SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Machine Learning in Kreĭn Spaces: An Exposition to the Main Representer Theorems

av

Andreas Karlsson

2025 - No K36

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Machine Learning in Kreĭn Spaces: An Exposition to the Main Representer Theorems

Andreas Karlsson

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Annemarie Luger

2025

Abstract

Kernel methods are powerful in machine learning because they enable non-linear learning with simpler linear algebra tools. Yet most theory is built around positive semidefinite kernels and Hilbert-space geometry. This expository thesis explains how learning with *indefinite* kernels can be made rigorous and tractable in reproducing kernel Kreĭn spaces (RKKS). The thesis builds a concise pathway from the geometry of Kreĭn spaces (fundamental decomposition and symmetry, strong topology via the associated Hilbert space) to reproducing kernels, where evaluations are continuous and the kernel-induced form coincides with the RKKS inner product. On this foundation, we synthesize two RKKS formulations of supervised learning: (1) a stabilization setup with a *weak representer theorem*, where stationary solutions lie in the data span, and (2) a variance-constrained minimization setup with a *strong representer theorem*, where optimizers lie in the span under strong-topology Tikhonov regularization. In both cases, the representer theorems make the infinite-dimensional problems collapse to finite-dimensional Gram-matrix algebra, introducing the kernel trick in the indefinite setting. A limitation of this setup is that naïve objectives can be unbounded below.

Sammanfattning

Kärnmetoder är kraftfulla verktyg för maskininlärning eftersom de översätter icke-linjär inlärning till en behändigare uppsättning av linjär-algebraiska verktyg. Ändå utgörs den huvudsakliga delen av teorin av positivt semidefinita kärnor och Hilbertrumsgeometri. Denna uppsats förklarar hur inlärning med *indefinita* kärnor kan göras på ett rigoröst och hanterbart sätt i Kreĭnrum med reproducerande kärna (RKKS). Uppsatsen utformar en väg från Kreĭn-rummens geometri (den fundamentala dekompositionen och symmetrin samt den starka topologin via det associerade Hilbertrummet) till reproducerande kärnor, där avbildningarna är kontinuerliga och formen som induceras av kärnan sammanfaller med den inre produkten i RKKS. Vi presenterar två RKKS-baserade problemformuleringar av övervakad inlärning: (1) en stabiliseringsansats med en *svag representersats*, där stationära lösningar ligger i spannet av kärnsektionerna, och (2) en variansbegränsad minimeringsansats med en *stark representersats*, där optimala lösningar ligger i spannet i den starka topologin. I båda fallen används representersatserna för att kollapsa de oändligdimensionella problemen till ändligdimensionell Gram-matris-algebra, vilket också introducerar kärntricket till det indefinita sammanhanget. En begränsning är att målfunktioner kan vara obegränsade nedåt om konstruerade utan eftertanke.

Acknowledgements

I am grateful to my supervisor, Annemarie Luger, for introducing me to the theory of indefinite inner products and Kreĭn spaces, and for her valuable feedback during the development of this thesis. I dedicate this thesis to my nephews — August, Ludvig, and Walter.

Contents

1 Introduction

Kernel methods are among the most successful tools in supervised learning because they allow nonlinear prediction with linear-algebraic machinery. In this survey, we present and synthesize the two main representer theorems central to supervised learning in indefinite inner product spaces. The motivation for this exposition is to illustrate the central role of kernels in machine learning problems, and to depart from the usual positive semidefinite kernels and Hilbert spaces commonly assumed for the theoretical setting. We ask which theoretical results give kernels their central role in learning theory, and why kernel methods are beneficial for supervised learning problems. To this end, we will explain how the geometry of Kreĭn spaces, the theory of reproducing kernels, and empirical risk minimization fit together into a coherent learning framework. As a result, this is an expository thesis: it claims no new results; rather, its contribution is a concise pathway from the basic definitions of indefinite inner products all the way up to a unified presentation of supervised learning in reproducing kernel Kreĭn spaces (RKKS). Along the way, we will clarify which functional–analytic ingredients are essential in this indefinite setting and why certain learning formulations must be adapted relative to the standard Hilbert–space case.

On the geometric side, in terms of scope, we pass from inner product spaces to Kreĭn spaces via the so-called fundamental decomposition and the fundamental symmetry J, and use the associated Hilbert space to induce the strong topology. We then define reproducing kernel Kreĭn spaces and their reproducing property. For a given RKKS, the reproducing kernel is unique, evaluation functionals are continuous by construction, and the RKKS inner product equals the kernel–induced pairing; thus the geometry and the kernel coincide. Conversely, any admissible Hermitian kernel that decomposes in this way generates (at least one) RKKS. The topic of functional analysis extends much further beyond what is covered here, but we restrict ourselves in scope to the necessary machinery in order to study learning problems.

On the learning side, we formulate two of the main supervised learning problems in reproducing kernel Kreĭn spaces. The first is a stabilization setup which induces a weak representer theorem in which stationary solutions to the problems lie in the data span. The second is a minimization setup with a strong representer theorem in which optimizers lie in the span with strong-topology Tikhonov regularization of the associated Hilbert space, and a variance constraint. The importance of these results is hard to overstate. In both the stabilization and minimization settingsl the kernel trick is prevalent. The infinite-dimensional search over a functional space reduces to finite-dimensional Gram-matrix algebra over real-valued coefficient vectors. Finding the optimal function in an RKKS collapses to finding scalars, which is a considerably more tractable problem both theoretically and practically. However, practitioners must still be mindful of how they formulate their learning problems. The indefinite nature of the RKKS inner product can render such problems ill-posed, for example, the minimization problem may be unbounded from below. This section of the thesis is limited to the theory of supervised learning problems. As a result, unsupervised learning problems and detailed practical considerations of machine learning are outside of the scope of the survey.

Throughout, we assume familiarity with linear algebra (e.g., vector spaces, matrices) and foundational analysis concepts (e.g., completeness, norms, continuity), together with elementary probability theoretical notions used in empirical risk minimization. These prerequisites should make the survey accessible to readers with an undergraduate degree in mathematics. Section 2 will then supply the functional-analytic background needed to study the supervised learning setup considered in Section 3.

Specifically, Section 2 develops the functional–analytic foundation: Kreĭn spaces, associated Hilbert spaces, RKKS, reproducing kernels, and the role of Gram matrices. The main references used for this section are authored by Aurelian Ghondea [Ghe22, Ghe13]. Section 3 formulates supervised learning in RKKS, presents the weak and strong representer theorems, and derives coefficient–space problems in terms of Gram matrices. The main literature used for this section consists of articles by Ong et al. and Oglic–Gärtner respectively [OMCS04, OG18]. Finally, we provide concluding remarks in Section 4.

2 Kreĭn Spaces and Reproducing Kernels

This chapter develops the machinery for working with indefinite inner products and their reproducing kernels. We begin by fixing notation for inner product spaces and describing how the sign distribution of Gram eigenvalues reflects the geometry in Section 2.1. We then introduce Kreĭn spaces in Section 2.2 via their fundamental decomposition, and then state the Riesz-Fréchet representation theorem for Kreĭn spaces. Finally, we define reproducing kernel Kreĭn spaces (RKKS) in Section 2.3, and further detail the reproducing property of reproducing kernels, and record some results regarding the existence and uniqueness of an RKKS. The chapter is self-contained and provides the functional-analytic foundation needed to study supervised learning problems in RKKS, which are covered in Section 3. We assume familiarity with foundational linear algebra concepts such as vector spaces, matrices, and subspaces, as well as a grasp of basic real and complex analysis concepts, for example, completeness, the supremum, and the real field $\mathbb{R}$ and the complex field $\mathbb{C}$.

2.1 Inner Product Spaces

Here we fix notation for (possibly indefinite) inner products, orthogonality, indices, and Gram matrices, which are used to analyze Kreĭn spaces later on. The primary source for this exposition is from Gheondea [Ghe22].

Definition 2.1 (Indefinite Inner product space [Ghe22, Chapter 1])**.** Let $\mathcal{X}$ be a complex vector space. A mapping $[\cdot,\cdot]_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is an *inner product* if it satisfies the following two properties:

(a) **Sesquilinearity.** The inner product is linear in the first variable, i.e. $[\alpha x + \beta y,\, z]_{\mathcal{X}} = \alpha[x,z]_{\mathcal{X}} + \beta[y,z]_{\mathcal{X}}$ and conjugate linear in the second: $[x,\, \alpha y + \beta z]_{\mathcal{X}} = \overline{\alpha}[x,y]_{\mathcal{X}} + \overline{\beta}[x,z]_{\mathcal{X}}$ for all $x, y, z \in \mathcal{X}$ and $\alpha, \beta \in \mathbb{C}$.

(b) **Hermitian symmetry.** The inner product is conjugate symmetric: $[x,y]_{\mathcal{X}} = \overline{[y,x]_{\mathcal{X}}}$ for all $x, y \in \mathcal{X}$.

Here $\overline{\alpha}$ denotes the complex conjugate of α. We call $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$ an *indefinite inner product space*, and the indefinite inner product $[\cdot,\cdot]_{\mathcal{X}}$ is sometimes referred to as a *form*.

Let $\mathcal{X}$ be an inner product space. For any vector $x \in \mathcal{X}$ we call $[x,x]_{\mathcal{X}}$ the *square* of x. A nonzero vector is positive if $[x,x]_{\mathcal{X}} > 0$, negative if $[x,x]_{\mathcal{X}} < 0$, or isotropic if $[x,x]_{\mathcal{X}} = 0$. The inner product space $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$ is *indefinite* if both positive and negative vectors exist. The space is *positive semidefinite (PSD)* if $[x,x]_{\mathcal{X}} \geq 0$ for all $x \in \mathcal{X}$ and it is *positive definite* if the inequality is strict for $x \neq 0$; negative (semi)definiteness is defined analogously [Ghe22].

