega(\mathcal{X})} C_k$.
The operator A_k is a self adjoint Hilbert-Schmidt operator, since k is symmetric and $\|A_k\|_{\mathcal{L}} \le \|A_k\|_{\mathrm{HS}} = \|k\|_{\mathrm{L}^2_\omega} \le \sqrt{\omega(\mathcal{X})} C_k$. It follows from Proposition 2.2.3 that A_k is compact. □

We have shown that for the integral operator A_k, the spectral Theorem applies. This is an important step in the proof of Mercer's theorem.

Theorem 2.2.5 (Mercer [Mer09])**.** Let k be a continuous kernel on a compact metric space $\mathcal{X}$ and let ω be a finite Borel measure on $\mathcal{X}$. Let $(\phi_k, \lambda_k)_k$ be the eigen-pairs of the operator A_k. The kernel can be written as

$$k(x,t) = \sum_{k=1}^{\infty} \lambda_k (\phi_k \otimes \phi_k)(x,t) \quad \forall x, t \in \mathcal{X}. \tag{2.3}$$

The convergence in (2.3) is absolute and uniform.

The eigen-pairs of the Hilbert-Schmidt operator A_k can be used to define the RKHS defined by the kernel k.

Corollary 2.2.6 (Spectral RKHS construction)**.** Under the same notation, we define a RKHS with kernel k as the vector space

$$H_k(\mathcal{X}) := \left\{ f \in \mathrm{L}^2(\mathcal{X}, \omega) \mid f = \sum_{j=1}^{\infty} \alpha_j \phi_j, \ \left(\frac{\alpha_j^2}{\lambda_j} \right)_{j \in \mathbb{N}} \in \ell^2(\mathbb{N}) \right\}, \tag{2.4}$$

equipped with the inner product

$$\langle f, g \rangle_{H_k} := \sum_{j=1}^{\infty} \frac{\langle f, \phi_i \rangle \langle g, \phi_i \rangle}{\lambda_j} = \sum_{j=1}^{\infty} \frac{\alpha_j \beta_j}{\lambda_j},$$

where $f = \sum_{j=1}^{\infty} \alpha_j \phi_j$ and $g = \sum_{k=1}^{\infty} \beta_k \phi_k$.

Proof. As done in [Wai19]; [Sai16]. □

It is in fact a general fact that any kernel can be written as an inner product

Proposition 2.2.7. Let $\mathcal{X}$ be a compact topological space and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a continuous map. Then k is a kernel if and only if there exists a feature map Φ such that $k(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\ell^2(\mathbb{N})}$.

Proof. Suppose that k is a kernel then it follows from Corollary 2.1.25 that there exists a unique RKHS H_k such that k has the reproducing property. Define $\tilde{\Phi} : \mathcal{X} \to \ell^2(\mathbb{N})$ by $\tilde{\Phi}(x) := k(x, \cdot)$. Since H_k is a seperable Hilbert space (Assumption 1), it is isometrically isomorphic to $\ell^2(\mathbb{N})$ under some map ξ. Then $\Phi := \xi \circ \tilde{\Phi}$ suffices, since

$$k(x,y) = \langle k(x,\cdot), k(y,\cdot) \rangle_{H_k} = \langle \tilde{\Phi}(x), \tilde{\Phi}(y) \rangle_{H_k} = \langle (\xi \circ \tilde{\Phi})(x), (\xi \circ \tilde{\Phi})(y) \rangle_{\ell^2(\mathbb{N})} = \langle \Phi(x), \Phi(y) \rangle_{\ell^2(\mathbb{N})}.$$

Conversely, the map k defined by $k(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\ell^2(\mathbb{N})}$ is a kernel since for $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{C}^n$ we have

$$\sum_{i,j=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_j) \overline{\mathbf{a}_j} = \langle \sum_{i=1}^{n} \mathbf{a}_i \Phi(\mathbf{x}_i), \sum_{i=1}^{n} \mathbf{a}_i \Phi(\mathbf{x}_i) \rangle_{\ell^2(\mathbb{N})} = \| \sum_{i=1}^{n} \mathbf{a}_i \Phi(\mathbf{x}_i) \|_{\ell^2(\mathbb{N})}^2 \geq 0.$$

□

It need not be the case that the feature map Φ maps into $\ell^2(\mathbb{N})$, it may also be some finite dimensional subspace thereof. In Example 2.1.14, the feature space was $\mathbb{R}^D$, which is a finite dimensional subspace of $\ell^2(\mathbb{N})$. One often calls this, finite dimensional, subspace F the feature space. From any feature map, one can define a kernel function canonically $k(x,y) := \langle \Phi(x), \Phi(y) \rangle_{\ell^2}$ and the converse path also holds, as we have seen in Proposition 2.2.7. Mercer's theorem allows us to construct the feature map explicitly

$$\Phi : \mathcal{X} \to \ell^2(\mathbb{N}), \tag{2.5}$$

$$x \mapsto (\sqrt{\lambda_1}\phi_1(x), \sqrt{\lambda_2}\phi_2(x), \ldots). \tag{2.6}$$

By definition we have

$$\|\Phi(x)\|_{\ell^2}^2 = \sum_{j=1}^{\infty} \lambda_j \phi_i(x)^2 = k(x,x) < \infty.$$

Moreover, the feature map provides an easy way to compute the inner product of the representation of $x, z \in \mathcal{X}$ in the feature space, i.e. ℓ^2, as follows

$$\langle \Phi(x), \Phi(z) \rangle_{\ell^2} = \sum_{j=1}^{\infty} \lambda_j \phi_i(x) \phi_i(z) = k(x,z)$$

so inner products boil down to kernel evaluations, we rediscovered the kernel trick.

2.2.2 RKHS and its elements

The Theorems presented in this chapter rely on the Interpolation Theorem A.1.1. To ensure we can apply this theorem, we make the following assumption throughout the remainder of this report.

Assumption 2 (Independence of kernels)**.** Assume that the set $\{k(x,\cdot) : x \in \mathcal{X}\}$ is linearly independent.

It turns out that this Assumption 2 is satisfied if and only if the kernel is strictly positive, meaning that there is a strict inequality in Equation (2.1).

Lemma 2.2.8. Let $\mathcal{X}$ be a set and let k be a kernel on $\mathcal{X}$. The set $\{k(x,\cdot) : x \in \mathcal{X}\}$ is linearly independent if and only if k is a positive kernel.

Proof. We will assume k is not a positive kernel and show that the set $\{k(x,\cdot) : x \in \mathcal{X}\}$ is linearly dependent. By assunption, there exists $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{R}^n$ such that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i \mathbf{a}_j k(\mathbf{x}_i, \mathbf{x}_j) = 0.$$

Using the reproducing property

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i \mathbf{a}_j k(\mathbf{x}_i, \mathbf{x}_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i \mathbf{a}_j \langle k(\mathbf{x}_i, \cdot), k(\mathbf{x}_j, \cdot) \rangle_{H_k} = \|\sum_{i=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \cdot)\|_{H_k}^2. \tag{2.7}$$

Thus $\sum_{i=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \cdot) = 0$, by definiteness of the norm. This shows that $\{k(x,\cdot) : x \in E\}$ is linearly dependent. By contraposition, we have shown that linear independence implies K to be positive definite. On the other hand, assume that $\{k(x,\cdot) : x \in E\}$ is linearly dependent, then there exists $x \in \mathcal{X}^n$ and $0 \neq \mathbf{a} \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \cdot) = 0$. By definiteness of the norm

$$0 = \|\sum_{i=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \cdot)\|_{H_k}^2 = \langle \sum_{i=1}^{n} \mathbf{a}_i k(\mathbf{x}_i, \cdot), \sum_{j=1}^{n} \mathbf{a}_j k(\mathbf{x}_j, \cdot) \rangle_{H_k} = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i \mathbf{a}_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{2.8}$$

which shows that k is not a positive definite kernel. We have shown that k is a positive kernel implies linear independence, by contraposition. □

Theorem 2.2.9 (Theorem 3.11 in [PR16])**.** Let H be a RKHS on $\mathcal{X}$ with positive kernel k and let $f : \mathcal{X} \to \mathbb{C}$ be a function. Then, $f \in H$ if and only if there exists a constant $c \geq 0$ such that $(x,y) \mapsto c^2 k(x,y) - f(x)\overline{f(y)}$ is a kernel function. Moreover, $\|f\|$ is the smallest possible value for c.

Proof. Suppose $f \in H$ and let $\|f\| \leq c$, by the reproducing property $\langle f, k(x,\cdot)\rangle = f(x)$. By the Interpolation Theorem A.1.1, for all $n \in \mathbb{N}$ there exists $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{b} \in \mathbb{C}^n$ such that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i\overline{\mathbf{b}_j} f(\mathbf{x}_i)\overline{f(\mathbf{y}_j)} \leq c^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i\overline{\mathbf{b}_j}\langle k(\mathbf{x}_i,\cdot), k(\mathbf{y}_j,\cdot)\rangle = c^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i\overline{\mathbf{b}_j} k(\mathbf{x}_i,\mathbf{y}_j).$$

It follows that $c^2 k(x,y) - f(x)\overline{f(y)}$ is a kernel function, with least possible constant $c = \|f\|$. Conversely, suppose that $c^2 k(x,y) - f(x)\overline{f(y)}$ is a kernel function. We conclude by A.1.1 that there exists $f \in H$ with $\|f\| \leq c$ such that $\|f\| \leq c$ and $\langle f, k(x,\cdot)\rangle = f(x)$ for any $x \in X$. ☐

A different way to write that $k(x,y) - f(x)\overline{f(y)}$ is a kernel function is to say that $k - f\overline{f} \succeq 0$ or alternatively $f\overline{f} \preceq k$. With this, we mean that for all $n \in \mathbb{N}$ and every $x \in \mathcal{X}^n$ and $\mathbf{a} \in \mathbb{C}^n$ we have

$$\sum_{i,j=1}^{n} \mathbf{a}_i \left(k(\mathbf{x}_i,\mathbf{x}_j) - f(\mathbf{x}_i)\overline{f(\mathbf{x}_j)}\right)\overline{\mathbf{a}}_j \geq 0$$

2.2.3 Sums and differences of kernels

In Observation 2.1.8, we saw that the sum and product of kernels are again kernels. It turns out that when the difference of kernels is again be a kernel, can be characterized by the inclusion of two RKHSs.

Definition 2.2.10 (Contractively contained)**.** Let $(H_1, \|\cdot\|_1)$ and $(H_2, \|\cdot\|_2)$ be Hilbert spaces. We say that H_1 is *contractively contained* in H_2 if H_1 is a subspace of H_2 and $\|f\|_2 \leq \|f\|_1$ for all $f \in H_1$.

Theorem 2.2.11 (Aronszajn's inclusion theorem)**.** Let k_1 and k_2 be kernels on a set $\mathcal{X}$. Then $H_{k_1} \subseteq H_{k_2}$ if and only if there exists a constant $c > 0$ such that $k_1 \preceq c^2 k_2$. Additionally, $\|f\|_2 \leq c\|f\|_1$ for any $f \in H_{k_1}$.

Proof. Suppose that $H_{k_1} \subseteq H_{k_2}$. Let $i : H_{k_1} \to H_{k_2}$ denote the inclusion operator. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in H_{k_1} converging to $f \in H_{k_1}$. Since the norm $\|\cdot\|_1$ is stronger than $\|\cdot\|_2$, we have $i(f_n) = f_n \to f$ in H_{k_2}. Then $(f_n, i(f_n)) \to (f, f)$ and by the closed graph Theorem, we conclude that i is bounded. Set $c = \|i\|$, take $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{a} \in \mathbb{C}^n$, by the reproducing property,

$$0 \leq \|\sum_{i=1}^{n} \mathbf{a}_i k_1(\mathbf{x}_i,\cdot)\|_1^2 = \sum_{i,j=1}^{n} \mathbf{a}_i\overline{\mathbf{a}_j} k_1(\mathbf{x}_i,\mathbf{x}_j), \tag{2.9}$$

$$= \sum_{i,j=1}^{n} \mathbf{a}_i\overline{\mathbf{a}_j}\langle k_1(\mathbf{x}_i,\cdot), k_2(\mathbf{x}_j,\cdot)\rangle_2, \tag{2.10}$$

$$= \langle\sum_{i=1}^{n} \mathbf{a}_i k_1(\mathbf{x}_i,\cdot), \sum_{j=1}^{n} \mathbf{a}_j k_2(\mathbf{x}_j,\cdot)\rangle_2, \tag{2.11}$$

$$\overset{\text{C.S}}{\leq} \|\sum_{i=1}^{n} \mathbf{a}_i k_1(\mathbf{x}_i,\cdot)\|_2\|\sum_{j=1}^{n} \mathbf{a}_j k_2(\mathbf{x}_j,\cdot)\|_2, \tag{2.12}$$

$$\leq c\|\sum_{i=1}^{n} \mathbf{a}_i k_1(\mathbf{x}_i,\cdot)\|_1\|\sum_{j=1}^{n} \mathbf{a}_j k_2(\mathbf{x}_j,\cdot)\|_2. \tag{2.13}$$

Canceling $\|\sum_{i=1}^{n} \mathbf{a}_i k_1(\mathbf{x}_i,\cdot)\|_1$ and squaring both sides results in

$$\sum_{i,j=1}^{n} \mathbf{a}_i\overline{\mathbf{a}_j} k_1(\mathbf{x}_i,\mathbf{x}_j) \leq c^2 \sum_{i,j=1}^{n} \mathbf{a}_i\overline{\mathbf{a}_j} k_2(\mathbf{x}_i,\mathbf{x}_j).$$

This implies that $k_1 \preceq c^2 k_2$ and since i was bounded, we have $\|i(f)\|_2 \leq \|i\|\|f\|_1 = c\|f\|_1$. Conversely, suppose that there exists a $c > 0$ such that $k_1 \preceq c^2 k_2$. Let $f \in H_{k_1}$, from Theorem 2.2.9 it follows that $f\overline{f} \preceq \|f\|_1^2 k_1$. Combining these results gives $f\overline{f} \preceq \|f\|_1^2 c^2 k_2$ and we conclude that $f \in H_{k_2}$. Using the same theorem, we conclude that $f \in H_{k_2}$ with $\|f\|_2 \leq c\|f\|_1$. □

Corollary 2.2.12 (Difference of kernels)**.** Let k_1 and k_2 be kernels on a set $\mathcal{X}$. Then H_{k_1} is contractively contained in H_{k_2} if and only if $k_2 - k_1$ is a kernel.

Proof. Both implications are a direct consequence of Theorem 2.2.11. □

Theorem 2.2.13 (RKHS from a sum of kernels)**.** Let k_1 and k_2 be kernels on a set $\mathcal{X}$. The space of functions $H = H_{k_1} + H_{k_2}$ with norm

$$\|f\|_H^2 := \min\{\|f_1\|_1^2 + \|f_2\|_2^2 \,:\, f = f_1 + f_2,\ f_i \in H_{k_i},\ i = 1, 2\}, \tag{2.14}$$

is a RKHS with kernel $k_1 + k_2$.

Proof. Let $k = k_1 + k_2$ and denote the elements $f, g \in H_{k_1} \oplus H_{k_2}$ as $f = (f_1, f_2)$ with $f_i \in H_{k_i}$ and similarly for g. Then, $H_{k_1} \oplus H_{k_2}$ is a Hilbert space with inner product

$$(\langle (f_1, f_2), (g_1, g_2)\rangle_{H_{k_1}\oplus H_{k_2}}) := \langle f_1, g_1\rangle_{H_{k_1}} + \langle f_2, g_2\rangle_{H_{k_1}}$$

for $f, g \in H_{k_1} \oplus H_{k_2}$ so that $(H_{k_1} \oplus H_{k_2}, \langle\cdot,\cdot\rangle_{H_{k_1}\oplus H_{k_1}})$ is a Hilbert space. Define the subspace $N = \{(h, -h) \subseteq H_{k_1} \oplus H_{k_2} \,:\, f \in H_1 \cap H_2\}$ where we assume that $H_{k_1} \cap H_{k_2}$ is non empty. Otherwise, the decomposition theorem applies immediately and we are done. Let $(f_n, -f_n)_{n\in\mathbb{N}}$ be a sequence in $H_{k_1} \cap H_{k_2}$ converging to (f, g). Using the definition of the norm, we see that $f_n \to f$ in H_{k_1} and that $-f_n \to g$ in H_{k_2}, therefore $f = -g$ and N is closed. Hence, by the decomposition theorem, we can write $H_{k_1} \oplus H_{k_2} = N \oplus N^\perp$ so that $(f_1, f_2) = (h, -h) + (g_1, g_2)$ with $(h, -h) \in N$ and $(g_1, g_2) \in N^\perp$. Define $H := H_{k_1} + H_{k_2}$ and the map $L : H_{k_1} \oplus H_{k_2} \to H$ with $L(f_1, f_2) = f_1 + f_2$, so that L is a linear surjection and that the kernel of L is N. This implies that the restriction of L to $N^\perp$, denoted $L|_{N^\perp} := L_\perp$, is a vector space isomorphism. Equip H with the inner product induced by the map $L_\perp$, which ensures that $(H, \langle\cdot,\cdot\rangle_H$ is a Hilbert space. Thus, for $f, g \in H$,

$$\langle f, g\rangle_H := \langle L_\perp^{-1}(f), L_\perp^{-1}(g)\rangle_{H_{k_1}\oplus H_{k_2}}.$$

What is left to do is to show that the norm is as in Equation (2.14), and that k satisfies the reproducing property. Let P be the orthogonal projection from $H_{k_1} \oplus H_{k_2}$ on $N^\perp$. Let $f = f_1 + f_2 \in H$ and observe that $\|f\|_H = \|P(f_1, f_2)\|_{H_{k_1}\oplus H_{k_2}}$. Take $f \in H$ so that $f = f_1 + f_2$, then $(f - Pf) \in (N^\perp)^\perp = N$, using the definitions of the norms and Pythagoras,

$$\begin{aligned}\|f_1\|_1^2 + \|f_2\|_2^2 = \|(f_1, f_2)\|^2_{H_{k_1}\oplus H_{k_2}} &= \|(f_1, f_2) - P(f_1, f_2) + P(f_1, f_2)\|^2_{H_{k_1}\oplus H_{k_2}},\\ &= \|(f_1, f_2) - P(f_1, f_2)\|^2_{H_{k_1}\oplus H_{k_2}} + \|P(f_1, f_2)\|^2_{H_{k_1}\oplus H_{k_2}},\\ &= \|(f_1, f_2) - P(f_1, f_2)\|^2_{H_{k_1}\oplus H_{k_2}} + \|f\|_H^2.\end{aligned}$$

We get $\|f\|_H^2 \leq \|f_1\|_1^2 + \|f_2\|_2^2$ with equality if and only if $f \in N^\perp$, which shows Equation (2.14). Take $(h, -h) \in N$, $x \in \mathcal{X}$, then

$$\langle (h, -h), (k_1(x, \cdot)), k_2(x, \cdot)\rangle_{H_1\oplus H_2} = h(x) - h(x) = 0,$$

and hence $(k_1(x, \cdot), k_2(x, \cdot)) \in N^\perp$ for all $x \in \mathcal{X}$. Let $f \in H$, since $L_\perp$ is a bijection, there exists unique elements $(h_1, h_2) \in N^\perp$ such that $[L_\perp(h_1, h_2)]\,(x) = f(x)$ for any $x \in \mathcal{X}$. Using the reproducing properties of k_1 and k_2 on H_{k_1} and H_{k_2}, respectively

$$\begin{aligned}\langle f, k(x, \cdot)\rangle_H = \langle L_\perp^{-1}(f), L_\perp^{-1}(k(x, \cdot))\rangle_{H_1\oplus H_2} &= \langle (h_1, h_2), (k_1(x, \cdot), k_1(x, \cdot))\rangle_{H_1\oplus H_2},\\ &= \langle h_1, k_1(x, \cdot)\rangle_{H_{k_1}} + \langle h_2, k_2(x, \cdot)\rangle_{H_{k_2}} = h_1(x) + h_2(x) = f(x).\end{aligned}$$

We have shown that k satisfies the reproducing property. Since $(H, \langle\cdot,\cdot\rangle_H)$ is a Hilbert space of functions on $\mathcal{X}$ that admits a reproducing kernel, it is a RKHS, by Theorem 2.1.17. □

A direct consequence of Theorem 2.2.13, which is also discussed in its proof, is the case when $H_{k_1} \cap H_{k_2} = \{0\}$, in which the sum of the underlying RKHS has a simple form.

Corollary 2.2.14. Let k_1 and k_2 be kernels on a set $\mathcal{X}$ and set $k = k_1 + k_2$. Suppose that $H_{k_1} \cap H_{k_2} = \{0\}$, then $H = H_{k_1} + H_{k_2} := \{f_1 + f_2 : f_i \in H_{k_i}, i = 1, 2\}$ with $\|f\|^2_{H_k} = \|f_1\|^2_1 + \|f_2\|^2_2$ is a RKHS with reproducing kernel k.

2.2.4 Composition operators

Let k_1 and k_2 be kernels defined on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Next, we study the composition operator between the RKHSs H_{k_1} and H_{k_2} induced by a map $\varphi : \mathcal{X}_1 \to \mathcal{X}_2$. First we give a general definition. Let $\mathbb{C}^{\mathcal{X}}$ be the set of all complex-valued functions from a set $\mathcal{X}$ to $\mathbb{C}$.

Definition 2.2.15 (Composition operator)**.** Let $\mathcal{X}$, $\mathcal{S}$ be sets and $\varphi : \mathcal{S} \to \mathcal{X}$ a map. The *composition operator* $T_\varphi : \mathbb{C}^{\mathcal{X}} \to \mathbb{C}^{\mathcal{S}}$ is defined as $T_\varphi f := f \circ \varphi$, where $f \in \mathbb{C}^{\mathcal{X}}$.

Let $\mathcal{F}(\mathcal{X}) \subseteq \mathbb{C}^{\mathcal{X}}$ be a function space with the standard pointwise operations and $\varphi : \mathcal{X} \to \mathcal{X}$. We say that the composition operator *preserves the space* if $T_\varphi : \mathcal{F}(\mathcal{X}) \to \operatorname{ran}(T_\varphi) \subseteq \mathcal{F}(\mathcal{X})$.
For the remainder of this section, we will restrict ourselves to a special case where k_1 and k_2 are kernels on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively and $\varphi : \mathcal{X}_1 \to \mathcal{X}_2$. Then, $T_\varphi : H_{k_2} \to H_{k_1}$ with $T_\varphi f = f \circ \varphi$. The composition operator is linear, but is it also bounded? To characterize the boundedness, we introduce the notion of a pullback and state the pullback theorem. For semantic reasons we write $k \circ \varphi$, where we actually mean $k \circ (\varphi \times \varphi)$.

Proposition 2.2.16. Let $\varphi : \mathcal{S} \to \mathcal{X}$ be a map. If k is a kernel function on $\mathcal{X}$, then $k \circ \varphi$ is a kernel on $\mathcal{S}$

Proof. Choose $s \in \mathcal{S}^n$ and $\mathbf{a} \in \mathbb{C}^n$. We have that $\{\varphi(\mathbf{s}_1), \ldots, \varphi(\mathbf{s}_n)\} = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ with $p \leq n$. Define $I_k = \{i \in \{1, \ldots, n\} : \varphi(\mathbf{s}_i) = \mathbf{x}_k\}$ for $1 \leq k \leq p$ and $\mathbf{b}_k = \sum_{i \in I_k} \mathbf{a}_i$. Then we have

$$\sum_{i,j=1}^{n} \mathbf{a}_i \overline{\mathbf{a}_j} k(\varphi(\mathbf{s}_i), \varphi(\mathbf{s}_j)) = \sum_{k,l=1}^{p} \sum_{i \in I_k} \sum_{j \in I_l} \mathbf{a}_i \overline{\mathbf{a}_j} k(\mathbf{x}_k, \mathbf{x}_l) = \sum_{k,l}^{p} \mathbf{b}_k \overline{\mathbf{b}_l} k(\mathbf{x}_k, \mathbf{x}_l) \geq 0.$$

This shows that $k \circ \varphi$ is a kernel function on S. □

Since $k \circ \varphi$ is a kernel, it induces its own RKHS, which we call the *pullback RKHS* of $H(k)$ along φ. The RKHS $H(k)$ is "pulled back" from $\mathcal{X}$ to $\mathcal{S}$ via the map φ.

Definition 2.2.17 (Pull-back)**.** Let $\mathcal{X}$ and $\mathcal{S}$ be sets, let $\varphi : \mathcal{S} \to \mathcal{X}$ be a map and let k be a kernel on $\mathcal{X}$. We call the RKHS $H(k \circ \varphi)$ the *pullback* of $H(k)$ along φ and we call the composition operator $H(k) \to H(k \circ \varphi)$ the *pullback map.*

The elements of the pullback RKHS are characterized in the following theorem.

Theorem 2.2.18 (Pullback theorem)**.** Let $\mathcal{X}$ and $\mathcal{S}$ be sets, let k a kernel on $\mathcal{X}$ and $\varphi : \mathcal{S} \to \mathcal{X}$ be a map. Then, $H(k \circ \varphi) = \{f \circ \varphi : f \in H(k)\}$ and for $u \in H(k \circ \phi)$ we have $\|u\|_{H(k \circ \phi)} = \min_{f \in H(k)} \{\|f\|_{H(k)} : u = f \circ \varphi\}$.

Proof. Let $f \in H(k)$, from Theorem 2.2.9 we have $f(x)\overline{f(y)} \preceq \|f\|^2_{H(k)} k(x, y)$ for any $(x, y) \in \mathcal{X} \times \mathcal{X}$. Hence, for any $(s, t) \in \mathcal{S} \times \mathcal{S}$ we have $f(\varphi(s))\overline{f(\varphi(t)} \preceq \|f\|^2_{H(k)} k(\varphi(s), \varphi(t))$. Following the converse of Theorem 2.2.9 for $k \circ \varphi$, which applies by Proposition 2.2.16, we immediately see that $f \circ \varphi \in H(k \circ \varphi)$ and $\|f \circ \varphi\|_{H(k \circ \varphi)} \leq \|f\|_{H(k)}$.

Any $u \in H(k\circ\varphi)$ can be written as $f\circ\varphi$, for some $f \in H(k)$. The reproducing property is then $\langle u, k(\varphi(t), \varphi(\cdot)\rangle_{H(k\circ\varphi)} = u(t)$. Let $\mathbf{t} \in \mathcal{S}^n$ and $\mathbf{a} \in \mathbb{C}^n$, set $u = \sum_{i=1}^n \mathbf{a}_i k(\varphi(\mathbf{t}_i), \varphi(\cdot))$ and note that

$$\begin{aligned}
\|u\|^2_{H(k\circ\varphi)} &= \sum_{i,j=1}^n \mathbf{a}_i\overline{\mathbf{a}_j}\langle k(\varphi(\mathbf{t}_i), \varphi(\cdot), k(\varphi(\mathbf{t}_j), \varphi(\cdot)\rangle_{H(k\circ\varphi)},\\
&= \sum_{i,j=1}^n \mathbf{a}_i\overline{\mathbf{a}_j} k(\varphi(\mathbf{t}_i), \varphi(\mathbf{t}_j)),\\
&= \sum_{i,j=1}^n \mathbf{a}_i\overline{\mathbf{a}_j}\langle k(\varphi(\mathbf{t}_i), \cdot), k(\varphi(\mathbf{t}_j), \cdot)\rangle_{H(k)},\\
&= \|\sum_{i=1}^n \mathbf{a}_i k(\varphi(\mathbf{t}_i), \cdot)\|^2_{H(k)}.
\end{aligned}$$

The map

$$\begin{aligned}
\zeta : \operatorname{span}\{k(\varphi(t), \varphi(\cdot) \,:\, t\in\mathcal{S}\} &\to H(k),\\
\sum_{i=1}^n \mathbf{a}_i k(\varphi(\mathbf{t}_i), \varphi(\cdot)) &\mapsto \sum_{i=1}^n \mathbf{a}_i k(\varphi(\mathbf{t}_i), \cdot)
\end{aligned}$$

is an isometry that extends to $H(k\circ\varphi)$ by Proposition 2.1.19. We see that $u = \zeta(u)\circ\varphi$, with $\|u\|_{H(k\circ\varphi)} = \|\zeta(u)\|_{H(k)}$, the result follows from the fact that $\|f\circ\varphi\|_{H(k\circ\varphi)} \le \|f\|_{H(k)}$ for any $f \in H(k)$. □

Corollary 2.2.19 (Restriction theorem)**.** Let k be a kernel on a set $\mathcal{X}$ and let $\mathcal{S}$ be a non-empty subset of $\mathcal{X}$. Denote $k|_S$, the kernel k restricted to the set $\mathcal{S}$. Then $u \in H(k|_S)$ if and only if u is the restriction of a function in $H(k)$ to the set $\mathcal{S}$. Moreover,

$$\|u\|_{H(k|_S)} = \min_{f\in H(k)}\{\|f\|_{H(k)} \,:\, u = f|_S\}.$$

Proof. Let $\varphi : \mathcal{S}\to\mathcal{X}$ be the inclusion map, that is $\varphi(s) = s$. Note that $H(k\circ\varphi) = H(k|_S)$ and apply the pullback Theorem 2.2.18. Conversely, suppose that $u = f|_S = f\circ\varphi$ for $f \in H(k)$. Let $t \in \mathcal{S}$ we get $|\delta_t(u)| = |\delta_{\varphi(t)}(f)| \le \|\delta_{\varphi(t)}\|\|f\|$ and point evaluations of u are bounded, hence $u \in H(k|_S)$. □

At this point, we are ready to state the characterization of the boundedness of the composition operator between two RKHSs.

Theorem 2.2.20 (Boundedness characterization of the composition operator)**.** Let k_1 and k_2 be kernels on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively and let $\varphi : \mathcal{X}_1 \to \mathcal{X}_2$ be a map. Then, the following are equivalent

(i) $\{f\circ\varphi \,:\, f\in H_{k_2}\} \subseteq H_{k_1}$,

(ii) $T_\varphi : H_{k_2} \to H_{k_1}$ is a bounded operator

(iii) there exists a constant $c > 0$ such that $k_2\circ\varphi \preceq c^2 k_1$.

Moreover, $c = \|T_\varphi\|$ is the smallest possible constant.

Proof. Let $f \in H_{k_2}$, let $x \in \mathcal{X}_1$ and suppose that T_φ is bounded, then by Cauchy-Schwarz

$$|\delta_x(f\circ\varphi)| = |\delta_{\varphi(x)}(f)| = |\langle f, T_\varphi k_2(x,\cdot)\rangle_{H_{k_2}}| \le \|f\|_{H_{k_2}}\|T_\varphi k_2(x,\cdot)\| \le \|f\|_{H_{k_2}}\|T_\varphi\|\|k_2(x,\cdot)\|_{H_{k_2}}.$$

So point evaluations of $f\circ\varphi$ are bounded and (ii) implies (i). Assume (iii) and let $f \in H_{k_2}$. From Theorem 2.2.9 we have that $f(x)\overline{f(y)} \preceq \|f\|^2_{H_{k_2}} k(x,y)$ for all $(x,y) \in \mathcal{X}_2\times\mathcal{X}_2$. Let $(s,t)\in\mathcal{X}_1\times\mathcal{X}_1$ and we have $f(\varphi(s))\overline{f(\varphi(t))} \preceq \|f\|^2_{H_{k_2}} k(\varphi(s),\varphi(t)) \preceq \|f\|^2_{H_{k_2}} c^2 k_1(s,t)$. Therefore $T_\varphi f = f\circ\varphi \in H_{k_1}$ and T_φ is bounded because $\|T_\varphi f\|_{H_{k_1}} \le c\|f\|_{H_{k_2}}$ and we have (ii). Note that (i) states $H(k_2\circ\varphi) \subseteq H_{k_1}$, which is equivalent to (iii) by Aronszajn's inclusion Theorem 2.2.11. As in the proof of Aronszajn's inclusion Theorem, one can choose the inclusion map to be T_φ and follow the same procedure, which shows the final statement. □

2.3 Vector-valued RKHS

Functions f in a RKHS H on some set $\mathcal{X}$ have values $f(x)$ in the real or complex numbers. This theory can be generalized to where the codomain of the functions in the RKHS is not the field $\mathbb{R}$ or $\mathbb{C}$ but a Hilbert space. This section builds up the necessary theory with some examples to later be used for applications in Section 2.4.

Definition 2.3.1 (vector-valued RKHS)**.** Let $\mathcal{C}$ be a Hilbert space and let $\mathcal{X}$ be a set. Denote $\mathcal{F}(\mathcal{X}, \mathcal{C})$ the vector space of functions from $\mathcal{X}$ to $\mathcal{C}$ under the usual pointwise addition and multiplication. Let $G \subseteq \mathcal{F}(\mathcal{X}, \mathcal{C})$ be a subspace, then G is called a a $\mathcal{C}$*-valued RKHS* on $\mathcal{X}$ provided that

(i) G is a Hilbert space.

(ii) for every $y \in \mathcal{X}$, the linear evaluation map $E_y : G \to \mathcal{C}$ given by $E_y(f) = f(y)$ is bounded.

We will also call a $\mathcal{C}$-valued RKHS a vector-valued RKHS, for some Hilbert space $\mathcal{C}$. The most basic example is taking copies of a scalar-valued RKHS.

Example 2.3.2 ($\mathcal{C} = \mathbb{C}^n$)**.** Let $\mathcal{C} = \mathbb{C}^n$ and H be a RKHS with kernel k on a set $\mathcal{X}$. Denote $H^n := \bigoplus_{i=1}^n H$, which is a Hilbert space, by Theorem A.1.2. For H^n to be a $\mathbb{C}^n$ valued RKHS, one needs to identify H^n with $\mathcal{F}(\mathcal{X}, \mathbb{C}^n)$. This can be via the following map,

$$\begin{aligned} L : H^n &\to \operatorname{ran}(L) \subseteq \mathcal{F}(\mathcal{X}, \mathbb{C}^n), \\ f &\mapsto L(f), \end{aligned}$$

where $L(f)$ is defined pointwise as $L(f)(x) = (f_1(x), \ldots, f_n(x)) \in \mathbb{C}^n$ for $x \in \mathcal{X}$ and $f \in H^n$. Point evaluations are bounded, let $x \in \mathcal{X}$ and $f \in H^n$

$$\|E_x(L(f))\|_{\mathbb{C}^n}^2 = \sum_{i=1}^n |f_i(x)|^2 \leq \sum_{i=1}^n \|f_i\|_H^2 \|k(x,\cdot)\|_H^2 = \|k(x,\cdot)\|_H^2 \sum_{i=1}^n |f_i|_H^2 = \|k(x,\cdot)\|_H^2 \|f\|_{H^n}^2. \tag{2.15}$$

We see that $\|E_x\|_{\mathcal{L}(H^n,\mathbb{C}^n)} \leq \|k(x,\cdot)\|_H$ and is bounded. We already conclude that $(H^n, \langle\cdot,\cdot\rangle_{H^n})$ is a $\mathbb{C}^n$ valued RKHS, the operator norm $\|E_x\|_{\mathcal{L}(H^n,\mathbb{C}^n)}$ can be determined with slightly more work. Let $f = (k(x,\cdot), 0, \ldots, 0) \in H^n$, then we have $\|E_x(L(f))\|_{\mathbb{C}^n} = |k(x,x)| = \|k(x,\cdot)\|_H^2 = \|f\|_{H^n} \|k(x,\cdot)\|_H$ and hence $\|E_y\| \geq \|k(x,\cdot)\|_H$, resulting in $\|E_x\|_{\mathcal{L}(H^n,\mathbb{C}^n)} = \|k(x,\cdot)\|_H$.

At this point, we have not yet defined what a kernel for a vector-valued RKHS should be. We review the scalar-valued case, as the the vector-valued case is a generalization of this. Let H be a scalar-valued RKHS on $\mathcal{X}$ that admits a kernel k. Once we fix $\mathbf{x} \in \mathcal{X}^n$, the kernel k induces a PSD matrix $\mathbf{K}$. A similar notion holds for the kernel of a vector-valued RKHS. To do so, we state what it means for matrices of operators to be positive.

Definition 2.3.3 (Positive operator)**.** Let $\mathcal{C}$ be a Hilbert space, an operator $T \in \mathcal{L}(\mathcal{C})$ is called *positive* if $\langle Tc, c\rangle_{\mathcal{C}} \geq 0$ for any $c \in \mathcal{C}$, written as $T \geq 0$.

Let $T = (T_{i,j})$ with $1 \leq i, j \leq n$ with $n \in \mathbb{N}$ where each $(T_{i,j}) \in \mathcal{L}(\mathcal{C})$ is a positive operator. We call T a matrix of operators on $\mathcal{C}$. One can identify, using the standard matrix-vector multiplication rules, T with an operator on $\mathcal{C}^n$. Let $\mathbf{c} \in \mathcal{C}^n$ and define

$$T(\mathbf{c}) := \begin{pmatrix} \sum_{j=1}^n T_{1,j}\mathbf{c}_j \\ \vdots \\ \sum_{j=1}^n T_{n,j}\mathbf{c}_j \end{pmatrix}.$$

The operator T is bounded, since

$$\|T(\mathbf{c})\|_{\mathcal{C}^n}^2 = \sum_{i=1}^n \left(\|\sum_{j=1}^n T_{i,j}\mathbf{c}_j\|_{\mathcal{C}}^2 \right) \overset{\triangle}{\leq} \sum_{i,j=1}^n \|T_{i,j}\mathbf{c}_j\|_{\mathcal{C}}^2 \leq \sum_{i,j=1}^n \|T_{i,j}\|_{\mathcal{L}(\mathcal{C})}^2 \|\mathbf{c}_j\|^2 \leq \|\mathbf{c}\|_{\mathcal{C}^n}^2 \sum_{i,j=1}^n \|T_{i,j}\|_{\mathcal{L}(\mathcal{C})}^2.$$

Definition 2.3.4 (Positive matrix of operators)**.** Let T be a matrix of operators on $\mathcal{C}$. We say that T is *positive*, written $T \geq 0$, if for any $\mathbf{c} \in \mathcal{C}^n$ the inner product

$$\langle T\mathbf{c}, \mathbf{c}\rangle_{\mathcal{C}^n} = \langle \Big(\sum_{j=1}^{n} T_{i,j}\mathbf{c}_j \Big)_{i=1}^{n}, \mathbf{c}\rangle_{\mathcal{C}^n} = \sum_{i,j=1}^{n} \langle T_{i,j}\mathbf{c}_j, \mathbf{c}_i\rangle_{\mathcal{C}} \geq 0.$$

Definition 2.3.5 (Operator-valued kernel)**.** Let $\mathcal{X}$ be a set and $\mathcal{C}$ be a Hilbert space and $k : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{C})$. Then k is an *operator-valued kernel* if $k(x, y) = k(y, x)^*$ for any $x, y \in \mathcal{X}$ and the matrix of operators defined by $K_{i,j} := k(x_i, x_j)$ is positive.

Let us look at the most simple example of an operator-valued kernel.

Example 2.3.6 (Simple example of an operator-valued kernel)**.** Let H be a scalar-valued RKHS with kernel k on a set $\mathcal{X}$ and let $\mathcal{C}$ be a Hilbert space. Then $\tilde{k} : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{C})$ defined by $\tilde{k}(x, y) = k(x, y)I_{\mathcal{C}}$, where $I_{\mathcal{C}}$ is the identity on $\mathcal{C}$, defines an operator-valued kernel.
Let $x, y \in \mathcal{X}$, we have $\tilde{k}(y, x)^* = \overline{k(y, x)}I_{\mathcal{C}}^* = k(x, y)I_{\mathcal{C}} = \tilde{k}(x, y)$. For the positivity, take $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{c} \in \mathcal{C}^n$, we get

$$\sum_{i,j=1}^{n} \langle \tilde{k}(\mathbf{x}_i, \mathbf{x}_j)c_j, c_i\rangle_{\mathcal{C}} = \sum_{i,j=1}^{n} k(\mathbf{x}_i, \mathbf{x}_j)\langle c_j, c_i\rangle_{\mathcal{C}} \geq 0.$$

since k is a scalar-valued kernel.

As in the scalar-valued case, an operator-valued kernel induces a vector-valued RKHS. The construction is explicit and similar to the scalar-valued case. To get some intuition of how this construction should be, we state the following proposition.

Proposition 2.3.7. Let G be a $\mathcal{C}$-valued RKHS on a set $\mathcal{X}$ with kernel k. The set $\text{span}\{E_x^* c : (x, c) \in \mathcal{X} \times \mathcal{C}\}$ is dense in G.

Proof. Call $D := \text{span}\{E_x^* c : (x, c) \in \mathcal{X} \times \mathcal{C}\}$ and let $f \in D^\perp$. Then $0 = \langle f, E_x^* c\rangle_G = \langle E_x f, c\rangle_{\mathcal{C}} = \langle f(x), c\rangle_{\mathcal{C}}$ for all $c \in \mathcal{C}$. As this holds for all $(x, c) \in \mathcal{X} \times \mathcal{C}$, we conclude that $f = 0$ and hence that D is dense in G. □

Lemma 2.3.8 (Moore's vector-valued theorem for RKHS)**.** Let $\mathcal{C}$ be a Hilbert space and let $k : \mathcal{X} \times \mathcal{X} \to \mathcal{L}(\mathcal{C})$ be an operator-valued kernel. There exists a Hilbert space $G \subseteq \mathcal{F}(\mathcal{X}, \mathcal{C})$ of $\mathcal{C}$ valued functions such that

- $[k_x c](\cdot) := k(\cdot, x)c \in G$ for any $x \in \mathcal{X}$, $c \in \mathcal{C}$,
- $\langle f, k_x c\rangle_H = \langle f(x), c\rangle_{\mathcal{C}}$ for any $f \in G$ and $c \in \mathcal{C}$.

We will also call this Hilbert space G_k.

Proof. Let $x \in \mathcal{X}$ and $c \in \mathcal{C}$, define $k_x : \mathcal{C} \to \mathcal{F}(\mathcal{X}, \mathcal{C})$ as $[k_x c](\cdot) := k(\cdot, x)c$. Let $G_{\text{pre}} := \text{span}\{k_x c : \in \mathcal{X}, c \in \mathcal{C}\}$. Let $f = \sum_{j=1}^{n} k_{\mathbf{x}_j}\mathbf{v}_j$ and $g = \sum_{i=1}^{n} k_{\mathbf{y}_i}\mathbf{w}_i$, where $\mathbf{x}, \mathbf{y} \in \mathcal{X}^n$ and $\mathbf{v}, \mathbf{w} \in \mathcal{C}^n$, be elements of G_{pre}. We may assume w.l.o.g. that f and g have an equal number of terms. Define the map

$$\langle \cdot, \cdot \rangle_{G_{\text{pre}}} : G_{\text{pre}} \times G_{\text{pre}} \to \mathbb{C},$$

$$\langle \cdot, \cdot \rangle_{G_{\text{pre}}}(f, g) \mapsto \langle f, g\rangle_{G_{\text{pre}}} := \sum_{i,j=1}^{n} \langle k(\mathbf{y}_i, \mathbf{x}_j)\mathbf{v}_j, \mathbf{w}_i\rangle_{\mathcal{C}}.$$

It is straightforward to check that $\langle \cdot, \cdot \rangle_{G_{\text{pre}}}$ is linear in the first component and conjugate linear in the second component. We have that $\langle f, f\rangle_{G_{\text{pre}}} \geq 0$, since k is an operator-valued kernel. Suppose that $f = 0$, meaning that $f(x) = 0$ for all $x \in \mathcal{X}$. Then, $\langle f, f\rangle_{G_{\text{pre}}} = \sum_{i,j=1}^{n} \langle k(\mathbf{x}_i, \mathbf{x}_j)\mathbf{v}_j, \mathbf{v}_i\rangle_{\mathcal{C}} = \sum_{i=1}^{n} \langle f(x), \mathbf{v}_i\rangle_{\mathcal{C}} = \sum_{i=1}^{n} \langle 0, \mathbf{v}_i\rangle_{\mathcal{C}} = 0$. Conversely, suppose that $\langle f, f\rangle_{G_{\text{pre}}} = \|f\|_{G_{\text{pre}}}^2 = 0$. Then, by the Cauchy-Schwarz inequality for quasi products we get $|\langle f, h\rangle| \leq \|f\|_{G_{\text{pre}}}\|h\|_{G_{\text{pre}}} = 0$ for any $h \in G_{\text{pre}}$. Let $h = k_y c$ for some $y \in \mathcal{X}$ and $c \in \mathcal{C}$, then $0 = \langle f, k_y c\rangle_{G_{\text{pre}}} = \langle \sum_{j=1}^{n} k(y, \mathbf{x}_j)\mathbf{v}_j, c\rangle_{\mathcal{C}} = \langle f(y), w\rangle_{\mathcal{C}}$. Since w is arbitrary, we conclude that $f(y) = 0$ for any $y \in \mathcal{X}$ and f is the zero function. The Hilbert space G is the completion of the inner product space $(G_{\text{pre}}, \langle \cdot, \cdot \rangle_{G_{\text{pre}}})$. □

It is not difficult to see that then G_k from Lemma 2.3.8 is a vector-valued RKHS.

Corollary 2.3.9. The Hilbert space G_k from Lemma 2.3.8 is a vector-valued RKHS. The evaluation functional $E_x = k_x^*$, the adjoint of the map $k_x \in \mathcal{L}(\mathcal{C}, G_k)$ for $x \in \mathcal{X}$.

Proof. Let $x \in \mathcal{X}$ and $c \in \mathcal{C}$. It is clear that $k_x : \mathcal{C} \to G_k$ is linear, using Cauchy-Schwarz $\|k_x c\|_{G_k}^2 = |\langle k(x,x)c, c\rangle_{\mathcal{C}}| \leq \|k(x,x)\|_{\mathcal{L}(\mathcal{C})}\|c\|_{\mathcal{C}}^2$ and we conclude that $\|k_x\|_{\mathcal{L}(\mathcal{C},G_k)} \leq \|k(x,x)\|_{\mathcal{L}(\mathcal{C})}^{1/2}$. Since k_x is a bounded linear map between Hilbert spaces, the adjoint exists. Let $f \in G_k$, the property $\langle f, k_x c\rangle_{G_k} = \langle f(x), c\rangle_{\mathcal{C}}$, we conclude that $E_x(f) = f(x) = k_x^*(f)$. It follows that $k(x,y) = k_x^* k_y$, for $x, y \in \mathcal{X}$. Since the evaluation map is bounded, we conclude that G_k is a vector-valued RKHS. □

Proposition 2.3.10. Let G_1 and G_2 be a $\mathcal{C}$-valued RKHSs on a set $\mathcal{X}$. Suppose that $k_1 = k_2$, then $G_1 = G_2$ as sets, and $\|f\|_{G_1} = \|f\|_{G_2}$.

Proof. Since both $k_i(x,y) = E_x^{(i)} E_y^{(i)*}$ with $x, y \in \mathcal{X}$ and $i = 1, 2$ are kernels for G_i, it follows from Proposition 2.3.7 that $D_i := \operatorname{span}\{E_x^{(i)*}c : (x,c) \in \mathcal{X} \times \mathcal{C}\}$ is dense in G_i. Since $E^{(1)*} = E^{(2)*}$, we have $G_1 = \overline{D_1} = \overline{D_2} = G_2$. Let $f \in G_1$, then $f = \sum_{i=1}^{\infty} E_{\mathbf{x}_i}^{(1)*}\mathbf{c}_i$ for $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{c} \in \mathcal{C}^n$ and $k_1 = k_2$ we have,

$$
\begin{aligned}
\|f\|_{G_1}^2 &= \lim_{n\to\infty} \|\sum_{i=1}^{n} E_{\mathbf{x}_i}^{(1)*}\mathbf{c}_i\|^2 = \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_i}^{(1)*}\mathbf{c}_i, E_{\mathbf{x}_j}^{(1)*}\mathbf{c}_j\rangle = \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_j}^{(1)} E_{\mathbf{x}_i}^{(1)*}\mathbf{c}_i, \mathbf{c}_j\rangle, \\
&= \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle k_1(\mathbf{x}_i, \mathbf{x}_j)\mathbf{c}_i, \mathbf{c}_j\rangle = \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle k_2(\mathbf{x}_i, \mathbf{x}_j)\mathbf{c}_i, \mathbf{c}_j\rangle = \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_j}^{(2)} E_{\mathbf{x}_i}^{(2)*}\mathbf{c}_i, \mathbf{c}_j\rangle, \\
&= \lim_{n\to\infty} \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_i}^{(2)*}\mathbf{c}_i, E_{\mathbf{x}_j}^{(2)*}\mathbf{c}_j\rangle = \lim_{n\to\infty} \|\sum_{i=1}^{n} E_{\mathbf{x}_i}^{(2)*}\mathbf{c}_i\|^2 = \|f\|_{G_2}^2.
\end{aligned}
$$

□

An operator-valued kernel gives rise to a vector-valued RKHS. In the scalar-valued case, the existence of a reproducing kernel in an abstract RKHS is guaranteed, by Riesz-Fréchet . However, using Corollary 2.3.9, an operator-valued kernel can always be constructed from an abstract vector-valued RKHS from Definition 2.3.1.

Proposition 2.3.11. Let $\mathcal{C}$ be a Hilbert space Let G be a $\mathcal{C}$ valued RKHS on a set $\mathcal{X}$. Then $k(x,y) := E_y E_x^*$ is an operator-valued kernel.

Proof. Let $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{c} \in \mathcal{C}^n$, we get

$$
\sum_{i,j=1}^{n} \langle k(\mathbf{x}_i, \mathbf{x}_j)\mathbf{c}_j, \mathbf{c}_i\rangle_{\mathcal{C}} = \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_i} E_{\mathbf{x}_j}^* \mathbf{c}_j, \mathbf{c}_i\rangle_{\mathcal{C}} = \sum_{i,j=1}^{n} \langle E_{\mathbf{x}_j}^* \mathbf{c}_j, E_{\mathbf{x}_i}^* \mathbf{c}_i\rangle_{\mathcal{C}} = \|\sum_{i=1}^{n} E_{\mathbf{x}_i}^* \mathbf{c}_i\|_{\mathcal{C}}^2 \geq 0.
$$

□

With Proposition 2.3.11, let us have a look how this definition breaks down in the case $\mathcal{C} = \mathbb{C}$. Let G be a $\mathbb{C}$ valued RKHS on a set $\mathcal{X}$. If $E_y E_x^* = k(x,y)$, then we expect $E_x^* = k(x,\cdot)$.
The kernel $k(x,y)$ is an element of $\mathcal{L}(\mathbb{C})$, and not of $\mathbb{C}$. However, there is a natural isomorphism between these two spaces by identifying linear maps $T \in \mathcal{L}(\mathbb{C})$ with the complex number $T(1) \in \mathbb{C}$. Furthermore, the element $k(x,\cdot) \in \mathcal{L}(\mathbb{C}, H)$ and not an element of H. Similarly, one can identify linear maps in $\mathcal{L}(\mathbb{C}, H)$ with elements of H uniquely. This can be done by mapping $S \in \mathcal{L}(\mathbb{C}, H)$ to $S(1) \in H$, which is one-to-one.
To see how the adjoint map $E_x^* : \mathbb{C} \to H$ acts on $\alpha \in \mathbb{C}$, we let $f \in H$ and $x \in \mathcal{X}$, then

$$
\langle f, E_x^*(\alpha)\rangle_H = \langle E_x(f), \alpha\rangle_{\mathbb{C}} = \langle f(x), \alpha\rangle_{\mathbb{C}} = \overline{\alpha}\langle f, k(x,\cdot)\rangle_H = \langle f, \alpha k(x,\cdot)\rangle_H.
$$

We see that $E_x^*(\alpha) = \alpha k(x, \cdot)$ and that $E_x^*(1)$ is naturally identified with $k(x, \cdot)$. The operator $E_y E_x^*(1) \in \mathcal{L}(\mathbb{C})$ boils down to

$$E_y E_x^*(1) = E_y(k(x, \cdot) = k(x, y).$$

With this identification, we can revisit Example 2.3.2, where we looked at copies of existing scalar-valued RKHSs. Using Proposition 2.3.11, we can find the kernel for this $\mathbb{C}^n$ valued RKHS.