By a *linear manifold* $\mathcal{A}$ we mean a non-empty subset of an inner product space $\mathcal{X}$ that is closed under vector addition and scalar multiplication, i.e. $\alpha x + \beta y \in \mathcal{A}$ for any $\alpha, \beta \in \mathbb{C}$ and any $x, y \in \mathcal{A}$. A linear manifold $\mathcal{A} \subseteq \mathcal{X}$ is positive (negative) if and only if the induced inner product space $(\mathcal{A}, [\cdot,\cdot]_{\mathcal{A}})$ is positive (negative) semidefinite. We use the terminology *strictly positive* (*strictly negative*) for the manifold if the corresponding inner product spaces are definite. A closed linear manifold is called a *subspace*, once a *topology* on $\mathcal{X}$ is determined (see Section 2.2). Two vectors $x, y \in \mathcal{X}$ are *orthogonal*, denoted by $x \perp y$, if $[x,y]_{\mathcal{X}} = 0$. Similarly, two manifolds $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ are orthogonal if $x \perp y$ for all $x \in \mathcal{A}$ and $y \in \mathcal{B}$. The *orthogonal companion* of a subset $\mathcal{A} \subseteq \mathcal{X}$ is the manifold $\mathcal{A}^{\perp} \subseteq \mathcal{X}$ for which

$$\mathcal{A}^{\perp} = \{\, x \in \mathcal{X} : x \perp y,\, y \in \mathcal{A} \,\}. \tag{2.1.1}$$

The *isotropic part* of a linear manifold $\mathcal{A} \subseteq \mathcal{X}$ is defined as its intersection with its companion, that is, $\mathcal{A}^0 = \mathcal{A} \cap \mathcal{A}^{\perp}$. If $\mathcal{A}^0 = \{0\}$ the manifold $\mathcal{A}$ is called *nondegenerate*, that is, if $[x,y]_{\mathcal{X}} = 0$ for all $y \in \mathcal{A}$ implies $x = 0$ for every $x \in \mathcal{A}$. Note in particular that $\mathcal{X}^0 = \mathcal{X}^{\perp}$, which is called the *radical* of $\mathcal{X}$. For nonempty manifolds $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ we define the *sum* of the linear manifolds by $\mathcal{A} + \mathcal{B} = \{a + b : a \in \mathcal{A},\, b \in \mathcal{B}\}$. That is, the manifold spanned by the vectors in $\mathcal{A} \cup \mathcal{B}$. Furthermore, if $\mathcal{A} \cap \mathcal{B} = \{0\}$ with respect to a positive definite inner product, then this sum is called a *direct sum* and we denote it as $\mathcal{A} \oplus \mathcal{B}$. If the inner product is indefinite, and in addition, $\mathcal{A} \perp \mathcal{B}$, then we call it an *orthogonal* direct sum and denote it by $\mathcal{A}\,[\oplus]\,\mathcal{B}$. Note that the notation for orthogonal direct sums in indefinite product spaces deviates somewhat from the standard Hilbert space (defined in Section 2.2) notation. An inner product space $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$ admits a *fundamental decomposition* if there exist two strictly positive and strictly negative orthogonal subspaces $\mathcal{P}, \mathcal{N} \subseteq \mathcal{X}$ such that the following orthogonal direct sum decomposition of the space holds

$$\mathcal{X} = \mathcal{P}\,[\oplus]\,\mathcal{X}^0\,[\oplus]\,\mathcal{N}. \tag{2.1.2}$$

For a linear manifold $\mathcal{A} \subseteq \mathcal{X}$ we define the *algebraic dimension* $\dim(\mathcal{A})$ as the largest number of linearly independent vectors contained in $\mathcal{A}$. If $\mathcal{A}$ contains linearly independent sets of arbitrarily large finite size, then we assign $\dim(\mathcal{A}) = \infty$. We use the algebraic dimension of the manifolds to define the indices of the inner product space [Ghe22].

Definition 2.2 (Indices and Inertia [Ghe22, Chapter 1])**.** Let $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$ be an inner product space. The *positive index* κ_+, *negative index* κ_-, and the isotropic index κ_0 of $\mathcal{X}$ are defined by

$$\begin{aligned}\kappa_+(\mathcal{X}) &= \sup\{\, \dim \mathcal{A} : \mathcal{A} \subseteq \mathcal{X} \text{ is positive definite} \,\}\\ \kappa_-(\mathcal{X}) &= \sup\{\, \dim \mathcal{A} : \mathcal{A} \subseteq \mathcal{X} \text{ is negative definite} \,\}\\ \kappa_0(\mathcal{X}) &= \dim \mathcal{X}^0.\end{aligned} \tag{2.1.3}$$

The ordered triple $(\kappa_-(\mathcal{X}),\ \kappa_0(\mathcal{X}),\ \kappa_+(\mathcal{X}))$ is called the *inertia* of the form, and the ordered pair $(\kappa_-(\mathcal{X}), \kappa_+(\mathcal{X}))$ is called the *signature* of $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$. We simply write $\kappa_-, \kappa_+, \kappa_0$ when the inner product space is apparent from the context.

Note that if an inner product space $\mathcal{X}$ is positive-semidefinite then $\kappa_- = 0$ (and vice versa), whereas if $\mathcal{X}$ is indefinite, then both $\kappa_\pm \neq 0$. The isotropic index κ_0 of $\mathcal{X}$ equals the dimension of the radical $\mathcal{X}^0$ and vanishes precisely when the form is nondegenerate. In that case, the fundamental decomposition (2.1.2) simplifies to $\mathcal{X} = \mathcal{P} [\oplus] \mathcal{N}$. Hence, from this point onward, if a signature (κ_-, κ_+) is mentioned, it is understood that the underlying inner product on $\mathcal{X}$ is nondegenerate. In this case, we say that the inner product space is *orthocomplemented* by the manifolds $\mathcal{P}$ and $\mathcal{N}$. A manifold $\mathcal{N}$ is the orthocomplement to the manifold $\mathcal{P}$ if $\mathcal{P} \perp \mathcal{N}$ and $\mathcal{P} \cap \mathcal{N} = \{0\} = \mathcal{X}^0$, and this is summarized by $\mathcal{P} [\oplus] \mathcal{N} = \mathcal{X}$. Furthermore, if $\mathcal{X} = \mathcal{P} [\oplus] \mathcal{N}$ is orthocomplemented, then for all $x \in \mathcal{X}$ there exist unique vectors $x_1 \in \mathcal{P}$ and $x_2 \in \mathcal{N}$ such that $x = x_1 + x_2$ [Ghe22].

Define the *quadratic form* $Q : \mathcal{X} \to \mathbb{C}$ as the mapping from each vector to its square; that is, $Q(x) = [x, x]_{\mathcal{X}}$. Let $i^2 = -1$; then by the *polarization identity*

$$[x, y]_{\mathcal{X}} = \frac{1}{4} \sum_{k=0}^{3} i^k Q(x + i^k y). \tag{2.1.4}$$

Thus, the inner product is uniquely determined by the quadratic form. The matrix representation of the quadratic form Q carries a special name [Ghe22].

Definition 2.3 (Gram matrix)**.** Let $(\mathcal{X}, [\cdot,\cdot]_{\mathcal{X}})$ be an inner product space and let $(x_i)_{i=1}^n$ denote a finite *ordered n-tuple* of elements in $\mathcal{X}$ where $i = 1, \ldots, n$. The

Gram matrix G on $(x_i)_{i=1}^n$ is defined as

$$G \;:=\; \Big([x_i, x_j]_{\mathcal{X}}\Big)_{i,j=1}^{n} \in \mathbb{C}^{n\times n}. \tag{2.1.5}$$

The quadratic form Q restricted to $\operatorname{span}\{x_1, \dots, x_n\}$ can be represented as $Q(x) = [x, x]_{\mathcal{X}} = \overline{\alpha}^\top G\alpha$ for some vector of scalars $\alpha \in \mathbb{C}^n$ and $x = \sum_i \alpha_i x_i$.

Note that because the indefinite inner product $[\cdot, \cdot]_{\mathcal{X}}$ is Hermitian, this also means that the Gram matrix is Hermitian, i.e., $G = \overline{G}^\top$, where $\overline{G}^\top$ denotes the *conjugate transpose* of G. Although the Gram matrix is a finite-dimensional representation of the (possibly infinite-dimensional) quadratic form on an inner product space $\mathcal{X}$, the sign distribution of the eigenvalues over *all* the finite samples $(x_i)_{i=1}^n \subseteq \mathcal{X}$ is inextricably linked to the geometry of $\mathcal{X}$.

Theorem 2.4 ([Ghe22, Prop. 1.1.16])**.** *For an inner product space $(\mathcal{X}, [\cdot, \cdot])$, the positive (negative) index $\kappa_+(\mathcal{X})$ $(\kappa_-(\mathcal{X}))$ is equal to the supremum of the number of positive (negative) eigenvalues of the Gram matrices G associated with all finite ordered tuples $(x_i)_{i=1}^n$ with elements from $\mathcal{X}$, counted with multiplicity*

$$\kappa_\pm(\mathcal{X}) = \sup_{\substack{n\geq 1\\ (x_i)_{i=1}^n\subseteq\mathcal{X}}} \left\{\textit{Number of } \pm \textit{ eigenvalues of } G\right\}. \tag{2.1.6}$$

Proof. The proof falls outside the scope of the thesis. See [Ghe22], Prop. 1.1.16 (Page 7) for additional details. $\square$

This result asserts that the signature (κ_-, κ_+) is recoverable from the distribution of the signs on the eigenvalues of its finite Gram matrices. The theorem also illustrates the strong link between the inner product space $(\mathcal{X}, [\cdot, \cdot]_{\mathcal{X}})$, the quadratic form Q, and the Gram matrix G. Similarly, the inertia $(\kappa_-, \kappa_0, \kappa_+)$ also describes the number of negative, null, and positive squares of Q, respectively.

2.2 Kreĭn Spaces

This section introduces Kreĭn spaces using their fundamental decomposition, the fundamental symmetry J, and the associated Hilbert space topology, culminating in a representation theorem in Kreĭn spaces. By a **Hilbert space** $(\mathcal{H}, \langle\cdot, \cdot\rangle_{\mathcal{H}})$ we mean a vector space with a positive-definite inner product that is *complete* (every Cauchy sequence converges in $\mathcal{H}$) with respect to the induced *norm* $\|x\|_{\mathcal{H}} := \sqrt{\langle x, x\rangle_{\mathcal{H}}}$.

Whenever we use the notation $\langle \cdot, \cdot \rangle$, it should be understood that the inner product is positive semidefinite [Ghe22].

Definition 2.5 (Kreĭn Space [Ghe22, Section 1.4])**.** A **Kreĭn space** $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ is a nondegenerate indefinite inner product space for which there exist subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ satisfying

1. The space admits a *fundamental decomposition* $\mathcal{K} = \mathcal{H}_+ [\oplus] \mathcal{H}_-$. That is, every $f \in \mathcal{K}$ can be written uniquely as $f = f_+ + f_-$ with $f_\pm \in \mathcal{H}_\pm$, and $\mathcal{H}_+$ and $\mathcal{H}_-$ are $[\cdot,\cdot]_{\mathcal{K}}$-orthogonal.
2. The form $[\cdot,\cdot]_{\mathcal{K}}$ is positive definite on $\mathcal{H}_+$ and negative definite on $\mathcal{H}_-$, i.e.
$$[f_+, f_+]_{\mathcal{K}} > 0 \quad \text{for all } 0 \neq f_+ \in \mathcal{H}_+, \quad [f_-, f_-]_{\mathcal{K}} < 0 \quad \text{for all } 0 \neq f_- \in \mathcal{H}_-.$$
3. With the inner products
$$\langle f_+, g_+ \rangle_{\mathcal{H}_+} := [f_+, g_+]_{\mathcal{K}} \quad \text{and} \quad \langle f_-, g_- \rangle_{\mathcal{H}_-} := -[f_-, g_-]_{\mathcal{K}},$$
the spaces $(\mathcal{H}_+, \langle \cdot,\cdot \rangle_{\mathcal{H}_+})$ and $(\mathcal{H}_-, \langle \cdot,\cdot \rangle_{\mathcal{H}_-})$ are Hilbert spaces.

The fundamental decomposition $\mathcal{H}_+ [\oplus] \mathcal{H}_-$ of a Kreĭn space is not necessarily unique, but the positive and negative indices are innate to the Kreĭn space. For any fundamental decomposition $\mathcal{K} = \mathcal{H}_+ [\oplus] \mathcal{H}_-$ one has $\kappa_\pm = \dim \mathcal{H}_\pm$, and hence the signature (κ_+, κ_-) is independent of the choice of fundamental decomposition. A Kreĭn space with finite negative index $\kappa_- < \infty$ is called a **Pontryagin space** and is denoted $(\Pi_\kappa, [\cdot,\cdot]_\Pi)$. Similarly, if $\mathcal{K}$ is finite-dimensional then the Kreĭn space is said to be *Minkowski* [OMCS04, Chapter 2.1]. Studies on statistical learning problems in Kreĭn spaces are often restricted to Pontryagin spaces for tractability (see e.g. [OMCS04, OG18]). Let $\mathcal{X}$ and $\mathcal{Y}$ be complex vector spaces. A mapping $T : D \to \mathcal{Y}$ is a *linear operator* if the domain D is a linear manifold of $\mathcal{X}$, and T is linear, i.e., $T(\alpha x + \beta y) = \alpha Tx + \beta Ty$ for $x, y \in D$ and $\alpha, \beta \in \mathbb{C}$. Unless otherwise stated, an *operator* is assumed to mean a linear operator.

Due to the indefinite nature of the Kreĭn inner product, we cannot introduce a norm topology as we would in the Hilbert case. That is, $\|x\| = \sqrt{[x,x]_{\mathcal{K}}}$ does not necessarily lead to a valid norm. In particular, $[x,x]_{\mathcal{K}}$ may be negative. Instead, we first fix our choice of fundamental decomposition $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$. Every $x \in \mathcal{K}$ can be written uniquely as $x = x_+ + x_-$ with $x_\pm \in \mathcal{H}_\pm$. We define the **fundamental symmetry** $J : \mathcal{K} \to \mathcal{K}$ as the linear operator $J(x_+ + x_-) := x_+ - x_-$. Then $J^2 = I$, and the equality

$$\langle x, y\rangle_J = [Jx, y]_{\mathcal{K}}, \qquad \text{for all } x, y \in \mathcal{K}, \tag{2.2.1}$$

defines a positive definite inner product. The pair $\mathcal{K}$ and $\langle x, y\rangle_J$ define the associated Hilbert space of $\mathcal{K}$ [Ghe22].

Definition 2.6 (Associated Hilbert Space [OMCS04, Definition 3])**.** Fix a fundamental decomposition $\mathcal{K} = \mathcal{H}_- [\oplus] \mathcal{H}_+$ of a Kreĭn space and let J denote its corresponding fundamental symmetry. The *associated Hilbert space* $(\mathcal{H}_{\mathcal{K}}, \langle\cdot,\cdot\rangle_J)$ is the same underlying vector space, i.e. $\mathcal{H}_{\mathcal{K}} := \mathcal{K}$, equipped with the inner product

$$\langle f, g\rangle_J = \langle f_+, g_+\rangle_{\mathcal{H}_+} + \langle f_-, g_-\rangle_{\mathcal{H}_-}, \quad \text{where } f_\pm, g_\pm \in \mathcal{H}_\pm. \tag{2.2.2}$$

Because $\mathcal{H}_{\mathcal{K}}$ is a Hilbert space, there exists a well-defined norm

$$\|x\|_J := \sqrt{\langle x, x\rangle_J} = \sup_{\substack{y\in\mathcal{K} \\ \|y\|_J \le 1}} |[x, y]|_{\mathcal{K}} \quad \text{for all } x \in \mathcal{K}, \tag{2.2.3}$$

called the **fundamental norm** of the Kreĭn space induced by J [Ghe22]. Writing $x = x_+ + x_-$ with $x_\pm \in \mathcal{H}_\pm$ and using that $J(x_+ + x_-) = x_+ - x_-$, the square of the norm can be written as $\|x\|_J^2 = [J(x_+ + x_-), x_+ + x_-]_{\mathcal{K}}$. Expanding the terms by sesquilinearity yields

$$\|x\|_J^2 = [x_+, x_+]_{\mathcal{K}} + [x_+, x_-]_{\mathcal{K}} - [x_-, x_+]_{\mathcal{K}} - [x_-, x_-]_{\mathcal{K}}. \tag{2.2.4}$$

But $[x_+, x_-]_{\mathcal{K}} = 0$ and $[x_-, x_+]_{\mathcal{K}} = 0$ by the $[\cdot,\cdot]_{\mathcal{K}}$-orthogonality of $\mathcal{H}_+$ and $\mathcal{H}_-$. Using $\langle\cdot,\cdot\rangle_{\mathcal{H}_+} = [\cdot,\cdot]_{\mathcal{K}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}_-} = -[\cdot,\cdot]_{\mathcal{K}}$, and the fact that Hilbert spaces $\mathcal{H}_\pm$ have norms such that $\langle\cdot,\cdot\rangle_{\mathcal{H}_+} = \|\cdot\|_{\mathcal{H}_+}^2$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}_-} = \|\cdot\|_{\mathcal{H}_-}^2$, respectively, we can write the squared norm succinctly as [OG18, OMCS04]

$$\|x\|_J^2 = \langle x_+, x_+\rangle_{\mathcal{H}_+} + \langle x_-, x_-\rangle_{\mathcal{H}_-} = \|x_+\|_{\mathcal{H}_+}^2 + \|x_-\|_{\mathcal{H}_-}^2, \tag{2.2.5}$$

for $x_\pm, y_\pm \in \mathcal{H}_\pm$. Thus, $\|\cdot\|_J$ defines a norm on $\mathcal{K}$, and the norm topology it induces is called the **strong topology** on $\mathcal{K}$. We now introduce a couple of additional standard concepts in order to tie the notions of fundamental decomposition, fundamental symmetry, and fundamental norm together. Let $(\mathcal{H}_1, \langle\cdot,\cdot\rangle_{\mathcal{H}_1})$ and $(\mathcal{H}_2, \langle\cdot,\cdot\rangle_{\mathcal{H}_2})$ be two Hilbert spaces. A linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ is *bounded* if there exists a constant $C \geq 0$ such that

$$\|Tx\|_{\mathcal{H}_2} \leq C\|x\|_{\mathcal{H}_1}, \quad \text{for all } x \in \mathcal{H}_1. \tag{2.2.6}$$

Equivalently, the *operator norm* $\|T\| := \sup_{\|x\|_{\mathcal{H}_1} \leq 1} \|Tx\|_{\mathcal{H}_2}$ is finite. The *(Hilbert) adjoint* of a bounded operator $A : \mathcal{H}_1 \to \mathcal{H}_2$ is an operator $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ such that

$$\langle Ax, y\rangle_{\mathcal{H}_2} = \langle x, A^*y\rangle_{\mathcal{H}_1} \quad \text{for all } x \in \mathcal{H}_1, y \in \mathcal{H}_2. \tag{2.2.7}$$

In the associated Hilbert space, we have $\langle Jx, y\rangle_J = [JJx, y]_{\mathcal{K}}$ for $x, y \in \mathcal{K}$. But $J^2 = I$, so $\langle Jx, y\rangle_J = [x, y]_{\mathcal{K}}$. By the orthogonality of $\mathcal{H}_+$ and $\mathcal{H}_-$, we can expand this to $[x_+, y_+]_{\mathcal{K}} + [x_-, y_-]_{\mathcal{K}}$, which is the same as $\langle x_+, y_+\rangle_{\mathcal{H}_+} - \langle x_-, y_-\rangle_{\mathcal{H}_-}$. Furthermore, $\langle x, Jy\rangle_J = [Jx, Jy]_{\mathcal{K}}$, which has the same expansion as that of $[x, y]_{\mathcal{K}}$. Hence, $\langle Jx, y\rangle_J = \langle x, Jy\rangle_J$ for all $x, y \in \mathcal{K}$, which means that the fundamental symmetry is *self-adjoint*, i.e., $J = J^*$. By definition of the fundamental symmetry, $J^2 = I$, so it follows that J is *unitary*, i.e., $J^{-1} = J^*$ because $J^2 = JJ^* = I$. Taken together, we have that $J^{-1} = J^* = J$.

By a *linear functional* $\psi : \mathcal{X} \to \mathbb{C}$, we mean a linear transformation that maps from a vector space $\mathcal{X}$ to the complex scalars $\mathbb{C}$, i.e., $\psi(\alpha x + \beta y) = \alpha\psi(x) + \beta\psi(y)$ for any $\alpha, \beta \in \mathbb{C}$ and $x, y \in \mathcal{X}$. If the linear functional $\psi : \mathcal{H} \to \mathbb{C}$ is bounded on a Hilbert space $\mathcal{H}$, then, by the **Riesz-Fréchet theorem of representation** on Hilbert spaces (see, for example, [Fri82, Theorem 6.2.4]), there exists a unique vector $y \in \mathcal{H}$ such that ψ is uniquely represented by an inner product, i.e., $\psi(x) = \langle x, y\rangle_{\mathcal{H}}$ for all $x \in \mathcal{H}$. The details of the Hilbert space version of this result fall outside the scope of this thesis. Rather, we will apply the Hilbert version of this result to prove additional results, such as the analogous representation theorem on a Kreĭn space $\mathcal{K}$, which arises from its associated Hilbert space $\mathcal{H}_{\mathcal{K}}$.