Example 2.3.12 (Example 6.4 from [PR16])**.** From Example 2.3.2, we have seen that $(H^n, \langle \cdot, \cdot \rangle_{H^n})$ is a $\mathbb{C}^n$ valued RKHS. Let $x, y \in \mathcal{X}$, we will work out the kernel $\tilde{k}(x, y) = \tilde{E}_y \tilde{E}_x^*$. The kernel for H will be denoted as k. The bounded linear maps $\mathcal{L}(\mathbb{C}^n)$ can be identified with the space $\mathbb{C}^{n \times n}$, by letting the linear maps acting on standard basis vector and constructing the matrix representation. Let $f \in H^n$, with abuse of notation we will write $f(x)$ instead of $(Lf)(x)$ as in Example 2.3.2. Let $\mathbf{a} \in \mathbb{C}^n$, we get

$$\langle f, \tilde{E}_x^*(\mathbf{a}) \rangle_{H^n} = \langle f(x), \mathbf{a} \rangle_{\mathbb{C}^n} = \sum_{i=1}^n f_i(x) \overline{\mathbf{a}_i} = \sum_{i=1}^n \langle f, \mathbf{a}_i k(x, \cdot) \rangle_H = \langle f, (\mathbf{a}_1 k(x, \cdot), \ldots, \mathbf{a}_n k(x, \cdot)) \rangle_{H^n}.$$

We see that $\tilde{E}_x^*(\mathbf{a}) = (\mathbf{a}_1 k(x, \cdot), \ldots, \mathbf{a}_n k(x, \cdot))$. Observe that $\tilde{E}_x^*(e_i)$ is the vector $k(x, \cdot)$ in the i-th component and zeroes elsewhere. As we can identify $\tilde{k} \in \mathcal{L}(\mathbb{C}^n)$ with complex square matrices of size $n \times n$, we evaluate the (i, j) component of $\tilde{k}$,

$$\langle \tilde{k}(x, y) e_i, e_j \rangle_{\mathbb{C}^n} = \langle \tilde{E}_y \tilde{E}_x^* e_i, e_j \rangle_{\mathbb{C}^n} = \langle \tilde{E}_x^* e_i, \tilde{E}_y^* e_j \rangle_{H^n} = \begin{cases} k(x, y) & i = j, \\ 0 & i \neq j. \end{cases}$$

Therefore, the kernel function $\tilde{k}(x, y)$ can be written as a product with the identity matrix I_n as follows

$$\tilde{k}(x, y) = \begin{bmatrix} k(x, y) & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & k(x, y) \end{bmatrix} = k(x, y) I_n.$$

Example 2.3.12 can be extended to the infinite dimensional case.

Example 2.3.13 (6.5 from [PR16])**.** Let H be RKHS on a set $\mathcal{X}$ that admits kernel k. Define the set

$$\ell^2(H) := \{(f_n)_{n \in \mathbb{N}} \,:\, f_n \in H \ \forall n \in \mathbb{N}, \sum_{i=1}^\infty \|f_i\|_H^2 < \infty\}$$

and the map

$$\langle \cdot, \cdot \rangle_{\ell^2(H)} : \ell^2(H) \times \ell^2(H) \to \mathbb{C},$$
$$(f, g) \mapsto \langle f, g \rangle_{\ell^2(H)} := \sum_{i=1}^\infty \langle f_i, g_i \rangle_H.$$

Using Cauchy-Schwarz, we see that the sum always converges

$$|\langle f, g \rangle_{\ell^2(H)}| \leq \sum_{i=1}^\infty |\langle f_i, g_i \rangle_H| \leq \sum_{i=1}^\infty \|f_i\|_H \|g_i\|_H \leq \sum_{i=1}^\infty \max(\|f_i\|_H^2, \|g_i\|_H^2) \leq \sum_{i=1}^\infty \|f_i\|_H^2 + \|g_i\|_H^2 < \infty$$

Using that $\langle \cdot, \cdot \rangle_H$ is an inner product, it is not hard to show that this map defines an inner product on $\ell^2(H)$. By defining pointwise limits, it follows that $(\ell^2(H), \langle \cdot, \cdot \rangle_{\ell^2(H)}$ is in fact a Hilbert space. Similar to the identification in Example 2.3.2, one can identify $\ell^2(H)$ with ℓ^2-valued functions. For

$f \in \ell^2(H)$ and $x \in \mathcal{X}$, write with abuse of notation, $f(x) = (f_1(x), f_2(x), \dots)$. One can show that $f(x)$ has finite ℓ^2-norm

$$\|E_x(f)\|_{\ell^2}^2 = \|f(x)\|_{\ell^2}^2 = \sum_{i=1}^{\infty} |f_i(x)|^2 = \sum_{i=1}^{\infty} |\langle f, k(x,\cdot)\rangle|^2 \leq \sum_{i=1}^{\infty} \|f_i\|_H^2 \|k(x,\cdot)\|_H^2 < \infty.$$

Since $E_x : H \to \ell^2$ is bounded, the space $\ell^2(H)$ is an ℓ^2 valued RKHS. Let $a = (a_n)_{n\in\mathbb{N}}$ be an ℓ^2 sequence, then

$$\langle f, E_x^*(a)\rangle_{\ell^2(H)} = \langle f(x), a\rangle_{\ell^2} = \sum_{i=1}^{\infty} f_i(x)\overline{a_i} = \sum_{i=1}^{\infty} \langle f, a_i k(x,\cdot)\rangle_{\ell^2(H)} = \langle f, (a_n k(x,\cdot)_{n\in\mathbb{N}}\rangle_{\ell^2(H)}. \quad (2.16)$$

Let $y \in \mathcal{X}$ the inner product between the kernel and the canonical basis vectors for ℓ^2, e_i and e_j

$$\langle E_y E_x^* e_i, e_j\rangle_{\ell^2} = \langle E_x^* e_i, E_y^* e_j\rangle_{\ell^2(H)} = \langle k(x,\cdot)e_i, k(y,\cdot)e_i\rangle_{\ell^2(H)} = \begin{cases} k(x,y) & i = j, \\ 0 & i \neq j. \end{cases}$$

Hence, the operator-valued kernel for $\ell^2(H)$ is given by $\tilde{k}(x,y) = k(x,y)I_{\ell^2}$, where I_{ℓ^2} is the identity operator on ℓ^2.

Finally, one can bound any $f(x) \in \mathcal{C}$ by the norm of f and the operator norm of $k(x,x)$.

Proposition 2.3.14. Let G be a $\mathcal{C}$-valued RKHS on a set $\mathcal{X}$ with kernel k and let $x \in \mathcal{X}^n$. Then $\|f(x)\|_{\mathcal{C}} \leq \|k(x,x)\|_{\mathcal{L}(\mathcal{C})}^{1/2} \|f\|_G$.

Proof. Using Cauchy-Schwarz and the fact that $f(x) = k_x^*(f)$

$$\begin{aligned} \|f(x)\|_{\mathcal{C}}^2 &= |\langle f(x), f(x)\rangle_{\mathcal{C}}| = |\langle k_x^* f, k_x^* f\rangle_{\mathcal{C}}| = |\langle f, k_x k_x^* f\rangle_{\mathcal{C}}| \leq \|k_x k_x^* f\|_G \|f\|_G, \\ &\leq \|k_x k_x^*\|_{\mathcal{L}(H)} \|f\|_G^2 = \|k_x^* k_x\|_{\mathcal{L}(C)} \|f\|_G^2 = \|k(x,x)\|_{\mathcal{L}(\mathcal{C})} \|f\|_G^2. \end{aligned}$$

□

2.4 Results for learning theory

We present two key abstract results with profound implications in learning theory. The first is the representer theorem, which is essential for deriving closed-form solutions when optimizing over a RKHS. This theorem leads to an optimal way of recovering a function that matches a set of data points. We show that a projection onto a finite-dimensional subspace of a RKHS can be viewed as a regression problem. The second result describes how elements in a specific vector-valued RKHS can be evaluated in terms of a Hilbert-Schmidt operator. Both of these results are used in addressing the estimation of the Koopman operator from snapshot data, which will be introduced in Chapter 4.

2.4.1 Representer theorem

Theorem 2.4.1 (Representer theorem)**.** Let k be a kernel with values in $\mathbb{R}$ on a set $\mathcal{X}$ and let H its associated RKHS. Fix $\mathbf{x} \in \mathcal{X}^n$ and let $\tau : \mathbb{C}^n \to \mathbb{R}_{\geq 0}$ and $\zeta : \mathbb{R}_{\geq 0} \to \mathbb{R}$ where ζ is nondecreasing and τ depends on f only through $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$. If the minimization problem

$$\min_{f\in H} \tau(f(\mathbf{x}_1), \dots, f(\mathbf{x}_n) + \zeta(\|f\|_H) \quad (2.17)$$

has a solution, it can be written as

$$f^*(\cdot) = \sum_{i=1}^{n} \mathbf{c}_i^{f^*} k(\mathbf{x}_i, \cdot), \quad (2.18)$$

with $\mathbf{c}^{f^*} \in \mathbb{R}^n$. If ζ is strictly increasing, then every solution is of the form (2.18).

Proof. Let $S := \text{span}\{k(\mathbf{x}_i, \cdot) : 1 \le i \le n\}$, which is a finite dimensional subspace of H and thus closed. Therefore, we may decompose $H = S \oplus S^{\perp}$ and any minimizer $f^* = f_S + f_{\perp}$ evaluated at a data point $\mathbf{x}_i$ results in,

$$f^*(\mathbf{x}_i) = \langle f, k(\mathbf{x}_i, \cdot)\rangle = \langle f_S, k(\mathbf{x}_i, \cdot)\rangle + \langle f_{\perp}, k(\mathbf{x}_i, \cdot)\rangle = f_S(\mathbf{x}_i).$$

We may bound the objective from below by,

$$\begin{aligned}\tau(f^*(x_1), \dots, f^*(x_n) + \zeta(\|f^*\|) &= \tau(f_S(x_1), \dots, f_S(x_n) + \zeta(\|f^*\|), \\ &\ge \tau(f_S(x_1), \dots, f_S(x_n) + \zeta(\|f_S\|). \end{aligned} \tag{2.19}$$

The bound follows since $\|f^*\|^2 = \|f_S\|^2 + \|f_{\perp}\|^2$ and ζ is nondecreasing. If f^* minimizes (2.17), then so does f_S, which is of the form in (2.18). If ζ is strictly increasing, we get a strict inequality in equation (2.19) and we conclude that f^* must be of the form (2.18). □

Example 2.4.2 (Regression in a RKHS)**.** Let k be a real kernel on $\mathcal{X}$ and the data $(\mathbf{x}, \mathbf{y}) \in (\mathcal{X} \times \mathbb{R})^N$. Consider the problem

$$\min_{f \in H_k} \|f\|_{H_k} \quad \text{s.t.} \quad f(\mathbf{x}_i) = \mathbf{y}_i,\ 1 \le i \le N.$$

By the representer theorem, the solution lies in in the finite dimensional subspace of H_k spanned by the canonical feature maps centered at the points $\mathbf{x}_i$. Imposing the constraint results in

$$f^*(\mathbf{x}_j) = \sum_{i=1}^{n} \mathbf{c}_i^{f^*} k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{y}_j, \quad 1 \le j \le N.$$

We see that the coefficients $\mathbf{c}^{f^*} = \mathbf{K}^{-1}\mathbf{y} \in \mathbb{R}^N$, where $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ for indices $1 \le i, j \le N$ and $\mathbf{y} \in \mathbb{R}^N$.

Example 2.4.3 (Projection on finite dimensional subspace of a RKHS)**.** Let $\mathbf{x} \in \mathcal{X}^N$, let k be a kernel on $\mathcal{X}$ and assume that $g \in H_k$. Let $V_{\mathbf{x}}$ be the subspace spanned by the functions $\text{span}\{k(\mathbf{x}_1, \cdot), \dots, k(\mathbf{x}_N, \cdot)\} \subset H_k$. Finding the best approximation of $f \in H_k$ in the subspace $V_{\mathbf{x}}$ results in the regression problem

$$\min_{f \in V} \|g - f\|^2_{H_k}.$$

It is clear that $f := P_{V_{\mathbf{x}}} g$, the projection onto the closed linear subspace $V_{\mathbf{x}}$ solves the problem. It is a standard result from Hilbert space theory that for $f = P_{V_{\mathbf{x}}} g$ we have $g - f \perp V_{\mathbf{x}}$. Thus,

$$0 = \langle g - f, k(\mathbf{x}_i, \cdot)\rangle_{H_k}, \quad 1 \le i \le N.$$

By the reproducing property, this is equivalent to $g(\mathbf{x}_i) = f(\mathbf{x}_i)$ for $1 \le i \le N$. Let $g_{\mathbf{x}} = (g(\mathbf{x}_1), \dots, g(\mathbf{x}_N))^T \in \mathbb{R}^{N \times 1}$ and since $f_{\mathbf{x}} = g_{\mathbf{x}}$ we get the formula for the projection map

$$\begin{aligned} P_{V_{\mathbf{x}}} &: H_k \to V_{\mathbf{x}}, \\ f(\cdot) &\mapsto [P_{V_{\mathbf{x}}} f](\cdot) := \sum_{i=1}^{N} (\mathbf{K}^{-1} f_{\mathbf{x}})_i k(\mathbf{x}_i, \cdot). \end{aligned}$$

Observe that the coefficients of the optimal solution to our regression problem in Example 2.4.2 is equivalent to the projection, with $g_{\mathbf{x}} = \mathbf{y}$. Thus, the regression problem in a RKHS can be viewed as a projection onto the subspace $V_{\mathbf{x}}$.

Example 2.4.4 (Regression in RKHS with a penalty term)**.** Let k be a real kernel on $\mathcal{X}$ and the data $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N \in \mathcal{X} \times \mathbb{R}$. Consider the problem

$$\min_{f \in H_k} \frac{1}{N} \sum_{i=1}^{N} (\mathbf{y}_i - f(\mathbf{x}_i))^2 + \lambda \|f\|^2_{H_k},$$

for some $\lambda > 0$. This optimization problem is of the form described in the representer theorem. Thus, the solution f^* is of the form (2.18). Observe that

$$\|f^*\|^2_{H_k} = \langle \sum_{i=1}^{N} \mathbf{c}_i^{f^*} k(\mathbf{x}_i, \cdot), \sum_{j=1}^{N} \mathbf{c}_j^{f^*} k(\mathbf{x}_j, \cdot) \rangle = (\mathbf{c}^{f^*})^T \mathbf{K} \mathbf{c}^{f^*}.$$

Evaluating f^* at $\mathbf{x}_j$ results in

$$f^*(\mathbf{x}_j) = \sum_{i=1}^{n} \mathbf{c}_i^{f^*} k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{e}_j^T \mathbf{K} \mathbf{c}^{f^*}.$$

Substituting this into the original problem results in an equivalent convex optimization problem

$$\begin{aligned}
\min_{\mathbf{c}^{f^*} \in \mathbb{R}^N} \frac{1}{N} \|\mathbf{y} - \mathbf{K}\mathbf{c}^{f^*}\|^2_{\mathbb{R}^N} + (\mathbf{c}^{f^*})^T \mathbf{K} \mathbf{c}^{f^*} &= \min_{\mathbf{c}^{f^*} \in \mathbb{R}^N} \frac{1}{N} (\mathbf{y} - \mathbf{K}\mathbf{c}^{f^*})^T (\mathbf{y} - \mathbf{K}\mathbf{c}^{f^*}) + \lambda (\mathbf{c}^{f^*})^T \mathbf{K} \mathbf{c}^{f^*}, \\
&= \min_{\mathbf{c}^{f^*} \in \mathbb{R}^N} \frac{1}{N} \mathbf{y}^T \mathbf{y} - \frac{2}{N} \mathbf{y}^T \mathbf{K} \mathbf{c}^{f^*} + \frac{1}{N} (\mathbf{c}^{f^*})^T \mathbf{K}^T \mathbf{K} \mathbf{c}^{f^*} + \lambda (\mathbf{c}^{f^*})^T \mathbf{K} \mathbf{c}^{f^*}, \\
&= \min_{\mathbf{c}^{f^*} \in \mathbb{R}^N} \frac{1}{N} \mathbf{y}^T \mathbf{y} + \frac{1}{N} (\mathbf{c}^{f^*})^T (\mathbf{K}^2 + \lambda N \mathbf{K}) \mathbf{c}^{f^*} - \frac{2}{N} \mathbf{y}^T \mathbf{K} \mathbf{c}^{f^*}.
\end{aligned}$$

Taking the gradient with respect to $\mathbf{c}^{f^*}$ and setting equal to zero results in

$$\begin{aligned}
\frac{2}{N} \mathbf{K}^2 \mathbf{c}^{f^*} + 2\lambda \mathbf{K} \mathbf{c}^{f^*} - \frac{2}{N} \mathbf{K} \mathbf{y} &= 0, \\
\mathbf{K}(\mathbf{K} + \lambda N I) \mathbf{c}^{f^*} &= \mathbf{K} \mathbf{y}.
\end{aligned}$$

Solving for $\mathbf{c}^{f^*}$ yields the coefficients

$$\mathbf{c}^{f^*} = (\mathbf{K} + \lambda N I)^{-1} \mathbf{y}.$$

2.4.2 Hilbert-Schmidt operators and vector-valued RKHS

We will show that the tensor product of two real separable Hilbert spaces H_1 and H_2, denoted $H_1 \otimes H_2$, is isometrically isomorphic to the class of Hilbert-Schmidt operators from H_2 to H_1, denoted $\mathrm{HS}(H_2, H_1)$. We write $\cong_\Xi$ to denote that two spaces are isometrically isomorphic under the map Ξ. To do so, we briefly introduce the concept of tensor products of Hilbert spaces. Let $v \in H_1$ and $e \in H_2$, define the bilinear map

$$\begin{aligned}
v_1 \otimes e_2 : (H_1 \times H_2) &\to \mathbb{R}, \\
(v_2, e_2) &\mapsto [v_1 \otimes e_1](v_2, e_2) = \langle v_1, v_2 \rangle_{H_1} \langle e_1, e_2 \rangle_{H_2}.
\end{aligned}$$

Let $\mathcal{E} := \mathrm{span}\{v \otimes e : (v, e) \in H_1 \times H_2\}$, the set of all finite linear combinations of such bilinear maps. We equip this space with the inner product

$$\langle v_1 \otimes e_1, v_2 \otimes e_2 \rangle_\otimes := \langle v_1, v_2 \rangle_{H_1} \langle e_1, e_2 \rangle_{H_2}.$$

One can readily check that this is in fact an inner product by using the properties of $\langle \cdot, \cdot \rangle_{H_1}$ and $\langle \cdot, \cdot \rangle_{H_2}$. The completion of $\mathcal{E}$ with respect to the inner product $\langle \cdot, \cdot \rangle_\otimes$ is called the tensor product of the Hilbert spaces H_1 and H_2, denoted $(H_1 \otimes H_2, \langle \cdot, \cdot \rangle_{H_1 \otimes H_2})$. Since both H_1 and H_2 are separable, it is not surprising that the tensor product is separable too, with a predictable basis.

Lemma 2.4.5 (Basis for tensor product)**.** Let H_1 and H_2 be two separable real Hilbert spaces with bases $(v_j)_{j=1}^\infty$ and $(e_j)_{j=1}^\infty$, respectively. Then, the basis for the tensor product $H_1 \otimes H_2$ is given by $(v_j \otimes e_j)_{j=1}^\infty$.

Proof. Define $\mathcal{B} := (v_j \otimes e_j)_{j=1}^{\infty}$ and observe that for any $a \in \mathbb{R}$ and $f \otimes g \in H_1 \otimes H_2$ we get $a(f \otimes g) = a \sum_{i=1}^{\infty} v_i \otimes e_i = \sum_{i=1}^{\infty} (av_i) \otimes e_i = (af) \otimes g$. Thus, an arbitrary element in $H_1 \otimes H_2$ may be written as $f \otimes g$. Define $h_n^1 := \sum_{i=1}^{n} \langle f, v_i \rangle v_i$ and $h_n^2 := \sum_{i=1}^{n} \langle g, e_i \rangle e_i$ and note that $h_n^1 \otimes h_n^2 \in \operatorname{span}\{\mathcal{B}\}$. To show that $\mathcal{B}$ is fundamental in $H_1 \otimes H_2$ we compute

$$\|f \otimes g - h_n^1 \otimes h_n^2\|_{H_1 \otimes H_2}^2 = \|f - h_n^1\|_{H_1}^2 \|g - h_n^2\|_{H_2}^2 \to 0 \quad (n \to \infty).$$

□

Theorem 2.4.6 ($H_1 \otimes H_2 \cong \mathrm{HS}(H_2, H_1)$)**.** Let H_1 and H_2 be separable Hilbert spaces. Then $H_1 \otimes H_2$ is isometrically isomorphic to $\mathrm{HS}(H_2, H_1)$.

Proof. Let $(v_i)_{i=1}^{\infty}$ and $(e_j)_{j=1}^{\infty}$ be ONB for H_1 and H_2, respectively, and Let $B \in H_1 \otimes H_2$ be arbitrary. Then, B can be written as $B = \sum_{i=1}^{\infty} f_i \otimes g_i$ with $f_i \in H_1$ and $g_i \in H_2$ for $i \geq 1$ and $\infty > \|B\|_{H_1 \otimes H_2}^2 = \|\sum_{i=1}^{\infty} f_i \otimes g_i\|_{H_1 \otimes H_2}^2 = \sum_{i=1}^{\infty} \|f_i\|_{H_1}^2 \|g_i\|_{H_2}^2$. Define the map

$$\begin{aligned} &\Xi : H_1 \otimes H_2 \to \mathcal{L}(H_2, H_1), \\ &\sum_{i=1}^{\infty} f_i \otimes g_i \mapsto \Xi(\sum_{i=1}^{\infty} f_i \otimes g_i)(\cdot) := \sum_{i=1}^{\infty} \langle \cdot, g_i \rangle_{H_2} f_i. \end{aligned}$$

The linearity is clear, let $h \in H_2$ and use Cauchy-Schwarz to show the boundedness

$$\|[\Xi(B)](h)\|_{H_1}^2 = \|\sum_{i=1}^{\infty} \langle h, g_i \rangle_{H_2} f_i\|_{H_1}^2 \leq \|h\|_{H_2}^2 \sum_{i=1}^{\infty} \|f_i\|_{H_1}^2 \|g_i\|_{H_2}^2 = \|h\|_{H_2}^2 \|\sum_{i=1}^{\infty} f_i \otimes g_i\|_{H_1 \otimes H_2}^2 = \|h\|_{H_2}^2 \|B\|_{H_1 \otimes H_2}^2.$$

By definition of the operator norms on $\mathcal{L}(H_2, H_1)$ and $\mathcal{L}(H_1 \otimes H_2, \mathcal{L}(H_2, H_1))$, we conclude that

$$\|\Xi(B)\|_{\mathcal{L}(H_2, H_1)} \leq \|B\|_{H_1 \otimes H_2}, \quad \text{and} \quad \|\Xi\|_{\mathcal{L}(H_1 \otimes H_2, \mathcal{L}(H_2, H_1))} \leq 1.$$

We claim that Ξ maps into $\mathrm{HS}(H_2, H_1)$, by expressing g_i as a linear combination of basis elements, consider

$$\begin{aligned} \|\Xi(B)\|_{HS(H_2,H_1)}^2 = \sum_{j=1}^{\infty} \|[\Xi(B)](e_j)\|_{H_1}^2 &= \sum_{j=1}^{\infty} \|\sum_{i=1}^{\infty} \langle e_j, g_i \rangle_{H_2} f_i\|_{H_1}^2, \\ &= \sum_{j=1}^{\infty} \|\sum_{i=1}^{\infty} \langle e_j, \sum_{k=1}^{\infty} \langle e_k, g_i \rangle_{H_2} e_k \rangle_{H_2} f_i\|_{H_1}^2, \\ &= \sum_{j=1}^{\infty} \|\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \langle g_i, e_k \rangle_{H_2} \langle e_j, e_k \rangle_{H_2} f_i\|_{H_1}^2, \\ &= \|\sum_{i=1}^{\infty} \sum_{k=1}^{\infty} \langle g_i, e_k \rangle_{H_2} f_i\|_{H_1}^2 \quad (j = k), \\ &\leq \sum_{i=1}^{\infty} \|f_i\|_{H_1}^2 \sum_{k=1}^{\infty} |\langle g_i, e_k \rangle_{H_2}|^2, \\ &= \sum_{i=1}^{\infty} \|f_i\|_{H_1}^2 \|g_i\|_{H_2}^2, \\ &= \|\sum_{i=1}^{\infty} f_i \otimes g_i\|_{H_1 \otimes H_2}^2 = \|B\|_{H_1 \otimes H_2}^2 < \infty. \end{aligned}$$

We conclude that Ξ maps into $HS(H_2, H_1)$ with $\|\Xi\|_{\mathcal{L}(H_1\otimes H_2, HS(H_2,H_1))} \leq 1$. On the other hand, by evaluating Ξ on $v_i \otimes e_i \in H_1 \otimes H_2$ we get

$$\begin{aligned}\|\Xi\|^2_{\mathcal{L}(H_1\otimes H_2, HS(H_2,H_1))} = \sup_{\substack{B\in H_1\otimes H_2\\ \|B\|\leq 1}} \|\Xi(B)\|^2_{HS(H_2,H_1)} &\geq \|\Xi(v_i\otimes e_j)\|^2_{HS(H_2,H_1)},\\ &= \sum_{k=1}^{\infty} \|\langle e_k, e_j\rangle v_i\|^2_{H_1},\\ &= \|v_i\|^2_{H_1} = 1.\end{aligned}$$

We conclude that Ξ is an isometry. To show that Ξ is surjective, it suffices to show that the range maps into a dense subset of $\mathrm{HS}(H_2, H_1)$. From Proposition 2.2.3 any $D \in \mathrm{HS}(H_2, H_1)$ is compact and hence finitely approximable, since H_1 is a Hilbert space. Hence, the set of finite rank operators between H_2 and H_1 is dense in $\mathrm{HS}(H_2, H_1)$. Any bounded finite rank operator $D : H_2 \to H_1$ with rank n can be written as $D(\cdot) = \sum_{i=1}^n \langle \cdot, e_i^D\rangle v_i^D$, for specific $(v_i^D)_{i=1}^n \in H_1$ and $(e_i^D)_{i=1}^n \in H_2$. Choose $z^D = \sum_{i=1}^n v_i^D \otimes e_i^D \in H_1 \otimes H_2$, then $\Xi(z^D) = D$ and Ξ is surjective as it maps in a dense set. Since Ξ is isometric and surjective, we conclude that $H_1 \otimes H_2$ is isometrically isomorphic to $\mathrm{HS}(H_2, H_1)$ under Ξ. □

Let k be a real-valued kernel on a set $\mathcal{X}$ and F be real-valued separable Hilbert space. From Example 2.3.6 we have seen that $\Gamma(x, y) = k(x, y)I_F$ defines an operator-valued kernel. The next theorem allows us to make the connection between the tensor product of a RKHS and a vector-valued RKHS. This theorem is a slight generalization of the case studied by Carmeli et. al. in [Car+10, Example 3.6(i)] where the F is replaced with H_k. Keep in mind that for the remainder of this thesis, we will assume (1- 2).

Theorem 2.4.7 ($G \cong F \otimes H_k$)**.** Let $(F, \langle\cdot,\cdot\rangle_F$ be a real-valued separable Hilbert space and k a continuous real scalar-valued kernel on a compact topological space $\mathcal{X}$. Let G be the F valued RKHS with kernel $\Gamma(x, y) := k(x, y)I_F$. Then $F \otimes H_k$ and G are isometrically isomorphic.