Theorem 2.7 (Riesz Representation on Kreĭn Spaces [Ghe22, Prop. 1.4.4.])**.** *Let $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ be a Kreĭn space and let $\psi : \mathcal{K} \to \mathbb{C}$ be a linear and bounded functional (with respect to the strong topology). Then there exists a unique $y \in \mathcal{K}$ such that $\psi(x) = [x, y]_{\mathcal{K}}$ for all $x \in \mathcal{K}$.*

Proof. Fix a fundamental symmetry $J : \mathcal{K} \to \mathcal{K}$ on the Kreĭn space $\mathcal{K}$ and define the positive definite inner product as $\langle x, y\rangle_J = [Jx, y]_{\mathcal{K}}$. The strong topology of $\mathcal{K}$ is exactly the norm topology on the associated Hilbert space $(\mathcal{H}_{\mathcal{K}}, \langle\cdot,\cdot\rangle_J)$. Therefore, boundedness of ψ on $\mathcal{K}$ means that ψ is bounded on the associated Hilbert space

$\mathcal{H}_\mathcal{K}$ as well. By the Riesz-Fréchet representation theorem in the associated Hilbert space, there exists a unique vector $z \in \mathcal{H}_\mathcal{K}$ such that

$$\psi(x) = \langle x, z \rangle_J = [Jx, z]_\mathcal{K}, \quad \text{for all } x \in \mathcal{K}.$$

Set $y := Jz$. Since the fundamental symmetry J is self-adjoint with respect to $(\mathcal{H}_\mathcal{K}, \langle \cdot, \cdot \rangle_J)$, we have $[Jx, z]_\mathcal{K} = [x, Jz]_\mathcal{K} = [x, y]_\mathcal{K}$ for all $x \in \mathcal{K}$. Hence $\psi(x) = [x, y]_\mathcal{K}$ for all $x \in \mathcal{K}$, which concludes the existence part of the proof. For uniqueness, suppose that $[x, y_1]_\mathcal{K} = [x, y_2]_\mathcal{K}$ for all $x \in \mathcal{K}$. Then also $[x, y_1 - y_2]_\mathcal{K} = 0$ for every x, and because the Kreĭn inner product $[\cdot, \cdot]_\mathcal{K}$ is nondegenerate, this means that $y_1 - y_2 = 0$. Hence y is unique. □

The following result identifies a one-to-one correspondence between fundamental decompositions, fundamental symmetries, and classes of fundamental norms [Ghe22, Remark 1.4.2].

Theorem 2.8 ([Ghe22, Theorem 1.4.1]). *Let $(\mathcal{K}, [\cdot, \cdot]_\mathcal{K})$ be an indefinite inner product space, not necessarily a Kreĭn space. The following are equivalent.*

- *(a)* ***Fundamental decomposition.*** *There exist two $[\cdot, \cdot]_\mathcal{K}$-orthogonal linear manifolds $\mathcal{H}_+$ and $\mathcal{H}_-$ of $\mathcal{K}$ such that $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$, where both $\mathcal{H}_+$ and $\mathcal{H}_-$ are Hilbert spaces with respect to the forms $\langle \cdot, \cdot \rangle_{\mathcal{H}_+} := [\cdot, \cdot]_\mathcal{K}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}_-} := -[\cdot, \cdot]_\mathcal{K}$, respectively.*
- *(b)* ***Fundamental symmetry.*** *There exists a fundamental symmetry $J : \mathcal{K} \to \mathcal{K}$ such that $(\mathcal{K}, \langle \cdot, \cdot \rangle_J)$ is a Hilbert space where $\langle x, y \rangle_J := [Jx, y]_\mathcal{K}$ for all $x, y \in \mathcal{K}$.*
- *(c)* ***Fundamental norm.*** *There exists a positive definite inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{K}$ such that $(\mathcal{K}, \langle \cdot, \cdot \rangle)$ is a Hilbert space and with an associated norm $\| \cdot \|$ that satisfies*

$$\|x\| = \sup_{\substack{y \in \mathcal{K} \\ \|y\| \leq 1}} |[x, y]_\mathcal{K}| \quad \textit{for all } x \in \mathcal{K}.$$

Proof.

$(a) \implies (b)$ By assumption we have $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$. Define a linear operator $J : \mathcal{K} \to \mathcal{K}$ by $J(x_+ + x_-) = x_+ - x_-$ for $x_\pm \in \mathcal{H}_\pm$. Let $\alpha, \beta \in \mathbb{C}$ and $x_\pm, y_\pm \in \mathcal{H}_\pm$ such that $x = x_+ + x_-$ and $y = y_+ + y_-$. The function J satisfies $J^2 = I$, which is evident from re-applying J to $Jx = x_+ - x_-$,

$$J^2 x = J(Jx) = J(x_+ - x_-) = x_+ + x_- = x.$$

Define $\langle x, y\rangle_J := [Jx, y]_\mathcal{K}$ for $x, y \in \mathcal{K}$. That this is sesquilinear and Hermitian follows from $[\cdot,\cdot]_\mathcal{K}$. Specifically, let $x, y, z \in \mathcal{K}$ and $\alpha, \beta \in \mathbb{C}$. Then $\langle\cdot,\cdot\rangle_J$ is linear in the first argument as

$$\begin{aligned}\langle \alpha x + \beta y, z\rangle_J &= [J(\alpha x + \beta y), z]_\mathcal{K}\\ &= [\alpha Jx + \beta Jy, z]_\mathcal{K}\\ &= \alpha[Jx, z]_\mathcal{K} + \beta[Jy, z]_\mathcal{K}\\ &= \alpha\langle x, z\rangle_J + \beta\langle y, z\rangle_J.\end{aligned}$$

It is also conjugate linear in the second argument

$$\begin{aligned}\langle x, \alpha y + \beta z\rangle_J &= [Jx, \alpha y + \beta z]_\mathcal{K}\\ &= \overline{\alpha}[Jx, y]_\mathcal{K} + \overline{\beta}[Jx, z]_\mathcal{K}\\ &= \overline{\alpha}\langle x, y\rangle_J + \overline{\beta}\langle x, z\rangle_J.\end{aligned}$$

Hence $\langle\cdot,\cdot\rangle_J$ is sesquilinear. To establish the Hermitian symmetry of $\langle\cdot,\cdot\rangle_J$ we start with

$$\overline{\langle z, x\rangle_J} = \overline{[Jz, x]_\mathcal{K}} = [x, Jz]_\mathcal{K}.$$

In addition, in our discussion on the self-adjointness of the fundamental symmetry we showed that $[x, Jz]_\mathcal{K} = \langle x_+, z_+\rangle_{\mathcal{H}_+} - \langle x_-, z_-\rangle_{\mathcal{H}_-} = [Jx, z]_\mathcal{K}$. Using the fact that $[Jx, z]_\mathcal{K} = \langle x, z\rangle_J$ by definition of J, it immediately follows that $\overline{\langle z, x\rangle_J} = \langle x, z\rangle_J$. Hence $\langle\cdot,\cdot\rangle_J$ is also Hermitian symmetric and $(\mathcal{K}, \langle\cdot,\cdot\rangle_J)$ is an inner product space according to Definition 2.1. Moreover, let $x = x_+ + x_- \in \mathcal{K}$; the inner product is of the form (2.2.5), i.e. $\langle x, x\rangle_J = \|x_+\|^2_{\mathcal{H}_+} + \|x_-\|^2_{\mathcal{H}_-}$, which is positive unless $x = 0$. Thus, $\langle\cdot,\cdot\rangle_J$ is positive definite, and so $\|\cdot\|_J = \sqrt{\langle\cdot,\cdot\rangle_J}$ defines a valid norm. It remains to show that $\mathcal{K}$ is complete. Let (x_n) be a Cauchy sequence in $\|\cdot\|_J$ and write $x_n = x_{n,+} + x_{n,-}$ with $x_{n,\pm} \in \mathcal{H}_\pm$. By construction of $\|\cdot\|_J$, for all n, m

$$\|x_{n,+} - x_{m,+}\|_{\mathcal{H}_+} \le \|x_n - x_m\|_J \quad \text{and} \quad \|x_{n,-} - x_{m,-}\|_{\mathcal{H}_-} \le \|x_n - x_m\|_J.$$

Therefore, the sequences $(x_{n,\pm})$ are Cauchy in $\mathcal{H}_\pm$. Because $\mathcal{H}_\pm$ are Hilbert spaces, hence complete, the sequences $(x_{n,\pm})$ converge in $\mathcal{H}_\pm$. That is, there exist $x_\pm \in \mathcal{H}_\pm$ with $x_{n,+} \to x_+$ and $x_{n,-} \to x_-$. Set $x = x_+ + x_- \in \mathcal{K}$. Then

$$\|x_n - x\|^2_J = \|x_{n,+} - x_+\|^2_{\mathcal{H}_+} + \|x_{n,-} - x_-\|^2_{\mathcal{H}_-} \longrightarrow 0.$$

Thus, $(\mathcal{K}, \langle\cdot,\cdot\rangle_J)$ is a complete inner product space with a positive definite inner product, hence also a Hilbert space. Specifically, it is the associated Hilbert space to $\mathcal{K}$. In addition, we have shown that $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ satisfies Definition 2.5, and is thus a Kreĭn space.

$(b) \implies (a)$ Let J be the fundamental symmetry and $\mathcal{K}$ be an indefinite inner product space with form $[\cdot,\cdot]_{\mathcal{K}}$. By definition, J is self-adjoint and satisfies $J^2 = I$. Define $J_\pm = \frac{1}{2}(I \pm J)$. Compute

$$J_\pm^2 = \left(\frac{I \pm J}{2}\right)^2 = \frac{1}{4}(I \pm 2J + J^2) = \frac{1}{4}(2I \pm 2J) = \frac{1}{2}(I \pm J) = J_\pm,$$

hence $J_\pm$ are *projections*. Since J is selfadjoint with respect to $\langle\cdot,\cdot\rangle_J$, we have

$$(J_\pm)^* = \left(\frac{1}{2}(I \pm J)\right)^* = \frac{1}{2}(I^* \pm J^*) = \frac{1}{2}(I \pm J) = J_\pm.$$

In particular, $J_\pm^* = J_\pm$, so $J_\pm$ are *orthogonal projections* on the associated Hilbert space $(\mathcal{K}, \langle\cdot,\cdot\rangle_J)$. Define the linear manifolds

$$\mathcal{H}_\pm := J_\pm\mathcal{K} = \{J_\pm x : x \in \mathcal{K}\}.$$

We also note the identities

$$J_+ - J_- = \frac{I+J}{2} - \frac{I-J}{2} = J,$$

as well as $J_+ + J_- = I$, and $J_+J_- = \frac{1}{4}(I+J)(I-J) = \frac{1}{4}(I - J^2) = 0$ (and also $J_-J_+ = 0$). Thus, for every $x \in \mathcal{K}$ we may write

$$x = (J_+ + J_-)x = J_+x + J_-x,$$

with $J_\pm x \in \mathcal{H}_\pm$. Thus, $\mathcal{K} = \mathcal{H}_+ + \mathcal{H}_- = \{x_+ + x_- : x_\pm \in \mathcal{H}_\pm\}$. Because $J_\pm$ are projections and $\mathcal{H}_\pm = J_\pm\mathcal{K}$, for any $x \in \mathcal{H}_\pm$ there exists $y \in \mathcal{K}$ such that $x = J_\pm y$. Then

$$J_\pm x = J_\pm(J_\pm(y)) = J_\pm^2 y = J_\pm y = x.$$

Thus $J_\pm x = x$ for all $x \in \mathcal{H}_\pm$. Suppose $x \in \mathcal{H}_+ \cap \mathcal{H}_-$ then $x = J_+x$ *and* $x = J_-x$, hence

$$x = J_+x = J_+(J_-x) = (J_+J_-)x = 0.$$

Thus $\mathcal{H}_+ \cap \mathcal{H}_- = \{0\}$ and $\mathcal{K} = \mathcal{H}_+ \oplus \mathcal{H}_-$ in our Kreĭn space notation. Now, we need to show that $\mathcal{H}_+$ and $\mathcal{H}_-$ are also $[\cdot,\cdot]_{\mathcal{K}}$-orthogonal to assert that $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$. Consider any $u \in \mathcal{H}_+$ and $w \in \mathcal{H}_-$. Then $u = J_+x$ and $w = J_-y$ for some $x, y \in \mathcal{K}$. Using that $J_\pm$ are self-adjoint and that $Ju = J_+x = u$, we compute

$$\begin{aligned}
[u,w]_{\mathcal{K}} &= \langle Ju, w\rangle_J \\
&= \langle u, w\rangle_J \\
&= \langle J_+x, J_-y\rangle_J \\
&= \langle x, (J_+)^* J_-y\rangle_J \\
&= \langle x, J_+J_-y\rangle_J \\
&= \langle x, 0\rangle_J \\
&= 0.
\end{aligned}$$

Since u, w were arbitrary, we have that $\mathcal{H}_+ \perp \mathcal{H}_-$ with respect to $[\cdot,\cdot]_{\mathcal{K}}$. As a result, we have the fundamental decomposition $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$. It remains to show that $\mathcal{H}_\pm$ are Hilbert spaces (i.e. complete with positive definite inner products $\langle\cdot,\cdot\rangle_{\mathcal{H}_+} := [\cdot,\cdot]_{\mathcal{K}} = \langle J\,\cdot\,,\cdot\rangle_J$ and $\langle\cdot,\cdot\rangle_{\mathcal{H}_-} := -[\cdot,\cdot]_{\mathcal{K}} = -\langle J\,\cdot\,,\cdot\rangle_J$). Let (u_n) be a Cauchy sequence in $\mathcal{H}_+$ with respect to the norm $\|\cdot\|_J = \sqrt{\langle\cdot,\cdot\rangle_J}$. Each term of the sequence is of the form $u_n = J_+x_n$ for some $x_n \in \mathcal{K}$. Since $u_n \in \mathcal{H}_+$ we have $J_+u_n = u_n$ for all n. As $(\mathcal{K}, \langle\cdot,\cdot\rangle_J)$ is a Hilbert space by assumption, every Cauchy sequence converges in $\mathcal{K}$ with respect to $\|\cdot\|_J$. In particular, there exists a $u \in \mathcal{K}$ such that $u_n \to u$. To show $u \in \mathcal{H}_+$ we first show that J_+ is bounded. Write $x = J_+x + J_-x$ for an arbitrary element $x \in \mathcal{K}$. Then by (2.2.5) we have

$$\|J_+x\|_J^2 \leq \|J_+x\|_J^2 + \|J_-x\|_J^2 = \|x\|_J^2.$$

Thus, J_+ is bounded according to (2.2.6) with $C = 1$ (the analogous case for J_- holds as well). Applying the boundedness inequality for J_+ and using that $J_+u_n = u_n$ for all n, we get

$$\|u_n - J_+u\|_J = \|J_+u_n - J_+u\|_J \leq \|J_+\|\|u_n - u\|_J \longrightarrow 0,$$

as $\|J_+\| \leq 1$ and $u_n \to u$. Using the standard result that Cauchy limits $u \in \mathcal{K}$

are unique, we have that $u = J_+ u \in \mathcal{H}_+$. Hence every Cauchy sequence in $\mathcal{H}_+$ converges in $\mathcal{H}_+$ (analogously for $\mathcal{H}_-$). Thus, $\mathcal{H}_\pm$ are complete. Finally, for $u \in \mathcal{H}_+$ and $w \in \mathcal{H}_-$ we have:

$$\langle u, u\rangle_{\mathcal{H}_+} = [u, u]_{\mathcal{K}} = \langle Ju, u\rangle_J = \langle u, u\rangle_J > 0, \quad \text{if } u \neq 0$$

and

$$\langle w, w\rangle_{\mathcal{H}_-} = -[w, w]_{\mathcal{K}} = -\langle Jw, w\rangle_J = \langle w, w\rangle_J > 0, \quad \text{if } w \neq 0$$

so the forms are positive definite, and hence we've shown $(\mathcal{H}_+, \langle\cdot,\cdot\rangle_{\mathcal{H}_+})$ and $(\mathcal{H}_-, \langle\cdot,\cdot\rangle_{\mathcal{H}_-})$ are Hilbert spaces and the desired fundamental decomposition $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$ holds.

$(b) \implies (c)$ By assumption, there exists a fundamental symmetry $J : \mathcal{K} \to \mathcal{K}$ such that $(\mathcal{K}, \langle\cdot,\cdot\rangle_J)$ is a Hilbert space. The inner product of a Hilbert space is positive definite by definition, so it just remains to show that the norm satisfies the desired property. Note that $\|\cdot\|_J = \sqrt{\langle\cdot,\cdot\rangle_J}$ defines the norm on $\mathcal{K}$. We already showed that J is unitary and self-adjoint, so $J^{-1} = J^* = J$, and

$$\|Jx\|_J^2 = \langle Jx, Jx\rangle_J = \langle x, x\rangle_J = \|x\|_J^2 \quad \text{for } x \in \mathcal{K}.$$

That is, $\|Jx\|_J = \|x\|_J$. Using the Cauchy-Schwarz inequality we have that $|\langle Jx, y\rangle_J| \leq \|x\|_J \|y\|_J$ for $x, y \in \mathcal{K}$. Taking the supremum over the unit ball $\|y\|_J \leq 1$ yields

$$\sup_{\substack{y \in \mathcal{K} \\ \|y\|_J \leq 1}} |\langle Jx, y\rangle_J| \leq \sup_{\substack{y \in \mathcal{K} \\ \|y\|_J \leq 1}} \|x\|_J \|y\|_J = \|x\|_J.$$

If $x = 0$, then equality holds trivially. If $x \neq 0$ then define $y_0 = Jx/\|Jx\|_J$. This element is normalized, i.e. $\|y_0\|_J = 1$. Because y_0 belongs to the unit ball,

$$\sup_{\substack{y \in \mathcal{K} \\ \|y\|_J \leq 1}} |\langle Jx, y\rangle_J| \geq |\langle Jx, y_0\rangle_J| = \frac{\langle Jx, Jx\rangle_J}{\|Jx\|_J} = \|Jx\|_J = \|x\|_J \qquad \text{for all } x \in \mathcal{K}.$$

The bidirectional inequality gives the equality $\sup_{\|y\|_J \leq 1} |\langle Jx, y\rangle_J| = \|x\|_J$.

Finally, because $\langle Jx, y\rangle_J = [x,y]_{\mathcal{K}}$, we have established that

$$\|x\| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |[x,y]_{\mathcal{K}}| \qquad \text{for all } x \in \mathcal{K},$$

which is what we had to show.

$(c) \implies (b)$ Let $\langle\cdot,\cdot\rangle$ be the given positive definite inner product on $\mathcal{K}$ such that $(\mathcal{K}, \langle\cdot,\cdot\rangle)$ is a Hilbert space with norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$ and

$$\|x\| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |[x,y]_{\mathcal{K}}| \qquad \text{for all } x \in \mathcal{K}.$$

Define the linear functional $\varphi_x : \mathcal{K} \to \mathbb{C}$ for which $\varphi_x(y) := [x,y]_{\mathcal{K}}$. Taking the norm gives

$$\|\varphi_x\| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |\varphi_x(y)| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |[x,y]_{\mathcal{K}}| = \|x\|,$$

which is finite. Hence, φ_x is a bounded linear functional on $\mathcal{K}$, and so by the Riesz-Fréchet representation theorem on the Hilbert space $(\mathcal{K}, \langle\cdot,\cdot\rangle)$ there exists a unique representer $v_x \in \mathcal{K}$ for which

$$\varphi_x(y) = [x,y]_{\mathcal{K}} = \langle v_x, y\rangle.$$

Define the operator $J : \mathcal{K} \to \mathcal{K}$ with $Jx = v_x$. Then

$$\varphi_x(y) = [x,y]_{\mathcal{K}} = \langle Jx, y\rangle.$$

Let $x, y, z \in \mathcal{K}$ and $\alpha, \beta \in \mathbb{C}$. We show that the operator is linear by

$$\begin{aligned}[\alpha x + \beta y, z]_{\mathcal{K}} &= \alpha[x,z]_{\mathcal{K}} + \beta[y,z]_{\mathcal{K}} \\ &= \alpha\langle Jx, z\rangle + \beta\langle Jy, z\rangle \\ &= \langle \alpha Jx + \beta Jy, z\rangle,\end{aligned}$$

for all $z \in \mathcal{K}$. Since the Riesz representer is unique, the vectors $J(\alpha x + \beta y)$ and $\alpha Jx + \beta Jy$ have the same inner products with all z, hence they coincide.