Proof. Define

$$\begin{aligned}&\Upsilon : F\otimes H_k \to G,\\ &\sum_{i=1}^{\infty} f_i\otimes h_i \mapsto \Upsilon(\sum_{i=1}^{\infty} f_i\otimes h_i)(\cdot) := \sum_{i=1}^{\infty}\langle h_i, k(\cdot,-)\rangle_{H_k} f_i = \sum_{i=1}^{\infty} h_i(\cdot) f_i.\end{aligned}$$

The linearity of Υ is clear and the boundedness follows readily from the Cauchy-Schwarz and the fact that $\sum_{i=1}^{\infty}\|f_i\|_F^2\|h_i\|^2_{H_k} < \infty$. The sets $\{f : f \in F\}$ and $\{k(x, \cdot) : x \in \mathcal{X}\}$ are fundamental in F and H_k, respectively. It follows from proposition 2.4.5 that $G_0 := \{f \otimes k(x, \cdot) : f \in F, x \in \mathcal{X}\}$ is fundamental in $F \otimes H_k$. The isometry follows since

$$\begin{aligned}\|\Upsilon(f\otimes k(x,\cdot))\|^2_G &= \langle k(x,\cdot)f, k(x,\cdot)f\rangle_G = \langle E_x^* f, E_x^* f\rangle_G = \langle E_x(E_x^*)f, f\rangle_F,\\ &= \langle k(x,x)I_F f, f\rangle_F = k(x,x)\langle f, f\rangle_F = \langle k(x,\cdot), k(x,\cdot)\rangle_{H_k}\langle f, f\rangle_F,\\ &= \langle f\otimes k(x,\cdot), f\otimes k(x,\cdot)\rangle_{F\otimes H_k},\\ &= \|f\otimes k(x,\cdot)\|^2_{F\otimes H_k}.\end{aligned}$$

The isometry of Υ can be extended via linearity and continuity to $F \otimes H_k$. What is left to show is that Υ is surjective. Consider the closure of the image of G_0

$$\overline{\Upsilon(G_0)} = \overline{\mathrm{span}}\{k(x,\cdot)f \,:\, f\in F,\, x\in\mathcal{X}\} = \overline{\mathrm{span}}\{E_x^* f \,:\, f\in F,\, x\in\mathcal{X}\}.$$

by proposition 2.3.7, it follows that $\overline{\mathrm{span}}\{E_x^* f : f \in F, x \in \mathcal{X}\} = G$, which means that Υ is surjective. We conclude, as Υ is an isometry and bijective. □

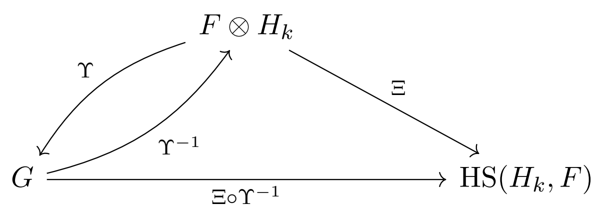

Figure 2.1: Chain of isomorphisms that shows $G \cong_{\Xi\circ\Upsilon^{-1}} \mathrm{HS}(H_k, F)$.

Combining Theorem 2.4.6 and 2.4.7 result in the isometry $G \cong_\Upsilon F \otimes H_k \cong_\Xi HS(H_k, F)$, which schematically can be seen as

Any element in G can be uniquely represented by a Hilbert-Schmidt operator. This viewpoint will be very useful in learning operators in a vector-valued RKHS. It states that any element of the vector-valued RKHS can be thought of as a Hilbert-Schmidt operator evaluated at the canonical feature map. The case where the isomorphism $G \cong HS(H_k, H_k)$ is considered is Corollary 4.5 in [MK20], here we present the extension where the image may be chosen freely as a real-valued separable Hilbert space.

Corollary 2.4.8 ("Operator reproducing property")**.** For every $g \in G$, there exists an operator $A := [\Xi \circ \Upsilon^{-1}](g) \in \mathrm{HS}(H_k, F)$ such that

$$g(x) = A\Big[k(x, \cdot)\Big], \tag{2.20}$$

for all $x \in \mathcal{X}$ and $\|F\|_G = \|A\|_{\mathrm{HS}(H_k,F)}$. Conversely, for any pair $g \in G$ and $A \in \mathrm{HS}(H_k, F)$ satisfying Equation 2.20 we have $A = (\Xi \circ \Upsilon^{-1})(g)$.

Proof. The first claim can be obtained via a direct computation using the isometries Ξ and Υ from Theorems 2.4.6 and 2.4.7, respectively. Since $\overline{\mathrm{span}}\{E_x^* f \,:\, f \in F, x \in \mathcal{X}\} = G$ we may express $g(\cdot) = \sum_{i=1}^\infty k(\mathbf{x}_i, \cdot) f_i$. Applying Υ^{-1} to g, we get $\Upsilon^{-1}(g) = \sum_{i=1}^\infty f_i \otimes k(\mathbf{x}_i, \cdot)$, applying Ξ and evaluating at $k(x, \cdot)$ for some $x \in \mathcal{X}$ results in

$$A\Big[k(x, \cdot]\Big) = \Xi\left[\sum_{i=1}^\infty f_i \otimes k(\mathbf{x}_i, \cdot)\right]\Big(k(x, \cdot)\Big) = \sum_{i=1}^\infty \langle k(x, \cdot), k(\mathbf{x}_i, \cdot)\rangle_{H_k} f_i = \sum_{i=1}^\infty k(\mathbf{x}_i, x) f_i = g(x).$$

Conversely, assume that $B \in HS(H_k, F)$ is any operator satisfying Equation 2.20. Then A and B must be equal as operators, since they agree on the fundamental set $\{k(x, \cdot) \,:\, x \in \mathcal{X}\}$. $\square$

Furthermore, Corollary 2.4.8 enables the explicit construction of the F-valued RKHS G, defined by the kernel $\Gamma(x, y) = k(x, y) I_F$, through the space of Hilbert-Schmidt operators $\mathrm{HS}(H_k, F)$. This construction is established via the chain of isomorphisms shown in Figure 2.1. Specifically, we have:

$$G = \{g : \mathcal{X} \to F \,:\, g(\cdot) = A\Big[k(x, \cdot)\Big],\, A \in \mathrm{HS}(H_k, F)\}.$$

Chapter 3

The Koopman operator on the Banach space $C(K)$

3.1 Preliminaries

3.1.1 Topological aspects

The powerset of a set $\mathcal{X}$, denoted $\mathcal{P}(\mathcal{X})$, is the set containing all subsets of $\mathcal{X}$.

Definition 3.1.1 (Topological space)**.** Let $\mathcal{X}$ be a set. A *topology* $\mathcal{O}$ on $\mathcal{X}$ is a subset $\mathcal{O} \subseteq \mathcal{P}(\mathcal{X})$ satisfying

1. $\emptyset \in \mathcal{O}$ and $\mathcal{X} \in \mathcal{O}$.
2. If $U, V \in \mathcal{O}$, then $U \cap V \in \mathcal{O}$.
3. Let A be an index set, which is a set whose elements are used as indices. Given $U_\alpha \in \mathcal{O}$, then $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{O}$.

Any set $U \in \mathcal{O}$ is called an *open set*. A set C is *closed* if $\mathcal{X} \setminus C$ is open. The pair $(\mathcal{X}, \mathcal{O})$ is called a *topological space*.

Example 3.1.2 (Examples of topologies)**.** Let $\mathcal{X}$ be a set, we state several topologies on can define on $\mathcal{X}$.

- The *discrete topology*, $\mathcal{O} := \mathcal{P}(\mathcal{X})$.
- The *chaotic topology*, $\mathcal{O} := \{\emptyset, \mathcal{X}\}$.
- Let d be a metric on $\mathcal{X}$, and define the open ball of radius $\varepsilon > 0$ centered at $x \in \mathcal{X}$ as $B_\varepsilon(x) := \{y \in \mathcal{X} : d(x, y) < \varepsilon\}$. The *standard topology* is defined as $\mathcal{O}_s := \{U \subseteq \mathcal{X} : \forall x \in U : \exists \varepsilon > 0 : B_\varepsilon(x) \subseteq U\}$.

Example 3.1.3 (Subspace topology)**.** A subset $\mathcal{Y} \subseteq \mathcal{X}$ of a topological space $(\mathcal{X}, \mathcal{O})$ is also called a *subspace*. One can canonically define a topology on $\mathcal{Y}$ called the *subspace topology*, as $\mathcal{O}|_{\mathcal{Y}} := \{U \cap \mathcal{Y} : U \in \mathcal{O}\}$. It is straightforward to verify that $(\mathcal{Y}, \mathcal{O}|_{\mathcal{Y}})$ is a topological space.

Definition 3.1.4 (Hausdorff and normal)**.** A topological space $(\mathcal{X}, \mathcal{O})$ is said to be

- *Hausdorff* if for any $x, y \in \mathcal{X}$, $x \neq y$, there exists disjoint open sets $U, V \in \mathcal{O}$ such that $x \in U$ and $y \in V$.
- *normal* if any disjoint two closed sets can be separated by disjoint open sets. That is, let C, D be disjoint closed sets. Then, there exist disjoint $U, V \in \mathcal{O}$ such that $C \subseteq U$, $D \subseteq V$.

Definition 3.1.5 (Topological convergence)**.** A sequence $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ is said to *converge in a topological space* $(\mathcal{X}, \mathcal{O})$ to some $x \in \mathcal{X}$ if for every $U \in \mathcal{O}$ with $x \in U$ there exists an $N \in \mathbb{N}$ such that for all $n > N$ we have $x_n \in U$.

Example 3.1.6 (Convergence with respect to different topologies)**.** We will discuss convergence of a sequence in topological spaces from Example 3.1.2.

- In the chaotic topology, any sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ converges against any point $x \in \mathcal{X}$. Since $\mathcal{X}$ is the only non-empty open set and $x_n \in \mathcal{X}$ for all $n \in \mathbb{N}$.
- Sequences in the discrete topology only converge if they are eventually constant.
- In the standard topology, a sequence $(x_n)_{n\in\mathbb{N}}$ converges to a limit point x if for any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$.

A desirable property to have is that limit points are unique. This is guaranteed to happen in Hausdorff topological spaces. Suppose on the contrary that a sequence $(x_n)_{n\in\mathbb{N}}$ in a Hausdorff topological space $(\mathcal{X}, \mathcal{O})$ has distinct limit points $x, y \in \mathcal{X}$. By the Hausdorff property, there exists disjoint open sets $U, V \in \mathcal{O}$ that contain x, y, respectively. There exist $N, M \in \mathbb{N}$ such that $x_n \in U$ for all $n > N$ and $x_n \in V$ for all $n > M$. However, $x_n \in U \cap V = \emptyset$ for $n > \max(N, M)$, a contradiction. A standard example of a topological space that is not Hausdorff is $(\mathbb{N}, \mathcal{O}_{\text{co}})$, the natural numbers equipped with the cofinite topology. That is $\mathcal{O}_{\text{co}} := \{U \subseteq \mathbb{N} : U = \emptyset \text{ or } \mathbb{N} \setminus U \text{ is finite}\}$. Suppose the cofinite topology has the Hausdorff property, then for distinct $x, y \in \mathbb{N}$ there exists disjoint open sets U, V that containing x and y, respectively. However, we have $U \subseteq V^{\text{c}}$ which cannot happen as V^{c} is either the empty set or contains finitely many points. A similar argument shows that $(\mathbb{N}, \mathcal{O}_{\text{co}})$ is not normal either. Let A be an index set, $C = (U_\alpha)_{\alpha\in A}$ *covers* $\mathcal{X}$ if $\mathcal{X} \subseteq \bigcup_{\alpha\in A} U_\alpha$.

Definition 3.1.7 (Compactness)**.** A topological space $(\mathcal{X}, \mathcal{O})$ is *compact* if every open cover of $\mathcal{X}$ has a finite subcover. That is, for every collection C of open subsets of $\mathcal{X}$ such that

$$\mathcal{X} \subseteq \bigcup_{U\in C} U$$

there exists a finite collection $I = \{U_1, \ldots, U_n\} \subseteq C$ such that

$$\mathcal{X} \subseteq \bigcup_{U\in I} U.$$

A subset $\mathcal{Y} \subseteq \mathcal{X}$ is compact if it is compact as a subspace with respect to the subspace topology. We call $\mathcal{Y}$ a *compact subset* of $\mathcal{X}$.

Further, we can prove that compact Hausdorff topological spaces are normal, to do so we need a small proposition.

Proposition 3.1.8. Let $(\mathcal{X}, \mathcal{O})$ be a Hausdorff topological space. Let $x \in \mathcal{X}$ and $\mathcal{Y}$ be a compact subset of $\mathcal{X}$. Then there exists disjoint open sets $U, V \in \mathcal{O}$ such that $x \in U$ and $\mathcal{Y} \subseteq V$.

Proof. For every $y \in \mathcal{Y}$, there exists disjoint open sets V_y and $U_{y,x}$ containing y and x, respectively. We have that $\cup_{y\in\mathcal{Y}} V_y$ is an open cover for $\mathcal{Y}$, whereas $U_{y,x}$ contains x for all $y \in \mathcal{Y}$. Since $\mathcal{Y}$ is compact, there exists a finite subcover $\cup_{i=1}^n V_{y_i}$ of $\mathcal{Y}$. Then

$$U := \bigcap_{i=1}^n U_{y_i,x}, \quad V := \bigcup_{i=1}^n V_{y_i}$$

are such that $x \in U$ and $\mathcal{Y} \subseteq V$. Suppose that there exists $z \in U \cap V$, then $z \in U_{y_i,x}$ for all i and $z \in V_j$ for some j, where $1 \leq i, j \leq n$. However, $U_{y_i,x} \cap V_{y_i} = \emptyset$ for all i, and we conclude that U and V are disjoint. $\square$

Let $(\mathcal{X}, \mathcal{O})$ be a compact topological space and let $\mathcal{Y}$ be closed. If C be a cover of $\mathcal{Y}$, then $C \cup \mathcal{Y}^c$ is an open cover of $\mathcal{X}$ and there exists a finite subcover $\{C_1, \dots, C_n\} \cup \mathcal{Y}^c$ of $\mathcal{X}$. Since $F \subseteq \mathcal{X} \subseteq \cup_{i=1}^n C_i \cup \mathcal{Y}^c$ it follows that $\cup_{i=1}^n C_i$ is a finite subcover of $\mathcal{Y}$. Thus, we have shown that a closed subset of a compact topological space is in turn compact with respect to the subspace topology.

Lemma 3.1.9. A compact Hausdorff topological space $(\mathcal{X}, \mathcal{O})$ is normal.

Proof. Let A and B be disjoint closed sets in $\mathcal{X}$. By Proposition 3.1.8 we have that for any $a \in A$ there exists disjoint open sets U_a and V_a such that $a \in U_a$ and $B \subseteq V_a$. Since B is a closed in a compact topological space, it is itself compact. Hence, there exists a finite subcover for $B \subseteq \cup_{i=1}^n V_{a_i}$. Then

$$U := \bigcap_{i=1}^n U_{a_i}, \quad V := \bigcup_{i=1}^n V_{a_i}$$

are disjoint open sets containing A and B, respectively. □

Lemma 3.1.10. Let $(\mathcal{X}, \mathcal{O})$ be a topological space. Then $(\mathcal{X}, \mathcal{O})$ is normal if and only if for any open U and any closed set $C \subseteq U$ there exists an open set V such that

$$C \subseteq V \subseteq \overline{V} \subseteq U.$$

Proof. Assume that $(\mathcal{X}, \mathcal{O})$ is normal and let C, U be a closed and open set, respectively, such that $C \subseteq U$. Since C and U^c are disjoint closed sets, there exists, by assumption, disjoint $V, W \in \mathcal{O}$ such that $C \subseteq V$ and $U^c \subseteq W$. Because V, W are disjoint and $W^c \subseteq U$, we conclude $V \subseteq W^c \subseteq U$. We conclude by observing that W^c is closed, $C \subseteq U$ and taking the closure in $V \subseteq W^c \subseteq U$.
On the other hand, let A, B be disjoint closed sets. Since $A \subseteq B^c$, there exists, by assumption, an open set D such that $A \subseteq D \subseteq \overline{D} \subseteq B^c$. From $\overline{D} \subseteq B^c$ we get $B \subseteq \overline{D}^c$. Since $D \subseteq \overline{D}$ we conclude $D \cap \overline{D}^c = \emptyset$. We have found disjoint open sets $U := D$ and $V := \overline{D}^c$ such that $A \subseteq U$ and $B \subseteq V$. We conclude that $(\mathcal{X}, \mathcal{O})$ is normal.

□

Definition 3.1.11 (Base of a topology)**.** Let $(\mathcal{X}, \mathcal{O})$ be a topological space. A collection of sets $\mathcal{B} \subseteq \mathcal{O}$ is called a *base* if each set in $\mathcal{O}$ can be written as a union of sets in $\mathcal{B}$. That is

$$\forall U \in \mathcal{O} \,:\, \exists I \,:\, U = \bigcup_{i \in I} V_i \quad V_i \in \mathcal{B}.$$

where I is a countable index set.

Definition 3.1.12 (Continuity)**.** Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces. A map $f : M \to N$ is called *continuous* if

$$\forall V \in \mathcal{O}_N : \mathrm{preim}_f(V) \in \mathcal{O}_M,$$

where

$$\mathrm{preim}_f(V) := \{m \in M \,:\, f(m) \in V\}$$

is the preimage of the set V with respect to the map f. One also writes $f^{-1}(V) := \mathrm{preim}_f(V)$.

The set of all $\mathbb{C}$ valued continuous functions on a topological space $(\mathcal{X}, \mathcal{O})$ is denoted by $\mathrm{C}(\mathcal{X})$.

Definition 3.1.13 (Homeomorphism)**.** Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces. A map $f : M \to N$ is called a *homeomorphism* if f is bijective and both f and f^{-1} are continuous.

Proposition 3.1.14 (Preimage of the union is the union of the preimage)**.** Let X, Y be sets and $f : X \to Y$ be a map. Let $(Y_i)_{i \in I}$ be a collection of subsets of Y over a countable index set I. We have

$$\mathrm{preim}_f \left(\bigcup_{i \in I} Y_i \right) = \bigcup_{i \in I} \mathrm{preim}_f \left(Y_i \right)$$

Proof. By definition of the preimage and the union, we get

$$\begin{aligned}\operatorname{preim}_f\left(\bigcup_{i\in I} Y_i\right) &\overset{\text{def}}{=} \left\{x \in X \,:\, f(x) \in \bigcup_{i\in I} Y_i\right\},\\ &= \left\{x \in X \,:\, \exists i \in I \ \text{ s.t. } \ f(x) \in Y_i\right\},\\ &\overset{\text{def}}{=} \bigcup_{i\in I}\left\{x \in X \,:\, f(x) \in Y_i\right\},\\ &\overset{\text{def}}{=} \bigcup_{i\in I} \operatorname{preim}_f(Y_i)\,.\end{aligned}$$

□

We end this section with several statements that we need later on. We do not give proofs of these statements, we refer the reader to any popular textbook on topology.

Proposition 3.1.15. Let $(\Omega, \mathcal{O})$, $(\Omega', \mathcal{O}')$ be a compact spaces. If $f : \Omega \to \Omega'$ is continuous and Ω' is Hausdorff, then $\overline{f(A)} = f(\overline{A})$ is compact for every $A \subseteq \Omega$. If in addition f is bijective, then f is a homeomorphism.

Lemma 3.1.16 (Urysohn)**.** Let $(\mathcal{X}, \mathcal{O})$ be a topological space. Then $(\mathcal{X}, \mathcal{O})$ is normal if and only if for any two non-empty closed disjoint subsets A and B of $\mathcal{X}$, there exists a continuous function $f : \mathcal{X} \to [0, 1]$ such that $f|_A = 0$ and $f|_B = 1$.

Urysohn's lemma tells that $\mathrm{C}(\mathcal{X})$ separates the points of $\mathcal{X}$ since singletons $\{x\}$ are closed.
The original version of Tietze states the result for a normal topological space. The spaces we consider are Hausdorff and compact. In Lemma 3.1.9, we have shown that such spaces are normal and therefore Tietze applies to our context.

Theorem 3.1.17 (Tietze)**.** Let $(\mathcal{X}, \mathcal{O})$ be a normal topological space and $\mathcal{Y} \subseteq \mathcal{X}$ a closed subset. Then, for $f \in \mathrm{C}(\mathcal{Y})$, there is a $g \in \mathrm{C}(\mathcal{X})$ such that $g|_{\mathcal{Y}} = f$.

3.1.2 Algebraic aspects

Definition 3.1.18 (Group)**.** Let S be a set and let $\star : S \times S \to S$ be a map that satisfies the following properties

1. (Associative) $\forall s, t, u \in S \,:\, s \star (t \star u) = (s \star t) \star u$.
2. (Neutral element) $\exists e_N \in S \,:\, \forall s \in S \,:\, e_N \star s = s \star e_N = s$. We call e_N the identity element.
3. (Inverse element) $\forall s \in S \,:\, \exists t \in S \,:\, s \star t = t \star s = e_I$. We denote $t := s^{-1}$, the inverse of s.

Then, the tuple $(S, \star)$ is called a *group*. If in addition the map $\star$ is associative, i.e.

$$\forall s, t \in S \,:\, s \star t = t \star s$$

we call $(S, \star)$ an abelian or commutative group.

Definition 3.1.19 (Field)**.** An *(algebraic) field* is a triple $(\mathbb{K}, \boxplus, \boxdot)$, where $\mathbb{K}$ is a set and $\boxplus, \boxdot$ are maps satisfying $\mathbb{K} \times \mathbb{K} \to \mathbb{K}$ satisfying the following axioms

- $(\mathbb{K}, \boxplus)$ is an abelian group.
- $(\mathbb{K}^*, \boxdot)$, where $\mathbb{K}^* = \mathbb{K} \setminus \{0\}$, is an abelian group.
- The maps $\boxplus$ and $\boxdot$ satisfy the distributive property

1. $\forall a, b, c \in \mathbb{K} : (a \boxplus b) \boxdot c = a \boxdot c \boxplus b \boxdot c$.

Definition 3.1.20 (Vector space)**.** Let $(\mathbb{K}, \boxplus, \boxdot)$ be a field. A $\mathbb{K}$-*vector space, or vector space* over $\mathbb{K}$ is a triple $(V, \oplus, \odot)$, where V is a set and

$$\oplus : V \times V \to V,$$
$$\odot : \mathbb{K} \times V \to V$$

are maps satisfying

1. $(V, \oplus)$ is an abelian group.
2. The map $\odot$ acts on $(V, \oplus)$:
   - $\forall \lambda \in \mathbb{K} : \forall v, w \in V : \lambda \boxdot (v \oplus w) = \lambda \boxdot v \oplus \lambda \boxdot w$;
   - $\forall \lambda, \mu \in \mathbb{K} : \forall v \in V : (\lambda \boxplus \mu) \odot v = \lambda \boxdot v \oplus \mu \boxdot v$;
   - $\forall \lambda, \mu \in \mathbb{K} : \forall v \in V : (\lambda \boxdot \mu) \odot v = \lambda \odot (\mu \odot v)$;
   - $\forall v \in V : 1 \odot v = v$, where 1 denotes the unit element of the abelian group $(\mathbb{K} \setminus \{0\}, \boxdot)$.

Example 3.1.21 ($\mathbb{C}$ as a vector space over $\mathbb{R}$)**.** The complex numbers $\mathbb{C}$ together with the standard addition and multiplication can be viewed as a vector space over the real numbers $\mathbb{R}$. To be complete, we write $(\mathbb{C}, +, \cdot)$, where $+ : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$.

Example 3.1.22 ($\mathrm{C}(\mathcal{X})$ as a vector space over $\mathbb{C}$)**.** Let $(\mathcal{X}, \mathcal{O})$ be a compact Hausdorff topological space. With $\mathrm{C}(\mathcal{X})$ we denote all continuous $\mathbb{C}$-valued functions on $\mathcal{X}$. Together with the standard pointwise addition and multiplication, the triple $(\mathrm{C}(\mathcal{X}), \oplus, \odot)$ is a vector space over $\mathbb{C}$. That is, for any $f, g \in \mathrm{C}(K)$, $\lambda \in \mathbb{C}$ and $x \in \mathcal{X}$ we have,

$$(f \oplus g)(x) := f(x) + g(x), \quad (\lambda \odot f)(x) := \lambda \cdot f(x).$$

Definition 3.1.23 (Vector subspace)**.** Let $(V, \oplus, \odot)$ be a vector space over $\mathbb{K}$ and let $U \subseteq V$ be nonempty. The triple $(U, \oplus|_{U \times U}, \odot|_{\mathbb{K} \times U})$ is a *vector subspace* if it is a vector space itself.

We will define an algebra as a unital algebra. All the algebras we will encounter will be algebras that contain a unit element.

Definition 3.1.24 (Algebra)**.** Let $(V, \oplus, \odot)$ be a vector space. Let $\cdot : V \times V \to V$ be a bilinear map. The quadruple $(V, \oplus, \odot, \cdot)$ is called an *algebra*. The bilinear map is called multiplication and a distinguished element $e \in A$ satisfying,

$$v \cdot e = e \cdot v = v, \quad \forall v \in V$$

is called the unit element. If the multiplication is commutative, $v \cdot u = u \cdot v$ for all $u, v \in V$, we call $(V, \oplus, \odot, \cdot)$ a commutative (unital) algebra.

Example 3.1.25 ($\mathrm{C}(\mathcal{X})$ as an algebra)**.** Recall Example 3.1.22 and add $\bigodot : \mathrm{C}(\mathcal{X}) \times \mathrm{C}(\mathcal{X}) \to \mathrm{C}(\mathcal{X})$, which is defined as the pointwise multiplication. That is, for $f, g \in \mathrm{C}(\mathcal{X})$ and $x \in K$,

$$(f \bigodot g)(x) := f(x) \cdot g(x).$$

The quadruple $(\mathrm{C}(\mathcal{X}), \oplus, \odot, \bigodot)$ is an algebra.

From an abstract point of view, we can view $\mathbb{C}$ as an algebra. This is overkill, but it does allow for good generalizations later on.

Example 3.1.26 ($\mathbb{C}$ as an algebra)**.** Recall that from Example 3.1.21 that we can view $\mathbb{C}$ as a vector space over $\mathbb{R}$. Now, we will make a distinction between $\cdot_{\mathbb{R}} : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$ and $\cdot_{\mathbb{C}} : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, which denotes the standard multiplication between the complex numbers. Then the quadruple $(\mathbb{C}, +, \cdot_{\mathbb{R}}, \cdot_{\mathbb{C}})$ is a (commutative) algebra.

Definition 3.1.27 (Algebra homomorphism)**.** A linear map $T : A \to B$ between two algebras $(A, \oplus_A, \odot_A, \cdot_A)$, $(B, \oplus_B, \odot_B, \cdot_B)$ is called *multiplicative* if

$$T(a \cdot_A b) = T(a) \cdot_B T(b), \quad \forall a, b \in A$$

and an *algebra homomorphism* if in addition $T(e_A) = e_B$.