Hence J is a linear operator. It remains to show that $J^2 = I$. First calculate,

$$\|Jx\| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |\langle Jx, y\rangle| = \sup_{\substack{y\in\mathcal{K}\\ \|y\|\le 1}} |[x,y]_{\mathcal{K}}| = \|x\|.$$

So J is bounded by (2.2.6) using $C = 1$ (in fact, J is the Gram operator introduced later in Example 2.11). Furthermore, by the relationship of $\langle\cdot,\cdot\rangle$ and $[\cdot,\cdot]_{\mathcal{K}}$ under J, and by the Hermitian symmetry of the inner products we have that

$$\langle x, Jy\rangle = \overline{\langle Jy, x\rangle} = \overline{[y,x]_{\mathcal{K}}} = [x,y]_{\mathcal{K}} = \langle Jx, y\rangle.$$

Therefore, $J = J^*$, so J is a self-adjoint operator. The linearity and norm preservation of J, combined with the polarization identity (2.1.4), give

$$\langle Jx, Jy\rangle = \frac{1}{4}\sum_{k=0}^{3} i^k\|J(x+i^k y)\|^2 = \frac{1}{4}\sum_{k=0}^{3} i^k\|x+i^k y\|^2 = \langle x, y\rangle.$$

As a result, $\langle x, y\rangle = \langle Jx, Jy\rangle = \langle J^2x, y\rangle$ for all $x, y \in \mathcal{K}$. Nondegeneracy of the inner product establishes $J^2 = I$ as desired. Note that this shows that J, as constructed, is invertible. Recall also that $[x,y]_{\mathcal{K}} = \langle Jx, y\rangle$ by construction. In particular,

$$[Jx, y]_{\mathcal{K}} = \langle J^2x, y\rangle = \langle x, y\rangle.$$

Consequently,

$$\langle x, y\rangle = \langle x, y\rangle_J := [Jx, y]_{\mathcal{K}},$$

and J is a fundamental symmetry.

□

The theorem states that any indefinite inner product space $\mathcal{K}$ that satisfies any of the conditions (a), (b), or (c) will satisfy all of them. In addition, each condition is a necessary and sufficient condition for $\mathcal{K}$ to be a Kreĭn space. It also follows from Theorem 2.8 that the choice of fundamental decomposition is the same as selecting a fundamental symmetry J. Though the fundamental decomposition (fundamental symmetry) is not necessarily unique, the next result shows that the strong topology introduced by any fundamental norm is independent of this decision [OG18].

Corollary 2.9 ([Ghe22, Corollary 1.4.3])**.** *Let J_1, J_2 be any two fundamental symmetries on a Kreĭn space $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$. The associated fundamental norms $\|x\|_{J_1}$ and $\|x\|_{J_2}$ are equivalent.*

Corollary 2.9 lets us unambiguously introduce a **strong topology** on the Kreĭn space $\mathcal{K} = \mathcal{H}_- \,[\oplus]\, \mathcal{H}_+$ induced by *any* fundamental norm. The strong topology of $\mathcal{K}$ is exactly the norm topology on the associated Hilbert space $(\mathcal{H}_\mathcal{K}, \langle\cdot,\cdot\rangle_J)$, making the strong topology canonical and independent of the choice of fundamental decomposition or symmetry J. Fixing a fundamental symmetry J defines the strong topology via $\langle x, y\rangle_J = [Jx, y]_\mathcal{K}$ and, as the Corollary shows, any two choices J_1, J_2 produce equivalent norms, hence the same strong topology. Therefore, any topological concerns in the Kreĭn space $\mathcal{K}$ (e.g. continuity, limits, openness, boundedness, compactness etc.) can be studied using the norm topology of the associated Hilbert space.

Remark 2.10 ([Ghe22, Theorem 1.4.11, and page 1]). In general in Kreĭn spaces, the strong topology cannot be recovered from the inner product alone, but must be introduced via a specific choice of fundamental symmetry J (fundamental decomposition). As the Corollary 2.9 shows, different choices yield equivalent norms. In contrast, in Pontryagin spaces Π_κ, the strong topology can be characterized directly in terms of the inner product (and is independent of the fundamental decomposition). However, the details of this are considered beyond the scope of this thesis.

Example 2.11 (Constructing a Kreĭn space from a Hilbert space [Ghe22, Example 1.1.10, Remark 1.3.3, and Prop. 2.1.9]). Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_\mathcal{H})$ be a Hilbert space and let $A : \mathcal{H} \to \mathcal{H}$ be a linear, self-adjoint, and bounded operator with respect to $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle_\mathcal{H}}$. Set $\mathcal{K}_A = \mathcal{H}$ and define the Hermitian form $[x, y]_A$ by

$$[x, y]_A := \langle Ax, y\rangle_\mathcal{H}, \quad \text{for all } x, y \in \mathcal{K}_A.$$

Thus $(\mathcal{K}_A, [\cdot,\cdot]_A)$ is a (possibly degenerate) indefinite inner product space. The operator A is called the *Gram operator* of the indefinite inner product $[\cdot,\cdot]_A$ with respect to the positive inner product $\langle\cdot,\cdot\rangle_\mathcal{H}$. It is possible to show, using the spectral theorem for bounded self-adjoint operators (the details of this result fall outside the scope of this survey), that $\mathcal{K}_A$ admits an orthogonal decomposition

$$\mathcal{K}_A = \mathcal{K}_+ \,[\oplus]\, \mathcal{K}_A^0 \,[\oplus]\, \mathcal{K}_-.$$

Here, $(\mathcal{K}_+, [\cdot,\cdot]_A)$ is strictly positive, and $(\mathcal{K}_-, [\cdot,\cdot]_A)$ is strictly negative, and $\mathcal{K}_A^0$ is the *null space/kernel* $\ker(A)$ of A. That is, $\mathcal{K}_A^0 = \ker(A) := \{x \in \mathcal{K}_A : Ax = 0\}$. All three subspaces are mutually orthogonal with respect to $[\cdot,\cdot]_A$. In particular, if the indices $\kappa_\pm(\mathcal{K}_A)$ are finite, then they equal the number of positive/negative eigenvalues of A, counted with multiplicities. Assume from now on that A also has a bounded inverse A^{-1}. Crucially, $\ker(A) = \{0\}$ by a standard linear algebra result. Hence $[\cdot,\cdot]_A$ is nondegenerate and $\kappa_0(\mathcal{K}_A) = 0$, so the orthogonal decomposition reduces to

$$\mathcal{K}_A = \mathcal{K}_+ \, [\oplus] \, \mathcal{K}_-,$$

where $\mathcal{K}_+$ is strictly positive and $\mathcal{K}_-$ is strictly negative for $[\cdot,\cdot]_A$. In general, these subspaces need not be closed or complete (hence not Hilbert spaces), so this is not automatically a fundamental decomposition. One may, however, show that this is in fact a fundamental decomposition in the case where A is boundedly invertible, but this goes beyond the scope of the thesis [Ghe22, see Prop. 2.1.9]. In that case, $\mathcal{K}_+$ and $\mathcal{K}_-$ are Hilbert spaces and the above decomposition is indeed fundamental. We say a self-adjoint operator A on a Hilbert space is *positive* if $\langle Ax, x\rangle_{\mathcal{H}} \geq 0$ and write $A \geq 0$. For any bounded self-adjoint operator, we say it has a *Jordan decomposition* if $A = A_+ - A_-$ for $A_\pm \geq 0$ in which $A_+A_- = A_-A_+ = 0$, and $A_\pm$ commute. Furthermore, we use $A^{1/2}$ to denote the (unique) positive *square-root* operator for which $(A^{1/2})^2 = A$ if $A \geq 0$. We need two auxiliary identities. First, $A^2 = (A_+ - A_-)^2 = (A_+ + A_-)^2$ because the cross-terms vanish, and $A_+ + A_- \geq 0$, so $|A| := (A^*A)^{1/2} = (A^2)^{1/2} = A_+ + A_-$. Second, $(A_+^{1/2} + A_-^{1/2})^2 = A_+ + A_- = |A| \geq 0$, so $|A|^{1/2} = A_+^{1/2} + A_-^{1/2}$. Using the Jordan decomposition of A we have on $\mathcal{K}_\pm$

$$[x, y]_A := \pm \langle A_\pm x, y\rangle_{\mathcal{H}}$$

for $x, y \in \mathcal{K}_\pm$. Moreover,

$$[x, x]_A = \pm \|A_\pm^{1/2} x\|^2 \quad \text{for all } x \in \mathcal{K}_\pm.$$

It is possible to show that the normed spaces $(\mathcal{K}_\pm, \|A_\pm^{1/2} \cdot \|)$ are complete, and that the normed space $(\mathcal{K}_A, \||A|^{1/2} \cdot \|)$ is complete as well. The details of this require more advanced topics, which go beyond this thesis. Interested readers are referred to [Ghe22, Prop. 2.1.9]. Define $J : \mathcal{K}_A \to \mathcal{K}_A$ by

$$Jx = \begin{cases} x & \text{if } x \in \mathcal{K}_+ \\ 0 & \text{if } x \in \mathcal{K}_A^0 \\ -x & \text{if } x \in \mathcal{K}_-. \end{cases}$$

Then the *polar decomposition* $A = J|A|$ holds. Recall that $\mathcal{K}_A^0 = \{0\}$ because A is invertible. Thus, if $x \in \mathcal{K}_A^0$ then $x = 0$, so $J^2x = 0 = x$. In addition, $J^2x = J(x) = x$ if $x \in \mathcal{K}_+$, and $J^2x = J(-x) = x$ if $x \in \mathcal{K}_-$. Hence, $J^2 = I$ on all of $\mathcal{K}_A$. Thus

$$[x, y]_A = \langle Ax, y\rangle_{\mathcal{H}} = \langle J|A|x, y\rangle_{\mathcal{H}}, \quad x, y \in \mathcal{K}_A.$$

With $\ker(A) = \{0\}$, the inner product $\langle |A|\cdot, \cdot\rangle_{\mathcal{H}}$ is positive definite and gives the norm $\sqrt{\langle |A|\cdot, \cdot\rangle_{\mathcal{H}}} = \||A|^{1/2}\cdot\|$. Moreover, by [Ghe22, Prop. 2.1.9], $\mathcal{K}_A$ is complete with respect to this norm, though we reiterate that this particular step goes beyond the scope of the thesis. Therefore, $(\mathcal{K}_A, \langle |A|\cdot, \cdot\rangle_{\mathcal{H}})$ is a Hilbert space and J is a fundamental symmetry on $\mathcal{K}_A$. Hence, by Theorem 2.8(b), $(\mathcal{K}_A, [\cdot, \cdot]_A)$ is a Kreĭn space. As the example shows, we may construct a Kreĭn space from a Hilbert space if A^{-1} is bounded. If A^{-1} is not bounded, we still obtain an indefinite inner product space with an orthogonal decomposition, but not a Kreĭn space in general: a nontrivial neutral part $\mathcal{K}_A^0$ may remain and/or the strict positive/negative subspaces $\mathcal{K}_\pm$ fail to be closed or complete (hence not Hilbert spaces) [Ghe22, Prop. 2.1.9]. $\square$

2.3 Reproducing Kernel Kreĭn Space

In this section we pass from Kreĭn spaces to spaces of functions with reproducing kernels. We define the associated reproducing property and present the main existence and uniqueness results that are relevant for this survey. Let X be a nonempty set and let $\mathcal{Y}$ be a (possibly indefinite) inner product space. We use $\mathcal{F}(X; \mathcal{Y})$ to denote the vector space of functions from X into $\mathcal{Y}$. At this point X is just a nonempty set, but X may itself carry an indefinite inner product such that Definition 2.1 holds. If this is the case, we swap in the notation $\mathcal{X}$ to indicate that it is an (indefinite) inner product space. If the domain and codomain coincide, we abbreviate $\mathcal{F}(\mathcal{X}) := \mathcal{F}(\mathcal{X}; \mathcal{X})$ [Ghe22].

Given a function $f : X \to \mathcal{Y}$, where $\mathcal{Y}$ is an inner product space, we say that f has *finite support* if the set $\operatorname{supp}(f) = \{x \in X : f(x) \neq 0\}$ is of finite cardinality. We denote by $\mathcal{F}_0(X; \mathcal{Y}) \subseteq \mathcal{F}(X; \mathcal{Y})$ the vector space of *finitely-supported functions* $f : X \to \mathcal{Y}$. Note that the inclusion turns into an equality if X is finite. The finite support of $f, g \in \mathcal{F}_0(X; \mathcal{Y})$ ensures that $f \pm g \in \mathcal{F}_0(X; \mathcal{Y})$, i.e., the sum remains finitely supported. In addition, for every $f \in \mathcal{F}_0(X; \mathcal{Y})$ there exists a unique representation for which $f = \sum_{x \in \operatorname{supp}(f)} \delta_x f_x$ holds, where $\delta_x : X \to \mathbb{C}$ is the *Kronecker delta*, i.e., $\delta_x(t) = 1$ if $t = x$ and 0 elsewhere. Similarly, $f_x := f(x) \in \mathcal{Y}$

is defined as the image of x under f. We will restrict ourselves to *scalar-valued* functions $f : X \to \mathbb{C}$ for the rest of this section, i.e., we fix the codomain to $\mathcal{Y} = \mathbb{C}$. Consequently, $f_x \in \mathbb{C}$ reduces to a scalar [Ghe22, Ghe13, Chapter 6.2 & Chapter 3, respectively.].

Definition 2.12 (Scalar-valued Kernel [Ghe22, Chapter 6.2])**.** Given a nonempty set X, a function $k : X \times X \to \mathbb{C}$ is called a scalar-valued *kernel*. The *adjoint kernel* is defined by $k^*(x, y) = k(y, x)^*$ for all $x, y \in X$. A kernel is called a *Hermitian kernel* if $k^* = k$.

Since the kernels we will consider are scalar-valued with values in $\mathbb{C}$, their adjoint kernels are given by the complex conjugates, i.e., $k^*(x, y) = \overline{k(y, x)}$ for all $x, y \in X$. Thus, a kernel is Hermitian if and only if $k(x, y) = \overline{k(y, x)}$. Given a Hermitian kernel $k : X \times X \to \mathbb{C}$ we define the *kernel-induced form* $[\cdot, \cdot]_k$ on $\mathcal{F}_0(X; \mathbb{C})$ by the double sum

$$[f, g]_k := \sum_{i,j} f_{x_i} \overline{g_{y_j}} k(y_j, x_i), \quad \text{for all } f, g \in \mathcal{F}_0(X; \mathbb{C}), \tag{2.3.1}$$

where $f = \sum_i f_{x_i} \delta_{x_i}$ and $g = \sum_j g_{y_j} \delta_{y_j}$ and all $f_{x_i}, g_{y_j} \in \mathbb{C}$ are scalars ([Ghe22, Theorem 6.2.1]). That this form is well-defined follows from the finite support of f and g. Sesquilinearity and Hermitian symmetry of $[\cdot, \cdot]_k$ follow from k being Hermitian. It follows that $[\cdot, \cdot]_k$ is an inner product such that $(\mathcal{F}_0(X; \mathbb{C}), [\cdot, \cdot]_k)$ is a (possibly indefinite) inner product space in the sense of Definition 2.1. Hence, we may define the *ranks of positivity/negativity* $\kappa_\pm(k)$ as the indices of the inner product space from Definition 2.2. With this definition, a kernel k is *positive (negative) semidefinite* if and only if $[f, f]_k \geq 0$ ($[f, f]_k \leq 0$) for all $f \in \mathcal{F}_0(X; \mathbb{C})$ [Ghe22].

Holding the second argument of a Hermitian kernel k constant, for some choice of $x \in X$, we define the *kernel section* $k(\cdot, x) : X \to \mathbb{C}$ by the function $k(\cdot, x) : y \mapsto k(y, x)$, and we use the shorthand $k_x(\,\cdot\,)$ for notational convenience. In addition, for any element $x \in X$ and for $f \in \mathcal{F}(X; \mathbb{C})$, we define the *evaluation functional* $\mathrm{Ev}_x : \mathcal{F}(X; \mathbb{C}) \to \mathbb{C}$ as the value of the function f at x. That is, $\mathrm{Ev}_x(f) = f(x)$.

Let X be a nonempty set, k a Hermitian kernel, and $(\mathcal{F}_0(X; \mathbb{C}), [\cdot, \cdot]_k)$ the inner product space with kernel-induced inner product via (2.3.1). Of particular interest are the linear subspaces $\mathcal{K}$ of $\mathcal{F}(X; \mathbb{C})$ for which $\mathcal{K}$ is a Kreĭn space and the evaluation functionals $\mathrm{Ev}_x : \mathcal{K} \to \mathbb{C}$ are continuous on the strong topology of $\mathcal{K}$.

Definition 2.13 (Reproducing Kernel Kreĭn Space (RKKS) [OMCS04, OG18, Definitions 4 and 2, respectively.])**.** Let X be a nonempty set. A Kreĭn space $(\mathcal{K}, [\cdot, \cdot]_\mathcal{K})$

is a *reproducing kernel Kreĭn Space (RKKS)* if $\mathcal{K} \subseteq \mathcal{F}(X;\mathbb{C})$ and, for every $x \in X$, the evaluation functional $\mathrm{Ev}_x : \mathcal{K} \to \mathbb{C}$ is continuous with respect to the strong topology of $\mathcal{K}$.

That is, the difference between a general Kreĭn space of functions and an RKKS is that all evaluation functionals Ev_x are continuous in an RKKS. Note that continuity on the strong topology of $\mathcal{K}$ implies that any such evaluation functional is also bounded. When the negative index of an RKKS is finite, i.e. $\kappa_- < \infty$, then the RKKS is called a *reproducing kernel Pontryagin space (RKPS)*. Furthermore, if $\kappa_- = 0$ (when the Kreĭn inner product is positive definite), the RKKS is called a *reproducing kernel Hilbert space (RKHS)*. As the name suggests, an RKKS satisfies a particular reproducing property, which we define next [Ghe13].

Definition 2.14 (Reproducing kernel/Reproducing property [OMCS04, Prop. 6])**.** Let X be a nonempty set and $\mathcal{K}$ be a Kreĭn space such that $\mathcal{K} \subseteq \mathcal{F}(X;\mathbb{C})$. A Hermitian kernel $k : X \times X \to \mathbb{C}$ is a *reproducing kernel* for $\mathcal{K}$ if, for each $x \in X$ and $f \in \mathcal{K}$, the kernel section $k_x(\,\cdot\,)$ belongs to $\mathcal{K}$ and satisfies the *reproducing property*

$$f(x) = [\,f,\, k_x(\,\cdot\,)]_{\mathcal{K}}.$$

In particular, if k is the reproducing kernel of an RKKS $\mathcal{K}$, then $k(x,y) = [k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}}$.

Theorem 2.15 (Existence and Uniqueness of the Reproducing Kernel [OMCS04, Prop. 6])**.** *If $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ is a reproducing kernel Kreĭn space, then there exists a unique reproducing kernel k that satisfies the reproducing property 2.14.*

Proof. Consider any $x \in X$. The evaluation map $\mathrm{Ev}_x : \mathcal{K} \to \mathbb{C}$ is a bounded linear functional, hence by the Riesz representation theorem on Kreĭn spaces (Theorem 2.7), there exists a unique element $k_x(\,\cdot\,) \in \mathcal{K}$ with

$$f(x) = \mathrm{Ev}_x(f) = [f, k_x(\,\cdot\,)]_{\mathcal{K}}\,, \quad \text{for all } f \in \mathcal{K}.$$

Define $k(x,y) := [k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}}$. For any $y \in X$, applying the representation formula at the point y and $f = k_x(\,\cdot\,)$ gives $k_x(y) = [k_x(\,\cdot\,), k_y(\,\cdot\,)]_{\mathcal{K}}$. By definition of k, this is exactly $k(y,x)$. Thus, $k_x(y) = k(y,x)$ for every $y \in X$, and therefore $k_x(\,\cdot\,) = k(\cdot, x)$. Hence,

$$f(x) = [f, k_x(\,\cdot\,)]_{\mathcal{K}} = [f, k(\cdot, x)]_{\mathcal{K}},$$

so the reproducing property holds and the existence part is complete. The kernel is also Hermitian because the Kreĭn inner product $[\cdot,\cdot]_{\mathcal{K}}$ is Hermitian, specifically

$$k(y,x) = [k_x(\,\cdot\,), k_y(\,\cdot\,)]_{\mathcal{K}} = \overline{[k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}}} = \overline{k(x,y)}.$$

To show uniqueness, suppose that two kernels k and m both satisfy the reproducing property 2.14. For each $x \in X$, it holds that $[f, k_x(\,\cdot\,)]_{\mathcal{K}} = [f, m_x(\,\cdot\,)]_{\mathcal{K}}$ for all $f \in \mathcal{K}$. Therefore,

$$[f, k_x(\,\cdot\,) - m_x(\,\cdot\,)]_{\mathcal{K}} = 0, \qquad \text{for all } f \in \mathcal{K}.$$

By nondegeneracy of the Kreĭn form $[\cdot,\cdot]_{\mathcal{K}}$ this implies $k_x(\,\cdot\,) = m_x(\,\cdot\,)$ for every $x \in X$. Consequently,

$$k(x,y) = [k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}} = [m_y(\,\cdot\,), m_x(\,\cdot\,)]_{\mathcal{K}} = m(x,y),$$

for all $x, y \in X$. Hence $k = m$, which completes the proof. □

This theorem shows that every RKKS has a unique reproducing kernel k. Moreover, it is possible to show that the inner product of the RKKS is also uniquely determined by the reproducing kernel (in fact, induced by (2.3.1)). Before we show this result, we state the following intermediate result.