Definition 3.1.28 (Algebra isomorphism)**.** An *algebra homomorphism* $T : A \to B$ is an algebra isomorphism if T is bijective.

Definition 3.1.29 (Subalgebra)**.** Let $(V, \oplus, \odot)$ be a vector space, and $\eta : V \times V \to V$ be a map. The quadruple $(V, \oplus, \odot, \eta)$ is called a *subalgebra*.

Definition 3.1.30 (Banach algebra)**.** Let $(V, \oplus, \odot, \cdot)$ be an algebra and $\|\cdot\|$ be a norm on V such that $(V, \oplus, \odot, \cdot, \|\cdot\|)$ is a Banach space. If

$$\forall a, b \in V \,:\, \|a \cdot b\| \leq \|a\| \|b\|$$

we call the quintuple $(V, \oplus, \odot, \cdot, \|\cdot\|)$ a *Banach algebra*.

Example 3.1.31 ($\mathrm{C}(\mathcal{X})$ as a Banach algebra)**.** One can define the supremum norm $\|\cdot\|_\infty$ on $\mathrm{C}(\mathcal{X})$, defined as

$$\|f\|_\infty = \sup_{x \in K} |f(x)|$$

for $f \in \mathrm{C}(\mathcal{X})$. Adding the supremum norm to the algebra from Example 3.1.25 results in a Banach algebra.

Definition 3.1.32 (Ideal)**.** Let $(V, \oplus, \odot, \cdot, \|\cdot\|)$ be a commutative Banach algebra. An *(algebra) ideal* of V is a vector subspace $I \subseteq V$ satisfying

$$f \in I,\, g \in V \quad \Longrightarrow \quad f \cdot g \in I.$$

We say that I is *closed* if it is closed with respect to the norm $\|\cdot\|$.

Observation 3.1.33. Since multiplication is continuous, the closure of an ideal is still an ideal.

Definition 3.1.34 (Proper)**.** Let I be an ideal of a commutative Banach algebra $(V, \oplus, \odot, \cdot, \|\cdot\|)$. The ideal I is called *proper* if $I \neq V$.

Lemma 3.1.35. An ideal I of a commutative Banach algebra $(V, \oplus, \odot, \cdot, \|\cdot\|)$ is proper if and only if $e \notin I$.

Proof. Suppose that I is a proper ideal of V and that $e \in I$. Take any $f \in V$, then $e \cdot f = f \in I$ which shows that $I = V$, a contradiction. On the other hand, let I be an ideal of V so that $e \notin I$, then it is clear that $I \neq V$. □

Definition 3.1.36 (Maximal)**.** A proper ideal I of a commutative Banach algebra $(V, \oplus, \odot, \cdot, \|\cdot\|)$ is called *maximal* if

$$I \subseteq J \subseteq V \quad \Longrightarrow \quad J = I \text{ or } J = V.$$

Definition 3.1.37 (Linear map)**.** Let $(A, \oplus_A, \odot_A)$, $(B, \oplus_B, \odot_B)$, be vector spaces over the same field $\mathbb{K}$, and $\xi : A \to B$ a map. The map ξ is called *linear* if

$$\forall \lambda \in \mathbb{K} \,:\, \forall a, b \in A \,:\, \xi(\lambda \odot_A a \oplus_A b) = \lambda \odot_B \xi(a) \oplus_B \xi(b).$$

Definition 3.1.38 (Involution)**.** An *involution* on a complex (the field over which is vector space is is $\mathbb{C}$) Banach algebra $(V, \oplus, \odot, \cdot, \|\cdot\|)$ is a map

$$* : V \to V, \tag{3.1}$$
$$x \mapsto *(x) := x^* \tag{3.2}$$

satisfying

$$(x^*)^* = x, \quad (x+y)^* = x^* + y^*, \quad (\lambda x)^* = \overline{\lambda} x^*, \quad (xy)^* = y^* x^*$$

for all $x, y \in V$ and $\lambda \in \mathbb{C}$.

Definition 3.1.39 (C^*-algebra)**.** Let $(V, \oplus, \odot, \cdot, \|\cdot\|)$ be a Banach algebra over $\mathbb{C}$ and $*$ be an involution defined on V. If

$$\|x^* \cdot x\| = \|x\|^2 \quad \forall x \in V,$$

we call the sextuple $(V, \oplus, \odot, \cdot, \|\cdot\|, *)$ a C^*-*algebra*.

It follows that in a C^*-algebra $\|x^*\| = \|x\|$ and $\|e\| = 1$.

Example 3.1.40 ($\mathrm{C}(\mathcal{X})$ as a C^*-algebra)**.** Observe that taking elementwise complex conjugates is an involution. Adding this operation to the Banach algebra of Example 3.1.31 results in a C^*-algebra.

Now we will make a remark on the notation. From now on, we will suppress the long notation for algebras, vector spaces and so on and instead just write the name of the set. That is, with "The Banach algebra A is ...", we actually mean "The Banach algebra $(A, \oplus, \odot, \cdot, \|\cdot\|)$ is ...". The explicit use of $\odot$ and $\cdot$ is clear from the context, a difference between these operations will be made when necessary.

3.2 The C^*-algebra $\mathrm{C}(K)$ and the Koopman operator

For notational convenience, we shall from now use K to be a compact Hausdorff topological space, instead of $(\mathcal{X}, \mathcal{O})$. In this chapter we will define a dynamical system and lay out its relation to the Koopman operator when acting on $\mathrm{C}(K)$.

Definition 3.2.1 (Topological dynamical system)**.** Let $(\mathcal{X}, \mathcal{O})$ be a compact topological space and $\varphi : \mathcal{X} \to \varphi(\mathcal{X}) \subseteq \mathcal{X}$ be continuous. The pair $(\mathcal{X}, \varphi)$ is called a topological (dynamical) system (TDS). A topological system is surjective if φ is surjective, and the system is invertible if φ is invertible, i.e., a homeomorphism.

Instead of studying $\varphi : K \to K$, we study its Koopman operator $T := T_\varphi$ defined by

$$T_\varphi f := f \circ \varphi$$

for $f \in \mathrm{C}(K)$. Thus, the Koopman operator is nothing more than the composition operator between f and φ.
The Koopman operator T commutes with each operation defined on the algebra $\mathrm{C}(K)$. That is, for any $f, g \in \mathrm{C}(K)$ and $\lambda \in \mathbb{C}$ we have

$$T(f+g) = Tf + Tg, \quad T(\lambda f) = \lambda(Tf),$$
$$T(fg) = (Tf)(Tg), \quad \overline{Tf} = T\overline{f}, \quad |Tf| = T|f|.$$

Since φ is continuous, the Koopman operator is continuous if f is continuous. The goal of this chapter is to show that the Koopman operator T_φ contains all information about φ. In other words, we can uniquely determine φ from given a Koopman operator. To do so, we need to study $\mathrm{C}(K)$ as a C^*-algebra.

3.2.1 The space $\mathrm{C}(K)$ as a Commutative C^*-Algebra

In certain applications, it suffices to view $\mathrm{C}(K)$ as a vector space with pointwise addition and scalar multiplication. We have seen from Example 3.1.40 that $\mathrm{C}(K)$ can be viewed as a C^*-algebra. The key point of viewing $\mathrm{C}(K)$ as a C^*-algebra, is that we can define ideals. We will show that the closed ideals of $\mathrm{C}(K)$ can easily be characterized. The maximal ideals can easily be spotted.
For a closed subset $F \subseteq K$ we define

$$I_F := \{f \in \mathrm{C}(K) \,:\, f \equiv 0 \text{ on } F\}.$$

It is clear that I_F is an ideal, since

$$\forall x \in K \,:\, \forall f \in I_F \,:\, \forall g \in \mathrm{C}(K) \,:\, (f \cdot g)(x) = f(x) \cdot g(x) = 0.$$

The one function $\mathbf{1}_K \in I_F$ if and only if F is empty, hence I_F is a proper ideal if and only if $F \neq \emptyset$. Furthermore, I_F is a closed ideal. To see this, let $(f_n)_{n\in\mathbb{N}}$ be a sequence in I_F converging to $f \in \mathrm{C}(K)$ and $x \in F$. By the triangle inequality

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \|f - f_n\|_\infty - 0 \to 0 \quad (n \to \infty),$$

which shows f is zero on F.
Next is an interesting result. It states that for all closed ideals of $\mathrm{C}(K)$, there exists a closed set $F \subseteq K$ such that $I = I_F$. Thus, *if* we are have a closed ideal of $\mathrm{C}(K)$, we know what it "looks" like. The outline of the proof is as follows. It easily follows that $I \subseteq I_F$. To show that $I_F \subseteq I$ we apply the following strategy. We choose an $f \in I_F$, construct a sequence $(g_n)_{n\in\mathbb{N}}$ in I and show that $g_n \to f$ as $n \to \infty$. Since I is closed, it must be that $f \in I$ which implies that $I_F \subseteq I$, which completes the proof.

Lemma 3.2.2. Let $I \subseteq \mathrm{C}(K)$ be a closed algebra ideal. Then, there is a closed subset $F \subseteq K$ such that $I = I_F$.

Proof. Define

$$F := \{x \in K \,:\, f(x) = 0 \text{ for all } f \in I\} = \bigcap_{f\in I}[f = 0] = \bigcap_{g\in I}\{x \in K \,:\, f(x) = 0\}.$$

It can be shown that F is closed, since the functions in $I \subseteq \mathrm{C}(K)$ are continuous. We have $I \subseteq I_F$, as I_F contains *all* functions $f \in \mathrm{C}(K)$ that are zero on F. Fix $f \in I_F$, $\varepsilon > 0$ and define $F_\varepsilon := [|f| \geq \varepsilon]$. We have $F \cap F_\varepsilon = \emptyset$ and therefore $F_\varepsilon \subseteq F^{\mathrm{c}}$. Let $(x_n)_{n\in\mathbb{N}} \subseteq F_\varepsilon$ be a sequence converging to $x \in K$. By the continuity of f

$$|f(x)| = \lim_{n\to\infty} |f(x_n)| > \lim_{n\to\infty} \varepsilon = \varepsilon,$$

which shows that $x \in F_\varepsilon$. Hence F_ε is a closed subset of a compact set K and thus compact. For each $x \in F_\varepsilon$, we can find $f_x \in I$ such that $f_x(x) \neq 0$. Since $\mathrm{C}(K)$ is conjugation invariant, we have $\overline{f_x} \in \mathrm{C}(K)$ and $f_x \cdot \overline{f_x} \in I$, as I is an ideal. By multiplying with a suitable constant, we may assume w.l.o.g., that $f_x \geq 0$ and $f_x(x) > 1$. We have the following open cover

$$F_\varepsilon \subseteq \bigcup_{x\in F_\varepsilon} \{y \in K \,:\, f_x(y) > 1\} = \bigcup_{x\in F_\varepsilon} [f_x > 1].$$

By the compactness of F_ε, there exists a finite subcover $F_\varepsilon \subseteq [f_1 > 1] \cup \ldots \cup [f_k > 1]$. Define $g := f_1 + \ldots + f_k \in I$. Since each $f_i \geq 0$, we have that $g \geq 0$. For any $x \in F_\varepsilon$, we have $x \in [f_j > 1]$ for $1 \leq j \leq k$. Therefore, we have the inclusion $F_\varepsilon = [|f| \geq \varepsilon] \subseteq [g \geq 1]$[1]. Define

$$g_n := \frac{fn}{1 + ng} \cdot g.$$

[1] This is a trivial inclusion.

We have $g_n \in I$, since $g \in I$, $fn \in \mathrm{C}(K)$ and $(1+ng)^{-1} \in \mathrm{C}(K)$, as $(1+ng)$ has no zeroes. Then,

$$|g_n - f| = |\frac{nfg}{1+ng} - \frac{f - fng}{1+ng}| = \frac{|f|}{1+ng} \leq \max\left\{\varepsilon, \frac{\|f\|_\infty}{1+n}\right\}.$$

Hence, $\|f - g_n\|_\infty \leq \varepsilon$ as $n \to \infty$. Because ε was arbitrary and I is closed, $f \in I$ and $I_F \subseteq I$. Since $I \subseteq I_F$ and $I_F \subseteq I$, we conclude that $I = I_F$. ☐

Using Lemma 3.2.2, we can show that the maximal ideals of $\mathrm{C}(K)$ have a special structure.

Lemma 3.2.3. An ideal I of $\mathrm{C}(K)$ is maximal if and only if $I = I_{\{x\}}$ for some $x \in K$.

Proof. Suppose that I is a maximal ideal. We claim that it suffices to show that I is closed. By Lemma 3.2.2, there exists a closed set F such that $I = I_F$. Suppose there exists a closed set $H \subset F$, then $I_F \subset I_H \subseteq \mathrm{C}(K)$ and I_F would not be maximal. Hence, such set H cannot exist and we conclude that $F = \{x\}$ for some $x \in K$. If I is closed, then $I = \overline{I}$ and we are done. The other case is that $\overline{I} = \mathrm{C}(K)$, then we can find a sequence $(f_n)_{n\in\mathbb{N}}$ in I that converges to $\mathbf{1} \in \mathrm{C}(K)$. Therefore, we can find a function f such that $\|\mathbf{1} - f\|_\infty < \varepsilon$. Then, $f = \mathbf{1} - (\mathbf{1} - f)$ has no zeroes and $\frac{1}{f} \in \mathrm{C}(K)$. Since I is an ideal, $f \cdot \frac{1}{f} = \mathbf{1} \in I$, from Lemma 3.1.35 we conclude that $I = \mathrm{C}(K)$. This is a contradiction, as we assumed that I is maximal. Therefore, I is closed and the assertion follows as discussed above.
Now suppose that $I = I_{\{x\}}$ for some $x \in K$, and suppose that $I_{\{x\}}$ is not maximal. Then there exists an ideal J such that

$$I_{\{x\}} \subseteq J \subseteq \mathrm{C}(K)$$

such that $I \neq J$ and $\mathrm{C}(K) \neq J$. Take any $f \in J$ with $f \notin I_{\{x\}}$. It must be that $f(x) \neq 0$ for all $x \in K$. Then, since f has no zeroes, $\frac{1}{f} \in \mathrm{C}(K)$ and $\mathbf{1} = f\frac{1}{f} \in J$. However, since J contains the identity, we have that $J = \mathrm{C}(K)$, a contradiction. We conclude that $I_{\{x\}}$ is maximal. ☐

Using Lemmas 3.2.3 and 3.2.2, we can show that the linear functionals on $\mathrm{C}(K)$ are multiplicative if and only if they equal some evaluation functional.

Lemma 3.2.4. A nonzero linear functional $\psi : \mathrm{C}(K) \to \mathbb{C}$ is multiplicative if and only if $\psi = \delta_x$ for some $x \in K$

Proof. If $\psi = \delta_x$ for some $x \in K$, it is clear that ψ then is multiplicative. Suppose that $\gamma : \mathrm{C}(K) \to \mathbb{C}$ is a nonzero multiplicative linear functional. Take $f \in \mathrm{C}(K)$, then $\gamma(f) = a$ for some $a \in \mathbb{C}$. By the linearity of γ, we see that $f/a \in \mathrm{C}(K)$ is such that $\gamma(f/a) = 1$. After rescaling, we may assume that $\gamma(f) = 1$. We see that

$$1 = \gamma(f) = \gamma(f \cdot \mathbf{1}) \overset{\text{multipl.}}{=} \gamma(f)\gamma(\mathbf{1}) = \gamma(\mathbf{1}).$$

The identity element of $\mathrm{C}(K)$, $\mathbf{1}$, is mapped to the identity element of $\mathbb{C}$, 1. Therefore, γ is an algebra homomorphism. Define $I = \ker(\gamma)$, it follows from the multiplicativity of γ that I is an ideal,

$$\forall g \in I \,:\, \forall h \in \mathrm{C}(K) \,:\, \gamma(g \cdot h) \overset{\text{multipl.}}{=} \gamma(g)\gamma(h) = 0\gamma(h) = 0 \quad \Longrightarrow \quad g \cdot h \in I.$$

Since γ is a nonzero mapping, I is a proper ideal. Moreover, we claim that I is maximal. Suppose it is not, then there exists an ideal J such that $I \subseteq J \subseteq \mathrm{C}(K)$ and $J \neq I$ and $J \neq \mathrm{C}(K)$. Take $g \in J$ such that $g \notin I$, thus $\gamma(g) \neq 0$. Then we may assume that $\gamma(g) = c \in \mathbb{C}$. By linearity, we may assume that $\gamma(g) = 1$ and it follows that $g = \mathbf{1}$. Since $\mathbf{1} \in J$ we have $J = \mathrm{C}(K)$, a contradiction. Therefore, I is maximal.
By Lemma 3.2.3, there exists an $x \in K$ such that $\ker(\gamma) = I_{\{x\}} = \{x \in K \,:\, f(x) = 0\} = \ker(\delta_x)$. Since $\gamma(f) = 1$ and $\gamma(\mathbf{1}) = 1$, we have

$$0 = \gamma(f) - \gamma(f) = \gamma(f) - \gamma(f)\gamma(\mathbf{1}) \overset{\text{lin.}}{=} \gamma(f - \gamma(f)\mathbf{1}).$$

Therefore, $f - \gamma(f)\mathbf{1} \in \ker(\gamma) = \ker(\delta_x)$. Since $f - \gamma(f)\mathbf{1} \in \ker(\delta_x)$, we have $\gamma(f) = f(x) = \delta_x(f)\ \forall f \in \mathrm{C}(K)$. Since f was arbitrary, we conclude that there exists an $x \in K$ such that $\gamma = \delta_x$. ☐

Theorem 3.2.5 (Collection of results)**.** Let K be a compact space, and let

$$\Gamma(\mathrm{C}(K)) := \left\{\gamma \in \mathrm{C}(K)' \,:\, \gamma \text{ algebra homomorphism}\right\}.$$

Then the map

$$\begin{aligned}\delta : K &\to \Gamma(\mathrm{C}(K)),\\ x &\mapsto \delta_x\end{aligned}$$

is a homeomorphism, where $\Gamma(\mathrm{C}(K))$ is endowed with the weak*-topology as a subset of $\mathrm{C}(K)'$.

Proof. Since the weak*-topology is Hausdorff, it follows from Proposition 3.1.15 that δ is a homeomorphism if it is continuous and bijective. Since $\Gamma(\mathrm{C}(K))$ is equipped with the weak*-topology it follows by definition that δ is continuous. Take any $x, y \in K$ and suppose that $\delta_x = \delta_y$, which means that

$$\forall f \in \mathrm{C}(K) \,:\, f(x) = \delta_x(f) = \delta_y(f) = f(y).$$

Since $\mathrm{C}(K)$ separates the points of K, there exists a function $h \in \mathrm{C}(K)$ such that $h(x) \neq h(y)$, which implies that $x = y$. We conclude that δ is injective. Suppose that $\gamma \in \Gamma(\mathrm{C}(K))$ is nonzero, by Lemma 3.2.4, it follows that there exists an $x \in K$ such that $\gamma = \delta_x$. Therefore, the map $\delta : K \to \Gamma(\mathrm{C}(K))$ is surjective. We have shown that δ is injective and surjective, hence a homeomorphism.

□

3.2.2 Koopman Operator

We return to the original setting of a topological dynamical system (K, φ) with its Koopman operator $T = T_\varphi$ defined as

$$\begin{aligned}T_\varphi : \mathrm{C}(K) &\to \mathrm{C}(K),\\ f &\mapsto T_\varphi f = f \circ \varphi.\end{aligned}$$

But first, we will look at a slightly more general case, for a map $\varphi : L \to K$, where we only assume that K is compact and L is any topological space. We will show that φ is continuous if and only if $f \circ \varphi \in \mathrm{C}(L)$ for any $f \in \mathrm{C}(K)$, i.e. $T_\varphi(\mathrm{C}(K)) \subseteq \mathrm{C}(L)$. One implication is clear, the idea behind the other one is to show that $\bigcup_{h \in I} \mathrm{preim}_{|h|}(V)$ form a base of the topology of K for some countable collection of continuous functions I.

Lemma 3.2.6. Let K be a compact space, Ω a topological space, and let $\varphi : \Omega \to K$ be a map. Then φ is continuous if and only if $f \circ \varphi$ is continuous for all $f \in \mathrm{C}(K)$.

Proof. Suppose that $f \in \mathrm{C}(K)$ and φ is continuous, then $f \circ \varphi$ is also continuous. On the other hand, suppose that $f \circ \varphi$ is continuous. Assume that $\mathbb{C}$ is equipped with the standard topology, then $V = (0, \infty)$, once canonically embedded into $\mathbb{C}$, is open. The map $|f \circ \varphi|$ is continuous since $f \circ \varphi$ is continuous, therefore the set $\mathrm{preim}_{|f \circ \varphi|}(V)$ is open in Ω for every $f \in \mathrm{C}(K)$. We may rewrite

$$\begin{aligned}\mathrm{preim}_{|f\circ\varphi|}(V) &= \{y \in \Omega \,:\, |(f \circ \varphi)(y)| \in V\},\\ &= \left\{y \in \Omega \,:\, \varphi \in \mathrm{preim}_{|f|}(V)\right\},\\ &\overset{\text{def.}}{=} \mathrm{preim}_\varphi\left(\mathrm{preim}_{|f|}(V)\right),\end{aligned}$$

and note that this set is open in Ω for any $f \in \mathrm{C}(K)$. Suppose that the sets of the form $\mathrm{preim}_{|f|}(V)$ form a base for the topology on K. This means that for any open set U of K, we can find a countable collection of functions I such that $U = \bigcup_{h \in I} \mathrm{preim}_{|h|}(V)$. From Lemma 3.1.14, we conclude

$$\mathrm{preim}_\varphi(U) = \mathrm{preim}_\varphi\left(\bigcup_{h\in I} \mathrm{preim}_{|h|}(V)\right) = \bigcup_{h \in I} \underbrace{\mathrm{preim}_\varphi\left(\mathrm{preim}_{|h|}(V)\right)}_{\text{open in } \Omega}.$$

A countable union of open sets is again open Ω. Since U was an arbitrary open set in K, we conclude that φ is continuous since we showed that $\mathrm{preim}_\varphi(U)$ is open in Ω. Next we will construct an open cover for an open set $U \subseteq K$ from sets of the form $\mathrm{preim}_{|f|}(V)$.
Let $V = (0, \infty)$, as before, from the continuity of f it follows that $\mathrm{preim}_{|f|}(V)$ is open in K. Let $U \subseteq K$ be a nonempty open set and choose $x \in U$. The sets $\{x\}$ and U^c are closed and disjoint, it follows from Urysohn that there exists a function $f \in \mathrm{C}(K)$ such that

$$f : K \to [0,1], \quad f\big|_{\{x\}} = 1, \quad f\big|_{U^c} = 0.$$

Since $f(x) = 1$, we have $x \in \mathrm{preim}_{|f|}(V) \subseteq U$. Hence, for every $x \in U$, we can find a function h_x, that depends on x, such that $x \in \mathrm{preim}_{h_x}(V) \subseteq U$. Then, we have the open cover

$$U \subseteq \bigcup_{x \in U} \mathrm{preim}_{h_x}(V).$$

Because each $\mathrm{preim}_{h_x}(V) \subseteq U$ we conclude that $U = \bigcup_{x \in U} \mathrm{preim}_{h_x}(V)$. Since U is arbitrary we have found a base for the topology on K.

□

We now return to when K and L are compact spaces. If φ is continuous, we are ensured that $T_\varphi f = f \circ \varphi$ maps into $\mathrm{C}(L)$. Hence the operator T_φ is an algebra homomorphism between $\mathrm{C}(K)$ and $\mathrm{C}(L)$ when φ is continuous. The next result states that **every** algebra homomorphism between $\mathrm{C}(K)$ and $\mathrm{C}(L)$ is such a Koopman operator. In other words, for any algebra homomorphism $T : \mathrm{C}(K) \to \mathrm{C}(L)$, there exists a continuous function $\varphi : L \to K$ such that $T = T_\varphi$. This is stated in the following theorem.

Theorem 3.2.7. Let K, L be (nonempty) compact spaces and let $T : \mathrm{C}(K) \to \mathrm{C}(L)$ be linear. Then, the following assertions are equivalent:

(i) T is an algebra homomorphism.

(ii) There is a continuous mapping $\varphi : L \to K$ such that $T = T_\varphi$.

In this case, φ in (ii) is uniquely determined and the operator has norm $\|T\| = 1$.

Proof. Let $T : \mathrm{C}(K) \to \mathrm{C}(L)$ be an algebra homomorphism. Let $y \in L$ and define

$$\eta_y := \delta_y \circ T : \mathrm{C}(K) \to \mathbb{C}, \qquad f \mapsto (\delta_y \circ T)(f) = \delta_y(T(f)) = (Tf)(y).$$

By a straightforward argument, it can be seen that η_y is an algebra homomorphism. By Theorem 3.2.5,

$$\forall y \in L \, : \, \exists!\, x_y \in K \, : \, \eta_y(f) = \delta_{x_y}(f).$$

We write $x_y \in K$, to stress the dependency on the choice of $y \in L$. The uniqueness of this dependency, which we call $\varphi : L \to K$, follows from the fact that δ from Theorem 3.2.5 is a homeomorphism. To conclude (ii), we have show is that $Tf = f \circ \varphi$ and that φ is continuous. Apply an arbitrary $f \in \mathrm{C}(K)$ to both sides of $\eta_y = \delta_{\varphi(y)}$, we get $(Tf)(y) = (f \circ \varphi)(y)$. Since $y \in L$ was arbitrary, we conclude that $Tf = f \circ \varphi$. Recall that $T : \mathrm{C}(K) \to \mathrm{C}(L)$, so that $f \circ \varphi = Tf \in \mathrm{C}(L)$ is continuous for any $f \in \mathrm{C}(K)$. By Lemma 3.2.6, we conclude that φ is continuous and we have shown that (i) implies (ii).
Conversely, let $T : \mathrm{C}(K) \to \mathrm{C}(L)$ be linear and suppose that there exists a continuous map $\varphi : L \to K$ such that $T = T_\varphi$. By definition of the Koopman operator, we have

$$\forall f, g \in \mathrm{C}(K) \, : \, T(fg) = T_\varphi(fg) = T_\varphi(f) T_\varphi(g) = T(f)T(g).$$

The one function on K, $\mathbf{1}_K$, gets mapped to $T\mathbf{1}_K = T_\varphi \mathbf{1}_K = \mathbf{1}_K \circ \varphi = \mathbf{1}_L$. We conclude that T is an algebra homomorphism. Let $f \in \mathrm{C}(K)$, then

$$\|Tf\| = \|T_\varphi f\|_\infty = \|f \circ \varphi\| = \sup_{y \in L} |(f \circ \varphi)(y)| \leq \sup_{x \in K} |f(x)| = \|f\|_\infty,$$

and we have $\|T\| \leq 1$. The lower bound is obtained by choosing $f = \mathbf{1}_K$,

$$\|T\| \geq \|T\mathbf{1}_K\|_\infty = \|\mathbf{1}_K \circ \varphi\|_\infty = 1.$$

□

The implications from Theorem 3.2.7 are important, as we can study something that is highly nonlinear, φ, and translate it to something that is linear, T_φ. We have shown that there is a 1-1 correspondence between these objects. Linear operators can be studied with our tools from linear analysis.
We end this chapter with an example of the Koopman operator acting on a permutation map of a finite dimensional space.

Example 3.2.8 (Permutations on a finite dimensional space)**.** Let $K = \{1, 2, \ldots, n\}$ with $n \in \mathbb{N}$ and equip K with the discrete topology, i.e. $\mathcal{O} = \mathcal{P}(K)$, so that we have a compact Hausdorff topological space. Let $\varphi : K \to K$ be a permutation on K, which is continuous since every map is continuous in the discrete topology. Let $f \in \mathrm{C}(K)$, $i \in K$ and let the Koopman operator $T_\varphi : \mathrm{C}(K) \to \mathrm{C}(K)$ act on f

$$(T_\varphi f)(i) = (f \circ \varphi)(i) = f(\varphi(i)).$$

We observe that the Koopman operator T_φ is the permutation matrix induced by the permutation φ.

Chapter 4

The Koopman operator between RKHS

This chapter aims to provide a general framework in which we compare several recent developments in the field of estimating the Koopman operator from data using the theory of RKHS. In Section 4.1, we explain the Koopman paradigm for estimating a dynamical system through observed snapshot data. We introduce the general Dynamic Mode Decomposition (DMD) and conclude with kernel Extended Dynamic Mode Decomposition (kEDMD) in Section 4.2. We cover a recent method, kernel ridge regression, which is used to approximate the Koopman operator and compare it with kEDMD. Lastly, we discuss the boundedness of the Koopman operator and limit the dynamics for which this is feasible for specific RKHSs.

4.1 The Koopman operator and dynamical systems

The setting of this chapter is as follows, the goal is to approximate an unknown topological dynamical system $(\varphi, \mathcal{X})$. Recall that this means that φ is a continuous map between a compact topological space $\mathcal{X}$. For simplicity we assume that $\mathcal{X} \subset \mathbb{R}^d$. We are given observed snapshot data

$$\{\mathbf{x}_i, \mathbf{y}_i = \varphi(\mathbf{x}_i)\}_{i=1}^N$$

such that $\varphi(\mathbf{x}_i) = y_i$ for $1 \leq i \leq N$. The map φ encapsulates the evolution of the data. We will say that $\varphi(\mathbf{x}_i) = y_i$ is an observed discrete time dynamical system. We will store these data in matrices,

$$\mathbb{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \quad \text{and} \quad \mathbb{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}^{d \times N}. \tag{4.1}$$

Instead of studying the observed data directly, we will study measurements of these states. Koopman operators are precisely the tool that allows us to do so. To define the Koopman operator we need a dynamical system $(\mathcal{X}, \varphi)$ and a Banach space of functions $\mathcal{F} : \mathcal{X} \to \mathbb{R}$ so that we can define the Koopman operator T on a domain $\mathcal{D}(T_\varphi) \subset \mathcal{F}$ such that,

$$\begin{aligned} T_\varphi : \mathcal{D}(T_\varphi) &\to \mathcal{F}, \\ g(\cdot) &\mapsto [T_\varphi(g)](\cdot) := (g \circ \varphi)(\cdot) = g(\varphi(\cdot)). \end{aligned}$$

We assume that the Koopman operator is bounded, an assumption we shall make some remarks on in Section 4.4. Elements of this Banach space $\mathcal{F}$ are called observables, so that $g(x)$ is a measure of the state $x \in \mathcal{X}$, for some $g \in \mathcal{F}$. The underlying dynamical system from which we have observed snapshot data does not define the Koopman operator uniquely. Rather, it depends on the choice of observables $\mathcal{F}$ together with the dynamical system. A canonical choice in the literature for $\mathcal{F}$ is the Hilbert space $\mathrm{L}^2(\mathcal{X}, \omega)$ for some measure ω. The measure ω is positive and the space is equipped with the standard inner product $\langle g_1, g_2 \rangle = \int_{\mathcal{X}} g_1(x)\overline{g_2(x)} \mathrm{d}\omega(x)$. To ensure that the Koopman operator is well defined on $\mathrm{L}^2(\mathcal{X}, \omega)$, we need to check that T_φ does not depend on the chosen representative in the equivalence classes of functions that are equal ω-almost everywhere. That is, if $g_1(\cdot) = g_2(\cdot)$ ω-almost everywhere, then we need that $g_1(\varphi(\cdot)) = g_2(\varphi(\cdot))$ ω-almost everywhere, too. Assuming that φ is non-singular with respect to ω is sufficient, which means that,

$$\forall X \subset \mathcal{X} : \omega(X) = 0 \implies \omega(\varphi^{-1}(X)) = \omega(\{x \in \mathcal{X} : \varphi(x) \in X\}) = 0.$$

Letting the Koopman operator act on our observed data results in,

$$T_\varphi g(\mathbf{x}_n) = g(\varphi(\mathbf{x}_n)) = g(\mathbf{x}_{n+1}),$$

which is the measurement of the next time step. The trade-off that is made in studying dynamical systems through the Koopman operator is that we gained linearity at the cost of the finite dimensionality of the data, as shown in Figure 4.1.

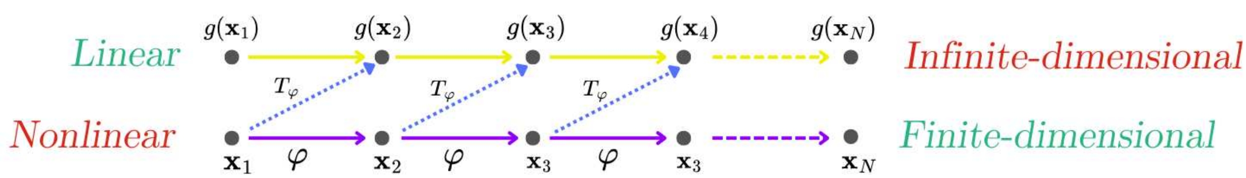

Figure 4.1: The concept of Koopman operators: By mapping the system to a space of observables, a nonlinear finite-dimensional system is transformed into a linear infinite-dimensional system.

Fix the Hilbert space $\mathcal{F} = \mathrm{L}^2(\mathcal{X}, \omega)$ and suppose that $g \in \mathrm{L}^2(\mathcal{X}, \omega)$ is an eigenfunction of T_φ with eigenvalue $\lambda \in \mathbb{C}$, then evaluating g on our observed data point x_n results in,

$$g(\mathbf{x}_n) = T_\varphi^n g(\mathbf{x}_0) = \lambda^n g(\mathbf{x}_0).$$

Thus, the observable evaluated at the data point x_n is a growth or decay by parameter λ multiplied by the value of the observable on x_0. The spectrum of the Koopman operator encapsulates information about the underlying dynamical system [Mez15]; [BMM12]; [Mez94]. The eigenvalues of an operator generalize to the notion of its spectrum,

$$\sigma(T_\varphi) := (\lambda \in \mathbb{C} : (T_\varphi - \lambda I) \text{ is not invertible}) \subset \mathbb{C},$$

where I denotes the identity operator. It is not difficult to show that the set of approximate eigenvalues, denoted $\sigma(T_\varphi)_{\text{ap}}$, is contained in the spectrum,

$$\sigma(T_\varphi)_{\text{ap}} = \left\{\lambda \in \mathbb{C}, : \exists (f_n)_{n\in\mathbb{N}} \subset \mathrm{L}^2(\mathcal{X}, \omega), \, \|f_n\| = 1 \, \forall n \in \mathbb{N}, \, \lim_{n\to\infty} \|T_\varphi f_n - \lambda f_n\| \to 0\right\} \subset \sigma(T_\varphi) \subset \mathbb{C}.$$

The approximate pseudoeigenvalues for $\varepsilon > 0$ are given by,

$$\sigma(T_\varphi)_{\text{ap},\varepsilon} = \left\{\lambda \in \mathbb{C}, : \exists (f_n)_{n\in\mathbb{N}} \subset \mathrm{L}^2(\mathcal{X}, \omega), \, \|f_n\| = 1 \, \forall n \in \mathbb{N}, \, \lim_{n\to\infty} \|T_\varphi f_n - \lambda f_n\| \leq \varepsilon\right\} \subset \sigma(T_\varphi)_{\text{ap}} \subset \mathbb{C}.$$

An observable $g \in \mathrm{L}^2(\mathcal{X}, \omega)$ with $\|g\| = 1$ and $\|T_\varphi g - \lambda g\| \leq \varepsilon$ is called a pseudoeigenfunction. By the structure of the Koopman operator, for $n \in \mathbb{N}$ we get,

$$\|T_\varphi^n g - \lambda^n g\| = \mathcal{O}(n\varepsilon).$$

It is more challenging to approximate the spectral properties of an infinite dimensional operator, T_φ, compared to a finite dimensional system. Dynamic Mode Decomposition (DMD) and the various extensions thereof aim to approximate spectral properties of T_φ and has numerous variants. We will introduce exact DMD, Extended DMD (EDMD) and kernel EDMD (kEDMD). For extensive overview of these methods, we refer to [Tu+14]; [WKR15]; [Wil+15], respectively. A review of the applications of the Koopman operator and what various other variants can be found in, e.g., [Bru+21]; [Col23].

4.2 Dynamic Mode Decomposition and its variants

4.2.1 Exact DMD

We will explain the original DMD algorithm, introduced by Schmidt [Sch10]. To fit to the Koopman paradigm, we need a dynamical system $(\mathcal{X}, \varphi)$ and a Banach space of functions $\mathcal{F}$. The dynamical system we consider is a discrete time dynamical system given by the observation data in equation (4.1). Having a finite amount of data results in that we cannot evaluate our data on an unobserved point. A way to resemble this in the choice of the domain of the Koopman operator is to equip the space of square integrable functions on $\mathcal{X}$ with the empirical point measure $\omega_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\mathbf{x}_i}$. One can check that this choice of measure is nonsingular with respect to φ. The observable function $g \in \mathrm{L}^2(\mathcal{X}, \omega_N)$ is taken to be the identity function, also called a full state observable. Given the snapshot data $\mathbb{X}, \mathbb{Y} \in \mathbb{R}^{d\times N}$ from equation (4.1), we aim to find a matrix $\mathbf{K}_{\text{DMD}}$ such that $\mathbb{Y} \approx \mathbf{K}_{\text{DMD}}\mathbb{X}$. The best fit that best describes the linear dynamics may be formulated as,

$$\tilde{\mathbf{K}}_{\text{DMD}} := \underset{\mathbf{K}_{\text{DMD}} \in \mathbb{C}^{d\times d}}{\arg\min} \|\mathbb{Y} - \mathbf{K}_{\text{DMD}}\mathbb{X}\|. \tag{4.2}$$

It is well known that $\tilde{\mathbf{K}}_{\text{DMD}} = \mathbb{Y}\mathbb{X}^{\dagger} \in \mathbb{C}^{d\times d}$, where $\dagger$ denotes the pseudo-inverse. The objective of the DMD algorithm is to approximate the leading spectral decomposition of the matrix $\tilde{\mathbf{K}}_{\text{DMD}}$. In doing so, the $\tilde{\mathbf{K}}_{\text{DMD}}$ is never explicitly computed. The dimension of the data is typically much larger than the number of observations points, that is $d \gg N$. The large size of the dimension d makes it intractable to compute the full spectrum. In exact DMD, devoloped by Tu et. al. [Tu+14], the SVD of rank $r \in \mathbb{N}$ of $\mathbb{X}$ is used to compute the pseudo-inverse of $\mathbb{X}$. The rank-r SVD of $\mathbb{X}$ is denoted as $\mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^*$, where the columns of $\mathbf{U}_r$ and $\mathbf{V}_r$ are orthonormal and $\mathbf{\Sigma}_r$ is diagonal. Tu et. al showed that the following procedure covers all of the nonzero eigenvalues of $\mathbf{K}_{\text{DMD}}$.

Algorithm 1 Exact DMD Algorithm [Tu+14]

Require: Snapshot data matrices $\mathbb{X} \in \mathbb{C}^{d\times M}$ and $\mathbb{Y} \in \mathbb{C}^{d\times N}$, rank $r \in \mathbb{N}$.
1: Compute the truncated singular value decomposition (SVD) of $\mathbf{X}$: $\mathbb{X} \approx \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^*$, where $\mathbf{U}_r \in \mathbb{C}^{d\times r}$, $\mathbf{\Sigma}_r \in \mathbb{R}^{r\times r}$ is diagonal, and $\mathbf{V}_r \in \mathbb{C}^{N\times r}$.
2: Compute the low-dimensional compression: $\tilde{f}_{\text{DMD}} = \mathbf{U}_r^*\mathbb{Y}\mathbf{V}_r\mathbf{\Sigma}_r^{-1} \in \mathbb{C}^{r\times r}$.
3: Perform the eigendecomposition of $\tilde{\mathbf{K}}_{\text{DMD},r}$: $\tilde{\mathbf{K}}_{\text{DMD},r}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$, where $\mathbf{W}$ contains eigenvectors and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues.
4: Compute the DMD modes: $\mathbf{\Phi} = \mathbb{Y}\mathbf{V}_r\mathbf{\Sigma}_r^{-1}\mathbf{W}$.
Ensure: Eigenvalues $\mathbf{\Lambda}$ and modes $\mathbf{\Phi} \in \mathbb{C}^{d\times r}$.

If the SVD in Algorithm 1 is exact, then the result should compute the eigenvalues of $\tilde{\mathbf{K}}_{\text{DMD}}$. In this case, we may write $\tilde{\mathbf{K}}_{\text{DMD}} = \mathbb{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^*$, we get,

$$\tilde{\mathbf{K}}_{\text{DMD}}\mathbf{\Phi} = \mathbb{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\underbrace{\mathbf{U}^*\mathbb{Y}\mathbf{V}\mathbf{\Sigma}^{-1}}_{=\tilde{\mathbf{K}}_{\text{DMD}}}\mathbf{W} = [\mathbb{Y}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}]\mathbf{\Lambda} = \mathbf{\Phi}\mathbf{\Lambda}.$$

The spectrum of $\tilde{\mathbf{K}}_{\text{DMD}}$, or an approximation thereof, encapsulates information of the dynamical system φ. This may appear odd, as the approximation in (4.2) assumes a linear case, whereas the dynamics are in practice often complex and nonlinear. For complicated systems, the spectrum of such linear methods can be too restricted. Imposing a prior nonlinear belief on the data can overcome this issue.

4.2.2 Extended Dynamic Mode Decomposition (EDMD)

A limitation of DMD is that it implicitly assumes the observables in $\mathrm{L}^2(\mathcal{X}, \omega_N)$ are full state observables, i.e. , $g(x) = x$. This restricts the approximation $\tilde{\mathbf{K}}_{\text{DMD}}$ of the Koopman operator to the subspace of the direct measurements $\mathbb{X}$. Thus, DMD is limited in its extrapolation. To address this issue, Williams et al. [WKR15] designed an extended version of DMD that allows a prior belief about the underlying dynamics to be imposed, dubbed the *Extended Dynamic Mode Decomposition* (EDMD). A predefined set of observable functions $\{\psi_1, \dots, \psi_M\}$ in $\mathrm{L}^2(\mathcal{X}, \omega)$ is selected. A common name for this set of functions is a dictionary, and we define $V_M = \operatorname{span}\{\psi_1, \dots, \psi_M\}$. For any collection of points $Z = (\mathbf{z}_i)_{i=1}^P$ in $\mathcal{X}$ and $P \in \mathbb{N}$, we introduce the notation,

$$\boldsymbol{\Psi}(\mathbf{z}_i) = [\psi_1(\mathbf{z}_i), \dots, \psi_M(\mathbf{z}_i)] \in \mathbb{C}^{1\times M}, \quad \boldsymbol{\Psi}_Z = \begin{bmatrix} \boldsymbol{\Psi}(\mathbf{z}_1) \\ \vdots \\ \boldsymbol{\Psi}(\mathbf{z}_P) \end{bmatrix} \in \mathbb{R}^{P\times M}.$$

Using this notation, any function in $g \in V_M$ evaluated at $x \in \mathcal{X}$ may be written as,

$$g(x) = \sum_{i=1}^{M} \mathbf{c}_i^g \psi_i(x) = \boldsymbol{\Psi}(x)\mathbf{c}^g.$$

for coordinates $\mathbf{c}^g \in \mathbb{C}^{M\times 1}$. The aim of EDMD is to find a finite dimensional approximation of T_φ restricted to the subspace V_M, that we call $\tilde{\mathbf{K}}_{\mathrm{EDMD}}$. For a given $g \in V_M$, we are after $\tilde{\mathbf{K}}_{\mathrm{EDMD}}\mathbb{R}^{M\times M}$ such that,

$$T_\varphi g \approx \sum_{i=1}^{M} (\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g)_i \, \psi_i.$$

This approximation can only be exact if T_φ is V_M invariant. Thus, we seek to minimize the following error term $R(g, x)$,

$$\begin{aligned} [T_\varphi g](x) = (g \circ \varphi)(x) = \boldsymbol{\Psi}(\varphi(x))\mathbf{c}^g &= \boldsymbol{\Psi}(x)\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g + \boldsymbol{\Psi}(\varphi(x))\mathbf{c}^g - \boldsymbol{\Psi}(x)\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g, \\ &= \boldsymbol{\Psi}(x)\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g + \underbrace{\left(\sum_{i=1}^{M} \psi_i(\varphi(x))\mathbf{c}_i^g - \boldsymbol{\Psi}(x)\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g\right)}_{:=R(g,x)}. \end{aligned}$$

Again, $R(g, x)$ can only equal zero if T_φ is V_M invariant. However, this is generally not the case for our chosen dictionary. From the observation data, the optimization problem becomes term to be minimized may be rewritten as,

$$\begin{aligned} \tilde{\mathbf{K}}_{\mathrm{EDMD}} &:= \underset{\mathbf{K}_{\mathrm{EDMD}} \in \mathbb{C}^{M\times M}}{\arg\min} \frac{1}{2} \sum_{i=1}^{N} |\boldsymbol{\Psi}(\varphi(\mathbf{x}_i))\mathbf{c}^g - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{K}_{\mathrm{EDMD}}\mathbf{c}^g|^2, \\ &= \underset{\mathbf{K}_{\mathrm{EDMD}} \in \mathbb{C}^{M\times M}}{\arg\min} \frac{1}{2} \sum_{i=1}^{N} |(\boldsymbol{\Psi}(y_i) - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{K}_{\mathrm{EDMD}})\, \mathbf{c}^g|^2. \end{aligned}$$

Approximating via the quadrature rule with nodes $\mathbb{X}$ and appropriate weights $(w_i)_{i=1}^N$, this least squares problem may be solved uniquely by $\tilde{\mathbf{K}}_{\mathrm{EDMD}} := \mathbf{G}^\dagger \mathbf{A}$, where,

$$\mathbf{G} = \sum_{i=1}^{N} w_i \boldsymbol{\Psi}(\mathbf{x}_i)^* \boldsymbol{\Psi}(\mathbf{x}_i) \in \mathbb{C}^{M\times M}, \quad \mathbf{A} = \sum_{i=1}^{N} w_i \boldsymbol{\Psi}(\mathbf{x}_i)^* \boldsymbol{\Psi}(y_i) \in \mathbb{C}^{M\times M} \tag{4.3}$$

where $*$ denotes the complex transpose. Let $\mathbf{D} = \operatorname{diag}(\frac{1}{M}, \dots, \frac{1}{M})$, for entries $1 \le i \le N$ and $1 \le j \le M$ of we may write,

$$\mathbf{G}_{ij} = \left[\boldsymbol{\Psi}_{\mathbb{X}}^* \mathbf{D} \boldsymbol{\Psi}_{\mathbb{X}}\right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} \overline{\psi_i(\mathbf{x}_k)} \psi_j(\mathbf{x}_k) = \langle \psi_j, \psi_i \rangle_{\mathrm{L}^2(\mathcal{X}, \omega_N)}, \tag{4.4}$$

$$\mathbf{A}_{ij} = \left[\boldsymbol{\Psi}_{\mathbb{X}}^* \mathbf{D} \boldsymbol{\Psi}_{\mathbb{Y}}\right]_{ij} = \frac{1}{N} \sum_{k=1}^{N} \overline{\psi_i(\mathbf{x}_k)} \psi_j(\mathbf{y}_k) = \langle T_\varphi \psi_j, \psi_i \rangle_{\mathrm{L}^2(\mathcal{X}, \omega_N)}. \tag{4.5}$$

Algorithmically, we have

Algorithm 2 EDMD Algorithm [WKR15]

Require: Snapshot data matrices $\mathbb{X} \in \mathbb{C}^{d\times N}$ and $\mathbb{Y} \in \mathbb{C}^{d\times N}$, quadrature weights $\{w_i\}_{i=1}^N$, and a dictionary of functions $\{\psi_m\}_{m=1}^M$.
1: Compute the matrices $\Psi_{\mathbb{X}}$, $\Psi_{\mathbb{Y}}$ and $\mathbf{D} = \text{diag}(w_1, \ldots, w_M)$.
2: Compute the matrices $\mathbf{G}$ and $\mathbf{A}$ according to equations (4.4) and (4.5), respectively.
3: Compute the EDMD matrix $\tilde{\mathbf{K}}_{\text{EDMD}} : \mathbf{G}^\dagger\mathbf{A} \in \mathbb{C}^{M\times M}$
4: Perform the eigendecomposition of $\tilde{\mathbf{K}}_{\text{EDMD}}$: $\tilde{\mathbf{K}}_{\text{EDMD}}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda}$, where $\mathbf{V}$ contains eigenvector coefficients and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.
Ensure: Eigenvalues $\boldsymbol{\Lambda}$ and eigenvector coefficients $\mathbf{V} \in \mathbb{C}^{M\times M}$.

Forming the matrices $\mathbf{G}$ and $\mathbf{G}$ both take $\mathcal{O}(NM^2)$ and computing the eigendecomposition of $\tilde{\mathbf{K}}_{\text{EDMD}}$ is $\mathcal{O}(M^3)$. For dictionaries with a large amount of functions, computations swiftly become intractible. The type of functions inside the dictionary are vital to a high quality approximation the Koopman operator. By choosing linear functions as observables, this algorithm is equivalent to standard DMD, developed by Tu et. al. except that in this particular choice of dictionary, we have $\tilde{\mathbf{K}}_{\text{EDMD}}^T = \tilde{\mathbf{K}}_{\text{DMD}}$. Hence, the name *extended* DMD is in place. Often it is the case that the size of the dictionary scales with the dimension of the state space $\mathbb{R}^d$. For example, suppose we are given dynamics in $\mathbb{R}^2$ from which we belief that the eigenfunctions of the Koopman operator are analytic. Then, we may choose all polynomials up to some degree, say 30. From Example 2.1.14, the set of all polynomials would have size $\binom{2+30}{30} = 496$, which makes computations feasible. However, suppose we now have a dynamical system in $\mathbb{R}^{128}$, then the set of all polynomials up to degree 30 is $\binom{128+30}{30} = \mathcal{O}(10^{32})$. Scaling poorly with the dimension of the state space is also referred to as the curse of dimensionality. To address this issue, one requires an algorithm where the size of the dictionary does not scale problematically with respect to the dimension d.

4.2.3 Kernel Extended Mode Decomposition (kEDMD)

Suppose we are given a real-valued kernel k over $\mathbb{R}^d$ and snapshot data $\mathbb{X}$ and $\mathbb{Y}$. Then, a canonical data-driven approach to choose a dictionary is the set $\{k(\mathbf{x}_1, \cdot), \ldots, k(\mathbf{x}_N, \cdot)\}$. Call $V_{\mathbb{X}} = \text{span}\{k(\mathbf{x}_1, \cdot), \ldots, k(\mathbf{x}_N, \cdot)\}$ and define $V_{\mathbb{Y}}$ analogously. We see that the size of the dictionary equals the number of observation points, $M = N$. A suitable kernel function is the only choice to make. When the dictionary contains the canonical feature maps of a kernel, we will call EDMD with this specific dictionary kernel EDMD (kEDMD). We will make a distinction for kEDMD with one and two dictionaries. First, we will cover the case where $V_{\mathbb{X}}$ is the dictionary present. Hereafter, we present a different version of EDMD with two dictionaries, $V_{\mathbb{X}}$ and $V_{\mathbb{Y}}$.

Before we continue, we introduce some notation. Let $L, P \in \mathbb{N}$, for sets of points $(\mathbf{z}_i)_{i=1}^L \subset \mathbb{R}^d$ and $(\mathbf{w}_j)_{j=1}^P \subset \mathbb{R}^P$, we define the kernel matrices $\mathbf{K}_{\mathbf{z},\mathbf{w}} \in \mathbb{R}^{L\times P}$ componentwise as $(\mathbf{K}_{\mathbf{z},\mathbf{w}})_{ij} := k(\mathbf{z}_i, \mathbf{w}_j)$ where $1 \le i \le L$ and $1 \le j \le P$. Let $h \in V_{\mathbf{w}}$ and define $h_{\mathbf{z}} := (h(\mathbf{z}_1), \ldots, \mathbf{z}_L)^T \in \mathbb{R}^{L\times 1}$. Since $h \in V_{\mathbf{w}}$ we can write $h = \sum_{i=1}^P \mathbf{c}_i^h k(\mathbf{w}_i, \cdot)$ and we call the element $\mathbf{c}^h \in \mathbb{R}^{P\times 1}$ the coordinates of h, with respect to the canonical basis of $V_{\mathbf{w}}$. It turns out that $K_{\mathbf{z},\mathbf{w}}\mathbf{c}^h$ consists of the function h, evaluated at the points of $\mathbf{z}$, for $1 \le i \le N$ we have,

$$\left(\mathbf{K}_{\mathbf{z},\mathbf{w}}\mathbf{c}^h\right)_i = \sum_{j=1}^P (\mathbf{K}_{\mathbf{z},\mathbf{w}})_{ij}\, \mathbf{c}_j^h = \sum_{j=1}^P k(\mathbf{z}_i, \mathbf{w}_j)\mathbf{c}_j^h = \left(\sum_{j=1}^P \mathbf{c}_j^h k(\mathbf{w}_j, \cdot)\right)(\mathbf{z}_i) = h(\mathbf{z}_i). \tag{4.6}$$

We have that $\mathbf{K}_{\mathbf{z},\mathbf{w}}\mathbf{c}^h = h_{\mathbf{z}}$.

Suppose that we choose only one dictionary, $V_{\mathbb{X}}$. This is a special case of EDMD where the dictionary is the set of kernel functions centered at the data $\mathbb{X}$. For a given $g \in V_{\mathbb{X}}$ with coordinates $\mathbf{c}^g$, the matrix $\tilde{\mathbf{K}}_{\mathrm{kEDMD}}$ aims to approximate,

$$(T_\varphi g)(\cdot) \approx \sum_{i=1}^{N} \left(\tilde{\mathbf{K}}_{\mathrm{kEDMD}} \mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot).$$

We read off the matrix $\tilde{\mathbf{K}}_{\mathrm{kEDMD}}$ immediately from Equations (4.4) and (4.5) to see that $\tilde{\mathbf{K}}_{\mathrm{kEDMD}} = \mathbf{K}_{\mathbb{X}}^{-1}\mathbf{K}_{\mathbb{X},\mathbb{Y}}$. We immediately see that the kernel trick can be used for a computationally efficient construction of the matrices $\mathbf{G}$ and $\mathbf{A}$, respectively. By being able to compute the inner product through a kernel evaluation, $\mathbf{G}$ and $\mathbf{A}$ can be computed in $\mathcal{O}(N^2M)$, an improvement from $\mathcal{O}(NM^2)$ if the size of the dictionary M is large.

Now we present kEDMD with two dictionaries, $V_{\mathbb{X}}$ and $V_{\mathbb{Y}}$. The objective in this case changes slightly. We desire the Koopman operator to map into $V_{\mathbb{X}}$, when acting on $V_{\mathbb{Y}}$. Given $g \in V_{\mathbb{Y}}$ with coordinates $\mathbf{c}^g$, estimate $\tilde{\mathbf{K}}_{\mathrm{kEDMD}} \in \mathbb{R}^{N\times N}$ such that

$$(T_\varphi g)(\cdot) \approx \sum_{i=1}^{N} \left(\tilde{\mathbf{K}}_{\mathrm{kEDMD}} \mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot).$$

Note that on g is a linear combination of kernel functions, centered on $V_{\mathbb{Y}}$. Going through the machinery of the EDMD procedure, we see that the formulae for the matrices $\mathbf{G}$ and $\mathbf{A}$ from Equations (4.4) and (4.5) change as follows,

$$\mathbf{G}_{ij} = \langle k(\mathbf{x}_j, \cdot), k(\mathbf{x}_i, \cdot)\rangle, \quad \mathbf{A}_{ij} = \langle T_\varphi k(\mathbf{y}_j, \cdot), k(\mathbf{x}_i, \cdot)\rangle, \quad 1 \le i, j \le N.$$

For this version of kEDMD, we see that the approximation of the Koopman operator is $\tilde{\mathbf{K}}_{\mathrm{kEDMD}} = \mathbf{K}_{\mathbb{X}}^{-1}\mathbf{K}_{\mathbb{Y}} \in \mathbb{R}^{N\times N}$. Recall that, for $g \in V_{\mathbb{Y}}$, the matrix $\tilde{\mathbf{K}}_{\mathrm{kEDMD}}$ aims to approximate

$$\sum_{i=1}^{N} \left(\tilde{\mathbf{K}}_{\mathrm{kEDMD}} \mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot) \approx (T_\varphi g)(\cdot) = g(\varphi(\cdot)) = \sum_{j=1}^{N} \mathbf{c}^g_j k(\mathbf{y}_j, \varphi(\cdot)).$$

We claim that the matrix represents the operator $P_{V_{\mathbb{X}}} T_\varphi\big|_{V_{\mathbb{Y}}}$, the Koopman operator followed by a projection onto $V_{\mathbb{X}}$, restricted to $V_{\mathbb{Y}}$. The formula for the projection from H_k onto $V_{\mathbb{X}}$, denoted $P_{V_{\mathbb{X}}}$, was explored in Example 2.4.3. Let $f \in H_k$, then with some abuse of notation $[P_{V_{\mathbb{X}}} T_\varphi f](\cdot) = \sum_{i=1}^{N} (\mathbf{K}_{\mathbb{X}}^{-1} f_{\varphi(\mathbb{X})})_i k(\mathbf{x}_i, \cdot)$. The coordinates of $P_{V_{\mathbb{X}}} T_\varphi\big|_{V_{\mathbb{Y}}}(g)$ with respect to the canonical basis of $V_{\mathbb{X}}$ are given by $\left(\mathbf{K}_{\mathbb{X}}^{-1} g_{\mathbb{Y}}\right)$. Note that $g_{\mathbb{Y}}$ can, as we have shown in Equation (4.6), be written as $g_{\mathbb{Y}} = \mathbf{K}_{\mathbb{Y}}\mathbf{c}^g$. Thus, the Koopman operator followed by the projection onto $V_{\mathbb{X}}$, restricted to $V_{\mathbb{Y}}$ is given by,

$$\left[P_{\mathbb{X}}\big|_{V_{\mathbb{Y}}}(g)\right](\cdot) = \sum_{i=1}^{N} \left(\mathbf{K}_{\mathbb{X}}^{-1}\mathbf{K}_{\mathbb{Y}}\mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot) = \sum_{i=1}^{N} \left(\tilde{\mathbf{K}}_{\mathrm{kEDMD}} \mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot).$$

In the same spirit, one can show that the operator that is represented in the case of only one dictionary $V_{\mathbb{X}}$ is $P_{V_{\mathbb{X}}}\big|_{V_{\mathbb{Y}}}$.

It turns out that the matrix $\tilde{\mathbf{K}}_{\mathrm{kEDMD}}$ with one dictionary does not encapsulate any form of composition with the dynamics φ. Instead, it is a change of coordinates from $V_{\mathbb{Y}}$ to $V_{\mathbb{X}}$ in the form of a projection onto the subspace $V_{\mathbb{X}}$. The matrix $\tilde{\mathbf{K}}_{\mathrm{kEDMD}}$ when electing two dictionaries, $V_{\mathbb{X}}$ and $V_{\mathbb{Y}}$ does reflect a composition in the operator that it represents.

4.3 Kernel ridge regression

In this section, we discuss a recent "DMD-free" method by Kostic et. al [Kos+22] that approximates the Koopman operator from data. This method depends on the developed theory in Section 2.4. The approach taken by Kostic et. al. is explained after which it will be put in the same context as introduced in Section 4.1. Hereafter, the comparison with the kEDMD is made. Let k be a real-valued continuous kernel on a compact set $\mathcal{X}$, so that Assumptions (1- 2) are satisfied, and $(\varphi, \mathcal{X})$ an unknown dynamical system. We assume that the Koopman operator maps $T_\varphi : \mathrm{L}^2(\mathcal{X}, \omega) \to \mathrm{L}^2(\mathcal{X}, \omega)$ boundedly. Assume further that ω is finite and that for any $x \in \mathcal{X}$ the $k(x,x) < \infty$ ω almost-everywhere. It follows that the inclusion operator $S_\omega : H_k \hookrightarrow \mathrm{L}^2(\mathcal{X}, \omega)$ is bounded, let $f \in H_k$

$$\|f\|^2_{\mathrm{L}^2(\mathcal{X},\omega)} = \int_{\mathcal{X}} |f(x)|^2 \mathrm{d}\omega(x) = \int_{\mathcal{X}} |\langle f, k(x,\cdot)\rangle_{H_k}|^2 \mathrm{d}\omega(x) \leq \int_{\mathcal{X}} \|f\|^2_{H_k} k(x,x) \mathrm{d}\omega(x) = \|f\|^2_{H_k} k(x,x)\omega(\mathcal{X}) < \infty.$$

The bound above holds ω-almost everywhere. It is shown in [SC08, Theorem 4.26] that that the adjoint of S_ω is the integral operator

$$\begin{aligned} S^*_\omega : \mathrm{L}^2(\mathcal{X}, \omega) &\to H_k, \\ f \mapsto (S^*_\omega f)(\cdot) &= \int_{\mathcal{X}} k(x,\cdot) f(x) \mathrm{d}\omega(x). \end{aligned}$$

We see that S^*_ω is $A_k|_{H_k}$, where A_k is defined as in Proposition 2.2.4. Since A_k is Hilbert-Schmidt, the restriction is Hilbert-Schmidt too. As the norms of an operator and its adjoint are equivalent, we conclude that S_ω is a Hilbert-Schmidt operator. Since by assumption T_φ is bounded,the composition

$$\hat{T}_\varphi := T_\varphi S_\omega : H_k \to \mathrm{L}^2(\mathcal{X}, \omega),$$

is a Hilbert-Schmidt operator too. Thus, it is appropriate to approximate $\hat{T}_\varphi$ by means of Hilbert-Schmidt operators. To do so, let $M \in \mathrm{HS}(H_k)$ and define the risk as

$$\mathcal{R}(M) := \|\hat{T}_\varphi - S_\omega M\|_{\mathrm{HS}(H_k, \mathrm{L}^2(\mathcal{X},\omega))}.$$

Our aim is to minimize the empirical risk, once we are given data. Error bounds from the risk to the empirical risk are given in the original paper by Kostic et. al. To fit the framework presented in Section 4.1, we assume the same snapshot data are present. To define the empirical risk, we need the following operators.

$$\begin{aligned} \hat{S} : H_k &\to \mathbb{R}^N, \\ f \mapsto \hat{S}(f) &:= \frac{1}{\sqrt{N}}(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)). \end{aligned}$$

Since the range of $\hat{S}$ is finite dimensional and each component of $\hat{S}(f)$ can be bounded, it follows that $\hat{S} \in \mathrm{HS}(H_k, \mathbb{R}^N)$. The adjoint can easily be computed, let $f \in H$ and $\mathbf{a} \in \mathbb{R}^N$,

$$\langle \hat{S} f, \mathbf{a}\rangle_{\mathbb{R}^N} = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{a}_i f(\mathbf{x}_i) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{a}_i \langle f, k(\mathbf{x}_i, \cdot)\rangle_{H_K} = \langle f, \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{a}_i k(\mathbf{x}_i, \cdot)\rangle_{H_k} = \langle f, \hat{S}^* \mathbf{a}\rangle_{H_k}.$$

In fact, even though it is true that $\hat{S}^*$ maps into H_k, we can make the stronger statement that $\hat{S}^* : \mathbb{R}^N \to V_{\mathbf{X}}$. The composition $\hat{S}\hat{S}^* : \mathbb{R}^N \to \mathbb{R}^N$ is a map between finite dimensional spaces, and thus can be represented as a matrix in $\mathbb{R}^{N \times N}$. It turns out that this matrix is precisely the kernel matrix $\mathbf{K}_{\mathbf{X}}$ multiplied by $1/N$. To see this, let $\mathbf{a} \in \mathbb{R}^N$, since k is symmetric we get,

$$\begin{aligned} \hat{S}\hat{S}^*(\mathbf{a}) = \frac{1}{\sqrt{N}} \hat{S} \left(\sum_{i=1}^N \mathbf{a}_i k(\mathbf{x}_i, \cdot) \right) = \frac{1}{N} \begin{bmatrix} \sum_{i=1}^N \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_1) \\ \vdots \\ \sum_{i=1}^N \mathbf{a}_i k(\mathbf{x}_i, \mathbf{x}_N) \end{bmatrix} &= \frac{1}{N} \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \ldots & k(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N, \mathbf{x}_1) & \ldots & k(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_N \end{bmatrix}, \\ &= \frac{1}{N} \mathbf{K}_{\mathbf{X}} \mathbf{a}. \end{aligned}$$

Similarly for $\hat{S}^*\hat{S} \in \mathrm{HS}(H_k)$, let $f \in H_k$ we get

$$\hat{S}^*\hat{S}(f) = \frac{1}{\sqrt{N}}\hat{S}^*\left[\,(f(\mathbf{x}_1,\dots,f(\mathbf{x}_N))\,\right] = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)k(\mathbf{x}_i,\cdot) = \frac{1}{N}\sum_{i=1}^{N}\langle f, k(\mathbf{x}_i,-)\rangle_{H_k} k(\mathbf{x}_i,\cdot),$$
$$= \frac{1}{N}\sum_{i=1}^{N}\Xi\left(k(\mathbf{x}_i,\cdot)\otimes k(\mathbf{x}_i,-)\right)(f).$$

Where we made use of the isomorphism Ξ defined in Theorem 2.4.6. For notational convenience, we will omit the map Ξ. To summarize, we have the following operators,

$$\hat{S} \in \mathrm{HS}(H_k,\mathbb{R}^N), \quad \hat{S}(f) = \frac{1}{\sqrt{N}}(f(\mathbf{x}_i))_{i=1}^N, \qquad \hat{S}\hat{S}^* \in \mathcal{L}(\mathbb{R}^N), \quad \hat{S}\hat{S}^*\mathbf{a} = \frac{1}{N}\mathbf{K}_{\mathbb{X}}\mathbf{a},$$
$$\hat{S}^* \in \mathrm{HS}(\mathbb{R}^N, V_{\mathbb{X}}), \quad \hat{S}^*(\mathbf{a}) = \sum_{i=1}^{N}\mathbf{a}_i k(\mathbf{x}_i,\cdot), \qquad \hat{S}^*\hat{S} \in \mathrm{HS}(H_k), \quad \hat{S}^*\hat{S}(f) = \frac{1}{N}\sum_{i=1}^{N} k(\mathbf{x}_i,\cdot)\otimes k(\mathbf{x}_i,-)(f).$$

Define analogous operators on $\mathbb{Y}$,

$$\hat{Z} \in \hat{Z}(H_k,\mathbb{R}^N), \quad \hat{Z}(f) = \frac{1}{\sqrt{N}}(f(\mathbf{y}_i))_{i=1}^N, \qquad \hat{Z}\hat{Z}^* \in \mathcal{L}(\mathbb{R}^N), \quad \hat{Z}\hat{Z}^*\mathbf{a} = \frac{1}{N}\mathbf{K}_{\mathbb{Y}}\mathbf{a},$$
$$\hat{Z}^* \in \hat{Z}(\mathbb{R}^N, V_{\mathbb{Y}}), \quad \hat{Z}^*(\mathbf{a}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{a}_i k(\mathbf{y}_i,\cdot), \quad \hat{Z}^*\hat{Z} \in \hat{Z}(H_k), \quad \hat{Z}^*\hat{Z}(f) = \frac{1}{N}\sum_{i=1}^{N} k(\mathbf{y}_i,\cdot)\otimes k(\mathbf{y}_i,-)(f).$$

We can read off the empirical input, output and cross covariances, which are given by $\hat{S}^*\hat{S}$, $\hat{Z}^*\hat{Z}$ and $\hat{S}^*\hat{Z}$, respectively. The respective empirical kernel matrices with respect to the snapshot data are given by $\hat{S}\hat{S}^*$, $\hat{Z}\hat{Z}^*$ and $\hat{Z}\hat{S}^*$. The Koopman operator that we can estimate can only be acting between the spaces, i.e. $M : V_{\mathbb{Y}} \to V_{\mathbb{X}}$, as we do not have knowledge of what is outside these data.
The Koopman can now be estimated by minimizing the empirical risk,

$$\hat{\mathcal{R}}(M) = \|\hat{Z} - \hat{S}M\|^2_{\mathrm{HS}(V_{\mathbb{Y}},\mathbb{R}^N)}$$

Our optimization problem boils down to,

$$\text{Given,}\quad \mathbb{X},\mathbb{Y},\quad \text{solve}\quad \min_{M\in\mathrm{HS}(V_{\mathbb{Y}},\mathbb{R}^N)}\hat{\mathcal{R}}(M) = \min_{M\in\mathrm{HS}(V_{\mathbb{Y}},V_{\mathbb{X}})}\frac{1}{N}\sum_{i=1}^{N}\|k(\mathbf{y}_i,\cdot) - M^*k(\mathbf{x}_i,\cdot)\|_{H_k}. \tag{4.7}$$

It can be shown that the empirical risk can be written as in the right hand side in Equation 4.7, which is connected with Conditional Mean Embeddings (CME) [Son+09]. Within a more grounded measure theoretical framework, one can view the Koopman operator as a conditional expectation. Unfortunately, we can merely motivate the presence of the adjoint by mentioning that this ensures all objects live inside the same space, namely $V_{\mathbb{Y}}$, in our framework. A connection between vvRKHS and Hilbert-Schmidt operators was explored in Section 2.4. It turns out that the connection between least squares regression problems and CME relies on Corollary 2.4.8, studied by Mollenhauer and Koltai in [MK20].
Through the isometry $\Xi \circ \Upsilon^{-1}$ from Corollary 2.4.8, we may formulate the problem by minimizing over $V_{\mathbb{X}}$ valued functions, instead of optimizing over for an operator $M \in \mathrm{HS}(V_{\mathbb{Y}}, V_{\mathbb{X}})$. Recall that optimization over functions in a scalar-valued RKHS is convenient, as a closed form solution can be found easily by means of the representer Theorem 2.4.1. We presented a regularized regression problem for a scalar-valued RKHS in Example 2.4.4. An analogous result holds true for regression in vector-valued RKHS, shown by Micchelli and Pontil [MP05]. Let $\gamma \geq 0$ and a Tikhonov regularization term to Equation (4.7). The estimator $\hat{M}$ to the altered problem,

$$\min_{M\in\mathrm{HS}(V_{\mathbb{Y}},V_{\mathbb{X}})}\hat{\mathcal{R}}(M) + \gamma\|M\|^2_{\mathrm{HS}(V_{\mathbb{Y}},V_{\mathbb{X}})}, \tag{4.8}$$

is given by $\hat{M} = \hat{S}^*(\mathbf{K}_{\mathbb{X}} + \gamma I_N)^{-1}\hat{Z}$, where I_N denotes the identity on $\mathbb{R}^{N\times N}$. The regularized problem in Equation (4.8) is referred to as Kernel Ridge Regression (KRR). To put KRR in the same framework as kEDMD, we let $\hat{M}$ act on an arbitrary element in its domain. Let $\hat{Z}$ act on some $g \in V_{\mathbb{Y}}$ with coefficients $\mathbf{c}^g$ with respect to the canonical basis. Note that evaluating an element in $V_{\mathbb{Y}}$ on $V_{\mathbb{Y}}$, which is what $\hat{Z}$ does up to a constant, is a special case of Equation (4.6). We immediately write,

$$\hat{Z}(g) = \frac{1}{\sqrt{N}}\mathbf{K}_{\mathbb{Y}}\mathbf{c}^g.$$

Thus, we see that the approximation of the Koopman operator $\hat{M}$ can be written as,

$$\begin{aligned}\hat{M}(g) = \hat{S}^*(\frac{1}{N}\mathbf{K}_{\mathbb{X}} + \gamma I_N)^{-1}\hat{Z}(g) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left((\frac{1}{N}\mathbf{K}_{\mathbb{X}} + \gamma I_N)^{-1}\frac{1}{\sqrt{N}}\mathbf{K}_{\mathbb{Y}}\mathbf{c}^g\right)_i k(\mathbf{x}_i, \cdot),\\ &= \sum_{i=1}^{N}\Bigg(\underbrace{(\mathbf{K}_{\mathbb{X}} + N\gamma I_N)^{-1}\mathbf{K}_{\mathbb{Y}}\mathbf{c}^g}_{:=\tilde{\mathbf{K}}_{\text{KRR}}}\Bigg)_i k(\mathbf{x}_i, \cdot).\end{aligned}$$

We claim that, for $\gamma = 0$, the matrix $\tilde{\mathbf{K}}_{\text{KRR}}$ represents the operator $P_{V_{\mathbb{X}}}T_\varphi\Big|_{V_{\mathbb{Y}}}$. Let $f \in H_k$, the projection onto $V_{\mathbb{X}}$ is given by $[P_{V_{\mathbb{X}}}f](\cdot) = \sum_{i=1}^{N}(\mathbf{K}_{\mathbb{X}}^{-1}f_{\mathbb{X}})_i k(\mathbf{x}_i, \cdot)$. Let $g \in V_{\mathbb{Y}}$ with coordinates $\mathbf{c}^g$. The coordinates of g after composing with φ and then projecting are, with some abuse of notation, $\mathbf{K}_{\mathbb{X}}^{-1}g_{\varphi(\mathbb{X})} = \mathbf{K}_{\mathbb{X}}^{-1}g_{\mathbb{Y}}$. Recall from the derivation in Equation (4.6) that this equals $\mathbf{K}_{\mathbb{X}}^{-1}\mathbf{K}_{\mathbb{Y}}\mathbf{c}^g = \tilde{\mathbf{K}}_{\text{KRR}}\mathbf{c}^g$.

Observe that for the case $\gamma = 0$ the matrices $\tilde{\mathbf{K}}_{\text{KRR}}$ and $\tilde{\mathbf{K}}_{\text{kEDMD}}$ coincide, in the case one chooses two dictionaries, $V_{\mathbb{X}}$ and $V_{\mathbb{Y}}$, for kEDMD. For $\gamma > 0$, the regularization term $N\gamma I_N$ is added.

4.4 Some caveats

We give a slightly more negative view of the approximation of the Koopman operator. As we have seen in the case for $\tilde{\mathbf{K}}_{\text{KRR}}$, the Koopman operator is approximated using a finite rank operator. The idea is that the Koopman operator is approximated more accurately when the number of sample points increases. It is therefore natural to question the compactness of the Koopman operator. It turns out that the compactness of the Koopman operator on $\mathrm{L}^2(\mathcal{X}, \omega)$ depends on the underlying measure ω.

First, we need the notion of an atom. Let (S, Σ) be a measurable space and let μ be a measure on that space. An measurable set A is an atom if $\mu(A) > 0$ and if for any measurable subset $B \subset A$ we have $0 \in \{\mu(B), \mu(A \setminus B\}$. A measure that does not contain atoms is called non-atomic. It was shown by Singh and Kumar that no compact Koopman operators exist on $\mathrm{L}^2(\mathcal{X}, \omega)$ if ω is non-atomic [SK79, Corollary 2.1]. Since the Lebesgue measure is non-atomic, it follows that the standard $\mathrm{L}^2(\mathbb{R})$ space does not contain any compact Koopman operators, no matter the underlying dynamics.

The choice of underlying Banach space on which the Koopman operator reflects in its properties. Boundedness of the Koopman operator between Banach spaces was characterized in Theorem 2.2.20. We saw that the Koopman operator is bounded if the space is preserved which can be a simple condition, depending on the underlying RKHS. Hence, the choice of RKHS is of great importance to what dynamics one can have for a bounded Koopman operator. Even for popular Gaussian kernels, the dynamics for which the Koopman operator is bounded are dramatically restricted.

4.4.1 Space preservation

We give two examples of RKHSs H_k and dynamics φ such that T_φ maps boundedly between H_k.

Example 4.4.1 (Rotation preserves the RKHS $H^2(\mathbb{D})$)**.** Consider the Hardy space $H^2(\mathbb{D})$ from Example 2.1.22 and the dynamical system $(\mathbb{D}, \varphi)$, where $\varphi(z) = ze^{i\pi/2}$. Let $f \in H^2(\mathbb{D})$ and $z \in \mathbb{C}$.

We immediately see that $\|f\|_{H^2(\mathbb{D})} = \|T_\varphi f\|_{H^2(\mathbb{D})}$ and we conclude that the dynamical system preserves the RKHS structure.

Example 4.4.2 (Sobolev space $H^s(\mathbb{R}^d)$)**.** Let $H^k(\mathbb{R}^d)$ with $k > d/2$. It is a well known result that this space is included in the bounded continuous functions on $\mathbb{R}^d$, which makes $H^k(\mathbb{R}^d)$ a RKHS. Let φ be a k times continously differentiable diffeomorphism on $\mathbb{R}^d$. Then the Koopman operator T_φ between $H^k(\mathbb{R}^d)$ is bounded. Let $f \in H^k(\mathbb{R}^d)$, we need to show that $f \circ \varphi \in H^k(\mathbb{R}^d)$. We will show that the norm of the first partial derivative is bounded, similar calculations can be done for higher order derivatives. To show this, we will use the fact that $\|\frac{\partial \varphi}{\partial x_i}\|_\infty \leq C$ for some $C > 0$. We use the symbols ∇ and D for the gradient and the Jacobian, respectively. By a change of variables and since $\varphi(\mathbb{R}^d) = \mathbb{R}^d$ we get,

$$\begin{aligned}
\int_{\mathbb{R}^d} \left| \frac{\partial}{\partial x_i} f \circ \varphi(z) \right|^2 \mathrm{d}z &= \int_{\mathbb{R}^d} \left| \sum_{j=1}^d \frac{\partial f}{\partial x_j}(\varphi(z)) \frac{\partial \varphi}{\partial x_i}(z) \right|^2 \mathrm{d}z, \\
&= \int_{\varphi(\mathbb{R}^d)} \left| \sum_{j=1}^d \frac{\partial f}{\partial x_j}(w) \frac{\partial \varphi}{\partial x_i}(\varphi^{-1}(w)) \right|^2 \left|\det D\varphi(\varphi^{-1}(w))\right|^{-1} \mathrm{d}w \quad (w = \varphi(z), \text{ so } \varphi^{-1}(w) = z), \\
&\leq \tilde{C} \int_{\mathbb{R}^d} |\nabla f(w)|^2 \, \mathrm{d}w, \\
&= \tilde{C} \|\nabla f\|_{L^2(\mathbb{R}^d)} < \infty.
\end{aligned}$$

Here $\tilde{C} > 0$ is some constant depending on C.

4.4.2 The Koopman operator between Gaussian RKHS

The Gaussian kernel and its corresponding RKHS turns out to be an interesting space to study the boundedness of the Koopman operator. The Gaussian kernel on $\mathbb{R}$ with parameter $\sigma > 0$ is denoted

$$k_\sigma(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right).$$

Example 4.4.3 (Inclusion of Gaussian kernels [Phi+24])**.** Let $\mathcal{X}$ be a compact subset of $\mathbb{R}$, let $M \gg 1$ and define $\varphi : \mathcal{X} \to \mathcal{X}$ by $\varphi(x) := x/M$. Let $\sigma > 0$ and let k_σ denote the Gaussian kernel, composing k_σ and φ results in,

$$(k_\sigma \circ \varphi)(x, y) = \exp\Big(-\frac{(\frac{x}{M} - \frac{y}{M})^2}{2\sigma^2}\Big) = \exp\Big(-\frac{(x-y)^2}{2(\sigma M)^2}\Big) = k_{\sigma M}(x, y).$$

This is another Gaussian kernel with parameter $\sigma M \gg \sigma > 0$. From [SHS06, Corollary 6] it follows that $H_{\sigma M} \hookrightarrow H_\sigma$ with constant $\sqrt{M}$. From Theorem 2.2.11 we have $\|f\|_{H_\sigma} \leq \sqrt{M}\|f\|_{H_{\sigma M}}$ and $k_{\sigma M} \preceq M k_\sigma$. In [SHS06, Corollary 7], it is shown that the inclusion is not surjective.

We generalize the Gaussian kernel to act on elements of $\mathbb{R}^d$, the parameter $\sigma \in \mathbb{R}$ is replaced with a PSD matrix $C \in \mathbb{R}^{d \times d}$. With k_C we denote Gaussian kernel on $\mathbb{R}^d$ with parameter C and is defined as,

$$k_C(\mathbf{x}, \mathbf{y}) := \exp\left(\|{-C^{-1}}(\mathbf{x} - \mathbf{y})\|^2\right).$$

For the special case $C = \sigma I_d$, where $\sigma > 0$ and I_d is the identity on $\mathbb{R}^d$, the kernel $k_{\sigma I_d} = \exp\left(-\|(\mathbf{x} - \mathbf{y})\|^2/\sigma^2\right)$.

Example 4.4.4 (Restriction on dynamics for bounded operator on H_C)**.** Let $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ be a map. It has been shown in [Gon+24, Corollary 1] that the Koopman operator T_φ is bounded between H_σ only if the underlying dynamics are affine, that is $\varphi(\mathbf{x}) = A\mathbf{x} + b$, where $A \in \mathbb{R}^{d \times d}$ and $b \in \mathbb{R}^d$.

Corollary 4.4.5 (Proposition 3.3 in [Phi+24])**.** Let $C_1, C_2 \in \mathbb{R}^{d\times d}$ with $d > 1$ be positive semidefinite. Then H_{C_1} can be continuously embedded in H_{C_2} if and only if $C_1^2 - C_2^2$ is positive semidefinite. In this case we have for $f \in H_{C_1}$

$$\|f\|_{C_2} \leq \left(\frac{\det(C_1)}{\det(C_2)}\right)^{1/2} \|f\|_{C_1} \quad \text{and,} \quad k_{C_1}^2 \preceq \frac{\det(C_1)}{\det(C_2)} k_{C_2}^2.$$

It follows from Example 4.4.4 that at least the dynamics need to be affine for the Koopman operator to be bounded between H_C. Combining this result with Corollary 4.4.5 we can impose further restrictions on the dynamics, given that the dynamics are invertible.

Corollary 4.4.6. Let $C \in \mathbb{R}^{d\times d}$ be a positive semidefinite matrix and φ be affine dynamics given by $\varphi(\mathbf{x}) = A\mathbf{x} + b$ for $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$. Furthermore, assume that A is invertible. Then, the Koopman operator T_φ is bounded between H_C if only if $(CA^{-1})^2 - C^2$ is positive semidefinite.

Proof. Composing the Gaussian kernel with the dynamics φ results in

$$(k_C \circ \varphi)(\mathbf{x}, \mathbf{y}) = \exp\left(\|-C^{-1}((A\mathbf{x} + b) - (A\mathbf{y} + b)\|^2\right) = \exp\left(\|-(A^{-1}C)^{-1}(\mathbf{x} - \mathbf{y})\|^2\right) = k_{A^{-1}C}(\mathbf{x}, \mathbf{y}).$$

It follows from Corollary 4.4.5 that $H_{A^{-1}C}$ can be continuously embedded into H_C if and only if $(A^{-1}C)^2 - C^2$ is PSD. □

Let us apply Corollary 4.4.6 to the special case $C = \sigma I_d$. We make a further assumption that A is diagonalizable and that $b = 0$. This means that we may write $A = PDP^{-1}$, where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{d\times d}$, $P = [\mathbf{p}_1, \ldots, \mathbf{p}_d] \in \mathbb{R}^{d\times d}$ and each $p_i \in \mathbb{R}^{d\times 1}$ is an eigenvector of A with eigenvalue λ_i, for $1 \leq i \leq d$. Then, the condition can be simplified,

$$(A^{-1}C)^2 - C^2 = \left(P^{-1}\mathrm{diag}(\lambda_1^{-2}, \ldots, \lambda_d^{-2})P\sigma^2 I_d\right) - \sigma^2 I_d = P^{-1}\mathrm{diag}(\sigma^2(\frac{1}{\lambda_1^2} - 1), \ldots, \sigma^2(\frac{1}{\lambda_d^2} - 1))P.$$

We see that this matrix is PSD if and only if $\lambda_i \geq 1$ for all $1 \leq i \leq d$. Since the elements of P form a basis for $\mathbb{R}^d$, we see have that any $\mathbf{x} \in \mathbb{R}^d$ can be written as $\mathbf{x} = \sum_{i=1}^d \mathbf{c}_i \mathbf{p}_i$, for some scalars $\mathbf{c} \in \mathbb{R}^d$. In this case, we have an expanding dynamical system,

$$\|\varphi(\mathbf{x})\|_{\mathbb{R}^d} = \|A\mathbf{x}\|_{\mathbb{R}^d} = \|\sum_{i=1}^d \mathbf{c}_i A\mathbf{p}_i\|_{\mathbb{R}^d} = \|\lambda\mathbf{x}\|_{\mathbb{R}^d} \geq \|x\|_{\mathbb{R}^d}.$$

This system may arise from a practical problem. Let $z(t, \mathbf{x}) : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ be continuous and let $B \in \mathbb{R}^{d\times d}$. We interpret the components of z as time and space, respectively. The partial derivative with respect to time and the boundary condition are given by,

$$\partial_t z(t, \mathbf{x}) = Bz(t, \mathbf{x}), \quad z(0, \mathbf{x}) = \mathbf{x}.$$

The solution to this ODE is given by $z(t, \mathbf{x}) = \exp(Bt)\,\mathbf{x}$. For a fixed $t = t_0 \in [0, \infty)$, the dynamical system $z(t_0, \mathbf{x})$ maps an input $\mathbf{x}$ forward in time by t_0 units. We get the map $z(t_0, \mathbf{x}) = \exp(Bt_0)\,\mathbf{x} = A\mathbf{x} = \varphi(\mathbf{x})$.

4.4.3 Identifiability for strictly positive kernels

It is desirable that the Koopman operator can uniquely be formed from the underlying dynamical system, a result we have seen holds true when acting between spaces of continuous functions on compact sets in Theorem 3.2.7. We will show that the composition map induced from a densely defined Koopman operator on a RKHS with a strictly positive kernel is injective. First, we define what it means for the Koopman operator to be densely defined.

Definition 4.4.7 (Densely defined)**.** Let k be a kernel and φ be map on $\mathcal{X}$. Define $\mathcal{D}(T_\varphi) := \{g \in H_k : g \circ \varphi \in H_k\}$, so that $T_\varphi : \mathcal{D}(T_\varphi) \to H_k$. We say that T_φ is *densely defined* on H_k if $\overline{\mathrm{span}}\,(\mathcal{D}(T_\varphi)) = H_k$.

Densely defined Koopman operators do not need to be bounded on H_k, but they are closed.

Lemma 4.4.8. If T_φ is a densely defined operator, T_φ is a closed operator.

Proof. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in H_k such that $\|f - f_n\|_{H_K} \to 0$ and $\|g-(T_\varphi f_n)\|_{H_k} \to 0$ as $n \to \infty$. To show that T_φ is closed, we need to show that $f \circ \varphi = g$. Let $x \in \mathcal{X}$, by the reproducing property,

$$g(x) = \langle g, k(x,\cdot)\rangle_{H_k} = \lim_{n\to\infty} \langle T_\varphi f_n, k(x,\cdot)\rangle_{H_k} = \lim_{n\to\infty} f_n(\varphi(x)) = \lim_{n\to\infty} \langle f_n, k(\varphi(x),\cdot)\rangle_{H_k} = f(\varphi(x)).$$

Since $x \in \mathcal{X}$ was arbitrary, we conclude that $g = T_\varphi f$. □

Theorem 4.4.9. Let $(\mathcal{X}, \varphi)$ and $(\mathcal{X}, \psi)$ maps on $\mathcal{X}$. Let k be strictly positive kernel on $\mathcal{X}$. Assume that T_φ and T_ψ are densely defined on H_k. Then, $\psi = \varphi$ if and only if $T_\psi = T_\varphi$.

Proof. Suppose that $\psi = \varphi$, then it is clear that $T_\psi = T_\varphi$. Conversely, suppose that $T_\psi = T_\varphi$. Let $h \in H_k$, and let $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{D}(T_\varphi)$ and $\mathcal{D}(T_\psi)$, respectively, converging to h. By Lemma 4.4.8, T_φ and T_ψ are closed. Therefore,

$$\lim_{n\to\infty} T_\varphi f_n = h \circ \varphi = h \circ \psi = \lim_{n\to\infty} T_\psi g_n.$$

Let $x \in \mathcal{X}$ and use the reproducing property,

$$\begin{aligned} h \circ \varphi(x) = h \circ \psi(x) \quad &\Longleftrightarrow \quad h(\psi(x)) - h(\varphi(x)) = 0, \\ &\Longleftrightarrow \quad \langle h, k(\psi(x),\cdot)\rangle_{H_k} - \langle h, k(\varphi(x),\cdot)\rangle_{H_k} = 0, \\ &\Longleftrightarrow \quad \langle h, k(\psi(x),\cdot) - k(\varphi(x),\cdot)\rangle_{H_k} = 0. \end{aligned}$$

By the definiteness of the inner product, we have $k(\psi(x),\cdot) = k(\varphi(x),\cdot)$. It follows from Lemma 2.2.8 that $\psi = \varphi$, as $x \in \mathcal{X}$ was arbitrary. □

Chapter 5

Conclusion

In this thesis, we gave an extensive overview of kernel methods combined with Koopman theory. We started by reviewing relevant RKHS theory, ending with vector-valued RKHS. We concluded the theory by presenting the scalar-valued representer theorem and an important isometry between a vector-valued RKHS and the class of Hilbert-Schmidt operators. Then, we took a more theoretical view of the Koopman operator between Banach spaces of continuous functions on compact sets. Hereafter, we posed the problem of estimating the properties of an unknown, complex, and perhaps chaotic dynamical system from observed snapshot data. The Koopman paradigm was introduced, where we motivated that estimating spectral properties of the Koopman operator encapsulate properties of the underlying dynamical system. The classic DMD algorithm, its extension EDMD, and kernel EDMD are explained. Then kEDMD and a recent development in the estimation of the Koopman operator through dynamical systems are put in the same framework and compared. We have shown that the "DMD-free" method in fact coincides with kEDMD for $\gamma = 0$. Lastly, we provide examples of RKHSs where the Koopman operator is bounded. For Gaussian RKHSs, we characterize dynamics for which it cannot be bounded and combine recently published results into a small Corollary 4.4.6.

By providing a broad review of existing literature, we were able to comment on various aspects of the Koopman operator. We presented the advantages of kernel methods but also laid out instances that limit their applicability.

A limitation of this work is that a more measure-theoretic or ergodic view is missing. Studying dynamical systems through these lenses and putting existing literature within this broadened framework would be an interesting path for future research.

Bibliography

[SH98] A. M.. Stuart and A.R. Humphries. *Dynamical systems and numerical analysis.* Cambridge University Press, 1998, p. 685. ISBN: 9780521645638.

[Koo31] B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space." In: *Proceedings of the National Academy of Sciences of the United States of America* 17.5 (May 1931), pp. 315–318. ISSN: 0027-8424. DOI: 10.1073/PNAS.17.5.315.

[Bru+21] Steven L. Brunton et al. "Modern Koopman Theory for Dynamical Systems". In: *SIAM Review* 64.2 (Feb. 2021), pp. 229–340. ISSN: 00361445. DOI: 10.1137/21M1401243. URL: https://arxiv.org/abs/2102.12086v2.

[GE22] Amin Ghadami and Bogdan I. Epureanu. "Data-driven prediction in dynamical systems: recent developments". In: *Philosophical Transactions of the Royal Society A* 380.2229 (2022), pp. 1429–1442. ISSN: 1364503X. DOI: 10.1098/RSTA.2021.0213. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2021.0213.

[Mez05] Igor Mezić. "Spectral properties of dynamical systems, model reduction and decompositions". In: *Nonlinear Dynamics* 41.1-3 (Aug. 2005), pp. 309–325. ISSN: 0924090X. DOI: 10.1007/S11071-005-2824-X/METRICS. URL: https://link.springer.com/article/10.1007/s11071-005-2824-x.

[Sch10] Peter J. Schmid. "Dynamic mode decomposition of numerical and experimental data". In: *Journal of Fluid Mechanics* 656 (2010), pp. 5–28. ISSN: 1469-7645. DOI: 10.1017/S0022112010001217. URL: https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/dynamic-mode-decomposition-of-numerical-and-experimental-data/AA4C763B525515AD4521A6CC5E10DBD4.

[Row+09] Clarence W. Rowley et al. "Spectral analysis of nonlinear flows". In: *Journal of Fluid Mechanics* 641 (2009), pp. 115–127. ISSN: 1469-7645. DOI: 10.1017/S0022112009992059. URL: https://www.cambridge.org/core/journals/journal-of-fluid-mechanics/article/spectral-analysis-of-nonlinear-flows/311041E1027AE7FEE7DDA36AC9AD4270.

[Bat+24] Pau Batlle et al. "Kernel methods are competitive for operator learning". In: *Journal of Computational Physics* 496 (Jan. 2024), p. 112549. ISSN: 0021-9991. DOI: 10.1016/J.JCP.2023.112549.

[Kos+22] Vladimir R. Kostic et al. "Learning Dynamical Systems via Koopman Operator Regression in Reproducing Kernel Hilbert Spaces". In: *Advances in Neural Information Processing Systems* 35 (May 2022). ISSN: 10495258. URL: https://arxiv.org/abs/2205.14027v4.

[HSS08] Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. "Kernel methods in machine learning". In: *https://doi.org/10.1214/009053607000000677* 36.3 (June 2008), pp. 1171–1220. ISSN: 0090-5364. DOI: 10.1214/009053607000000677. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-36/issue-3/Kernel-methods-in-machine-learning/10.1214/009053607000000677.full.

[Aro50] N. Aronszajn. "Theory of Reproducing Kernels". In: *Transactions of the American Mathematical Society* 68.3 (1950), p. 337. ISSN: 00029947. DOI: 10.2307/1990404.

[KW70] George S. Kimeldorf and Grace Wahba. "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines". In: *The Annals of Mathematical Statistics* 41.2 (Apr. 1970), pp. 495–502. ISSN: 0003-4851. DOI: 10.1214/AOMS/1177697089.

[KW71] George Kimeldorf and Grace Wahba. "Some results on Tchebycheffian spline functions". In: *Journal of Mathematical Analysis and Applications* 33.1 (Jan. 1971), pp. 82–95. ISSN: 0022-247X. DOI: 10.1016/0022-247X(71)90184-3.

[SHS01] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. "A generalized representer theorem". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2111 (2001), pp. 416–426. ISSN: 16113349. DOI: 10 . 1007 / 3 - 540 - 44581 - 1{\ _ }27 / COVER. URL: https://link.springer.com/chapter/10.1007/3-540-44581-1_27.

[Gon+24] Efrain Gonzalez et al. "The Kernel Perspective on Dynamic Mode Decomposition". In: *Transactions on Machine Learning Research* 122.5 (May 2024), pp. 553–557. ISSN: 2835-8856. URL: https://arxiv.org/abs/2106.00106.

[Köh+24] Frederik Köhne et al. "L-infinity-error bounds for approximations of the Koopman operator by kernel extended dynamic mode decomposition". In: (Mar. 2024). URL: https://arxiv.org/abs/2403.18809v2.

[Mer09] James Mercer. "Functions of positive and negative type, and their connection the theory of integral equations". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209.441-458 (Jan. 1909), pp. 415–446. ISSN: 0264-3952. DOI: 10.1098/RSTA.1909.0016. URL: https://royalsocietypublishing.org/doi/10.1098/rsta.1909.0016.

[Sch00] B. Scholkopf. "The Kernel Trick for Distances". In: *Advances inNeural Information Processing Systems* 13 (2000). URL: https://is.mpg.de/fileadmin/user_upload/files/publications/tr-2000-51_1839%5B0%5D.pdf.

[SC08] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Ed. by Bernhard Scholkopf, Michael Jordan, and Jon Kleinberg. Springer New York, NY, 2008, pp. 124–132. ISBN: 978-0-387-77241-7. DOI: https://doi.org/10.1007/978-0-387-77242-4. URL: www.springer.com/series/3816.

[Wai19] Martin J. Wainwright. "Reproducing kernel Hilbert spaces". In: *High-Dimensional Statistics*. Cambridge University Press, Feb. 2019, pp. 383–415. DOI: 10 . 1017 / 9781108627771 . 012. URL: https://www.cambridge.org/core/books/highdimensional-statistics/reproducing-kernel-hilbert-spaces/7592A2B68EBB6F57B0AD042F8E812BEB.

[Sai16] Saburou Saitoh. "Theory of Reproducing Kernels". In: vol. 44. Developments in Mathematics. Tokyo: Springer Singapore, 2016, pp. 135–150. ISBN: 978-981-10-0529-9. DOI: 10 . 1007 / 978 - 1 - 4757 - 3741 - 7{\ _ }10. URL: http://link.springer.com/10.1007/978-981-10-0530-5.

[PR16] Vern I. Paulsen and Mrinal Raghupathi. "An introduction to the theory of reproducing kernel Hilbert spaces". In: *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces* (Apr. 2016), pp. 1–182. DOI: 10.1017/CBO9781316219232. URL: https://www.cambridge.org/core/books/an-introduction-to-the-theory-of-reproducing-kernel-hilbert-spaces/C3FD9DF5F5C21693DD4ED812B531269A.

[Car+10] C. Carmeli et al. "Vector valued reproducing kernel Hilbert spaces and universality". In: *https://doi.org/10.1142/S0219530510001503* 8.1 (Nov. 2010), pp. 19–61. ISSN: 02195305. DOI: 10.1142/S0219530510001503.

[MK20] Mattes Mollenhauer and Péter Koltai. "Nonparametric approximation of conditional expectation operators". In: (Dec. 2020). URL: https://arxiv.org/abs/2012.12917v3.

[Mez15] Igor Mezić. "On applications of the spectral theory of the Koopman operator in dynamical systems and control theory". In: *Proceedings of the IEEE Conference on Decision and Control* 54rd IEEE, CDC (Feb. 2015), pp. 7034–7041. ISSN: 25762370. DOI: 10.1109/CDC.2015.7403328. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7403328.

[BMM12] Marko Budišić, Ryan Mohr, and Igor Mezić. "Applied Koopmanism". In: *Chaos (Woodbury, N.Y.)* 22.4 (Oct. 2012). ISSN: 1089-7682. DOI: 10.1063/1.4772195. URL: https://pubmed.ncbi.nlm.nih.gov/23278096/.

[Mez94] I. Mezić. "On the geometrical and statistical properties of dynamical systems : theory and applications". In: (1994). DOI: 10.7907/SM03-1C46.

[Tu+14] Jonathan H. Tu et al. "On dynamic mode decomposition: Theory and applications". In: *Journal of Computational Dynamics* 1.2 (Dec. 2014), pp. 391–421. ISSN: 2158-2491. DOI: 10.3934/JCD.2014.1.391. URL: https://www.aimsciences.org/article/doi/10.3934/jcd.2014.1.391.

[WKR15] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. "A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition". In: *Journal of Nonlinear Science* 25.6 (Dec. 2015), pp. 1307–1346. ISSN: 14321467. DOI: 10.1007/S00332-015-9258-5/FIGURES/14. URL: https://link.springer.com/article/10.1007/s00332-015-9258-5.

[Wil+15] Matthew O. Williams et al. "A kernel-based method for data-driven koopman spectral analysis". In: *Journal of Computational Dynamics* 2.2 (Dec. 2015), pp. 247–265. ISSN: 2158-2491. DOI: 10.3934/JCD.2015005. URL: https://www.aimsciences.org/en/article/doi/10.3934/jcd.2015005%20https://www.aimsciences.org/en/article/doi/10.3934/jcd.2015005?viewType=HTML%20https://www.aimsciences.org/article/doi/10.3934/jcd.2015005.

[Col23] Matthew J. Colbrook. "The Multiverse of Dynamic Mode Decomposition Algorithms". In: *Handbook of Numerical Analysis* 25 (Nov. 2023), pp. 127–230. ISSN: 15708659. DOI: 10.1016/bs.hna.2024.05.004. URL: https://arxiv.org/abs/2312.00137v2.

[Son+09] Le Song et al. "Hilbert space embeddings of conditional distributions with applications to dynamical systems". In: *ACM International Conference Proceeding Series* 382 (2009). DOI: 10.1145/1553374.1553497. URL: https://dl.acm.org/doi/10.1145/1553374.1553497.

[MP05] Charles A. Micchelli and Massimiliano Pontil. "On learning vector-valued functions". In: *Neural Computation* 17.1 (Jan. 2005), pp. 177–204. ISSN: 08997667. DOI: 10.1162/0899766052530802.

[SK79] R. K. Singh and Ashok Kumar. "Compact composition operators". In: *Journal of the Australian Mathematical Society* 28.3 (1979), pp. 309–314. ISSN: 1446-8107. DOI: 10.1017/S1446788700012258. URL: https://www.cambridge.org/core/journals/journal-of-the-australian-mathematical-society/article/compact-composition-operators/E376E24E70E5455230581295C473CE78.

[Phi+24] Friedrich Philipp et al. "Invariance of Gaussian RKHSs under Koopman operators of stochastic differential equations with constant matrix coefficients". In: *arXiv* (May 2024), arXiv:2405.14429. DOI: 10.48550/ARXIV.2405.14429. URL: https://ui.adsabs.harvard.edu/abs/2024arXiv240514429P/abstract.

[SHS06] Ingo Steinwart, Don Hush, and Clint Scovel. "An explicit description of the reproducing Kernel Hilbert spaces of Gaussian RBF kernels". In: *IEEE Transactions on Information Theory* 52.10 (Oct. 2006), pp. 4635–4643. ISSN: 00189448. DOI: 10.1109/TIT.2006.881713.

Appendix A

Functional Analysis

A.1 Selection of results

Theorem A.1.1 (Interpolation theorem)**.** Let H be a Hilbert space, $\mathcal{X}$ a non-empty set and $c \geq 0$. Consider the maps $T : \mathcal{X} \to \text{ran}(T) \subseteq H$, with $T(x) := f_x$ and assume that $\overline{\text{span}}\{\text{ran}(T)\} = H$. Then, the following statements are equivalent

- There exists an element $g \in H$ such that $\|g\| \leq c$ and $\langle f_x, g\rangle = \alpha_x$ for any $x \in X$
- For any $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{b} \in \mathbb{C}^n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i \overline{\mathbf{b}}_j \alpha_{\mathbf{x}_i}\overline{\alpha_{\mathbf{x}_j}} \leq c^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i \overline{\mathbf{b}}_j \langle f_{\mathbf{x}_i}, f_{\mathbf{x}_j}\rangle \tag{A.1}$$

Proof. If such an element $g \in H$ exists, by Cauchy-Schwarz

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i \overline{\mathbf{b}}_j \alpha_{\mathbf{x}_i}\overline{\alpha_{\mathbf{x}_j}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \langle \mathbf{b}_i f_{\mathbf{x}_i}, g\rangle \overline{\langle \mathbf{b}_j f_{\mathbf{x}_j}, g\rangle} = \langle \sum_{i=1}^{n} \mathbf{b}_i f_{\mathbf{x}_i}, g\rangle \overline{\langle \sum_{j=1}^{n} \mathbf{b}_j f_{\mathbf{x}_j}, g\rangle} &= \left|\langle \sum_{i=1}^{n} \mathbf{b}_i f_{\mathbf{x}_i}, g\rangle\right|^2, \\ &\leq \|g\|^2 \|\sum_{i=1}^{n} \mathbf{b}_i f_{\mathbf{x}_i}\|^2, \\ &\leq c^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{b}_i \overline{\mathbf{b}_j} \langle f_{\mathbf{x}_i}, f_{\mathbf{x}_j}\rangle.\end{aligned}$$

Conversely, assume that equation (A.1) holds. Define $S : \text{span}(\text{ran}(T)) \to \mathbb{C}$ as $S(\sum_{i=1}^{n} \mathbf{b}_i f_{\mathbf{x}_i}) = \sum_{i=1}^{n} \mathbf{b}_i \alpha_i$. This defines a well defined linear functional if the elements f_x are linearly independent, with norm bounded above by c. By taking the closure, this bounded linear functional can be extended to H and we conclude by Riesz-Fréchet that there exists a unique $g \in H$ such that $S(f_x) = a_x = \langle f_x, g\rangle$ for any $f_x \in \text{ran}(T)$.

□

Theorem A.1.2 (Direct sum of Hilbert spaces)**.** Let $(H_i, \langle \cdot, \cdot\rangle_{H_i})_{i=1}^{M}$ be a sequence of Hilbert spaces. Then, the set

$$\bigoplus_{i=1}^{M} H_i := \left\{ h \in \prod_{i=1}^{M} H_i \ : \ \sum_{i=1}^{M} \|h_i\|_{H_i}^2 < \infty \right\}$$

equipped with the inner product

$$\langle f, g\rangle_{\bigoplus_{i=1}^{M} H_i} := \sum_{i=1}^{M} \langle f_i, g_i\rangle_{H_i}$$

is a Hilbert space. This Hilbert space is called the *direct sum of* $(H_i)_{i=1}^M$

Proof. The sesquilinearity and the definiteness of the inner product $\langle \cdot, \cdot \rangle_{\bigoplus_{i=1}^M H_i}$ readily follow from the inner products on H_i. Let $(f^{(n)})_{n\in\mathbb{N}}$ be a Cauchy sequence in $\bigoplus_{i=1}^M H_i$. Since $\|\cdot\|_{\bigoplus_{i=1}^M H_i}$ is stronger than each $\|\cdot\|_{H_i}$ we have that $(f_i^{(n)})_{n\in\mathbb{N}}$ is Cauchy in H_i for $1 \leq i \leq M$ and therefore has a limit $\lim_{n\to\infty} f_i^{(n)} = f_i$. Define $f = (f_i)_{i=1}^M$, then $\|f - f^{(n)}\|^2_{\bigoplus_{i=1}^M H_i} = \lim_{m\to\infty} \sum_{i=1}^M \|f_i^{(m)} - f_i^{(n)}\|^2_{H_i} \to 0$. Hence, there exists an $m \in \mathbb{N}$ such that $\|f\|_{\bigoplus_{i=1}^M H_i} \leq \|f - f^{(m)}\|_{\bigoplus_{i=1}^M H_i} + \|f^{(m)}\|_{\bigoplus_{i=1}^M H_i} < \varepsilon + \|f^{(m)}\|_{\bigoplus_{i=1}^M H_i} < \infty$ and $f \in \bigoplus_{i=1}^M H_i$. □