Lemma 2.16 (Density of kernel sections [Ghe13, Chapter 3; *RK4*])**.** *Let X be a nonempty set and let $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ be an RKKS with reproducing kernel k, such that $\mathcal{K} \subseteq \mathcal{F}(X;\mathbb{C})$. The linear span $D := \operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$ is dense in $\mathcal{K}$ with respect to the strong topology of $\mathcal{K}$.*

The proof of this result is beyond the scope of this thesis. Instead, we will use the lemma to show that the kernel-induced form (2.3.1) coincides with the inner product of the RKKS when restricted to $\mathcal{K}$.

Corollary 2.17 (Characterization result; Uniqueness of Kernel-induced Form)**.** *Let $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ be an RKKS of functions defined on a nonempty set X with reproducing kernel k. Then $[\cdot,\cdot]_{\mathcal{K}}$ is uniquely determined by k. Specifically, let $[\cdot,\cdot]'_{\mathcal{K}}$ be another Kreĭn form such that k is also the reproducing kernel of the RKKS $(\mathcal{K}, [\cdot,\cdot]'_{\mathcal{K}})$. If the strong topologies induced by $[\cdot,\cdot]_{\mathcal{K}}$ and $[\cdot,\cdot]'_{\mathcal{K}}$ are equivalent, then*

$$[f,g]'_{\mathcal{K}} = [f,g]_{\mathcal{K}}, \quad \textit{for all } f, g \in \mathcal{K}.$$

Proof. Let $D = \text{span}\{k_x(\cdot) : x \in X\}$, which is dense in $\mathcal{K}$ by Lemma 2.16. We begin by studying arbitrary finite sums $f, g \in D$. These functions are of the form $f = \sum_i \alpha_i k_{x_i}(\,\cdot\,)$ and $g = \sum_j \beta_j k_{y_j}(\,\cdot\,)$ for some $\alpha_i, \beta_j \in \mathbb{C}$. Since k is a reproducing kernel for both $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ and $(\mathcal{K}, [\cdot,\cdot]'_{\mathcal{K}})$, the reproducing property 2.14 gives

$$[k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}} = k(x,y) = [k_y(\,\cdot\,), k_x(\,\cdot\,)]'_{\mathcal{K}}, \qquad \text{for all } x, y \in X.$$

Since f and g are finite sums, sesquilinearity of the forms $[\cdot,\cdot]_{\mathcal{K}}$ and $[\cdot,\cdot]'_{\mathcal{K}}$ implies that

$$[f,g]_{\mathcal{K}} = \left[\sum_i \alpha_i k_{x_i}(\,\cdot\,), \sum_j \beta_j k_{y_j}(\,\cdot\,)\right]_{\mathcal{K}} = \sum_{i,j} \alpha_i \overline{\beta}_j [k_{x_i}(\,\cdot\,), k_{y_j}(\,\cdot\,)]_{\mathcal{K}} = \sum_{i,j} \alpha_i \overline{\beta}_j k(y_j, x_i),$$

and that

$$[f,g]'_{\mathcal{K}} = \left[\sum_i \alpha_i k_{x_i}(\,\cdot\,), \sum_j \beta_j k_{y_j}(\,\cdot\,)\right]'_{\mathcal{K}} = \sum_{i,j} \alpha_i \overline{\beta}_j [k_{x_i}(\,\cdot\,), k_{y_j}(\,\cdot\,)]'_{\mathcal{K}} = \sum_{i,j} \alpha_i \overline{\beta}_j k(y_j, x_i).$$

So the two forms coincide on D. Furthermore, D is dense in $\mathcal{K}$ with respect to the *common* strong topology of $[\cdot,\cdot]_{\mathcal{K}}$ and $[\cdot,\cdot]'_{\mathcal{K}}$, which coincides by assumption. Hence, for any $f, g \in \mathcal{K}$ there exist sequences $f_n, g_n \in D$ with $f_n \to f$ and $g_n \to g$ in this topology. By Theorem 2.8, because $\mathcal{K}$ is an RKKS, it also has a fundamental symmetry J for which $[f,g]_{\mathcal{K}} = \langle Jf, g\rangle_J$. By the Cauchy-Schwarz inequality: $|[f,g]_{\mathcal{K}}| \le \|f\|_J \|g\|_J$; thus $[\cdot,\cdot]_{\mathcal{K}}$ is jointly continuous in the strong topology on $\mathcal{K}$. The same argument holds for $[\cdot,\cdot]'_{\mathcal{K}}$. Therefore,

$$[f,g]_{\mathcal{K}} = \lim_{n\to\infty} [f_n, g_n]_{\mathcal{K}} \qquad \text{and} \qquad [f,g]'_{\mathcal{K}} = \lim_{n\to\infty} [f_n, g_n]'_{\mathcal{K}}.$$

But for each n: $[f_n, g_n]_{\mathcal{K}} = [f_n, g_n]'_{\mathcal{K}}$, and passing to the limit yields $[f,g]_{\mathcal{K}} = [f,g]'_{\mathcal{K}}$ on $\mathcal{K}$. Since $f, g \in \mathcal{K}$ were arbitrary, the two forms coincide on all of $\mathcal{K}$. □

Consider the fundamental decomposition $\mathcal{K} = \mathcal{H}_- \,[\oplus]\, \mathcal{H}_+$ of an RKKS $\mathcal{K}$ as the orthogonal direct sum of two Hilbert spaces $\mathcal{H}_\pm$. Since the evaluation functionals of $\mathcal{K}$ are continuous with respect to the strong topology, the associated Hilbert space is an RKHS. As $\mathcal{H}_\pm$ are Hilbert subspaces of the associated Hilbert space, both are also RKHS. Hence, each $\mathcal{H}_\pm$ admits a unique reproducing kernel $k_\pm : X \times X \to \mathbb{C}$. As each $\mathcal{H}_\pm$ is an RKHS, their reproducing kernels $k_\pm$ are positive semidefinite. Since kernel sections $k_x(\,\cdot\,)$ belong to $\mathcal{K}$, they also have unique decompositions $k_x(\,\cdot\,) = k_{x,+}(\,\cdot\,) + k_{x,-}(\,\cdot\,)$ with $k_{x,\pm}(\,\cdot\,) \in \mathcal{H}_\pm$ (likewise for $k_y(\,\cdot\,)$) [OMCS04, Prop. 6]. Furthermore, $k(x,y) = [k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}}$, and

$$k(x,y) = [k_y(\,\cdot\,), k_x(\,\cdot\,)]_{\mathcal{K}} = \underbrace{\langle k_{y,+}(\,\cdot\,), k_{x,+}(\,\cdot\,)\rangle_{\mathcal{H}_+}}_{k_+(x,y)} - \underbrace{\langle k_{y,-}(\,\cdot\,), k_{x,-}(\,\cdot\,)\rangle_{\mathcal{H}_-}}_{k_-(x,y)}, \tag{2.3.2}$$

for all $x, y \in X$. It follows that reproducing kernels are themselves decomposable across the fundamental decomposition, i.e., $k(x,y) = k_+(x,y) - k_-(x,y)$ for all $x, y \in X$ [OMCS04, Prop. 6]. We now state the converse to Theorem 2.15, which shows that while every RKKS has a unique reproducing kernel, a given kernel (satisfying certain conditions) may correspond to more than one RKKS. In this sense, the map from RKKS to reproducing kernels is "surjective" but not injective [OMCS04].

Theorem 2.18 (RKKS Construction Theorem [OMCS04, Prop. 7])**.** *Let X be a nonempty set and let $k : X \times X \to \mathbb{C}$ be a Hermitian kernel that admits a decomposition $k = k_+ - k_-$ where $k_\pm$ are positive semidefinite kernels on $X \times X$. Then there exists at least one RKKS $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$ of functions on X for which k is the reproducing kernel.*

Proof. The proof of this theorem is not covered in this thesis. The interested reader is referred to [OMCS04, Prop. 7]. □

The results of Theorem 2.15, Corollary 2.17, and Theorem 2.18 underline the central role of the reproducing kernel in RKKS theory. In an RKKS $(\mathcal{K}, [\cdot,\cdot]_{\mathcal{K}})$, there exists a unique reproducing kernel k, by Theorem 2.15. Moreover, the Kreĭn form is *exactly* the kernel-induced form (2.3.1), via Corollary 2.17, so it is uniquely determined by the reproducing kernel k. Hence, we may denote the RKKS form $[\cdot,\cdot]_{\mathcal{K}}$ as $[\cdot,\cdot]_k$, which is what we will do from now on. Conversely, any admissible Hermitian kernel $k = k_+ - k_-$, where $k_\pm$ are positive semidefinite, can be used to construct a (not necessarily unique) RKKS $(\mathcal{K}, [\cdot,\cdot]_k)$ with k as its reproducing kernel, by Theorem 2.18 [OMCS04].

Let X be a nonempty set and let $(\mathcal{K}, [\cdot,\cdot]_k)$ be an RKKS of functions $f : X \to \mathbb{C}$ with reproducing kernel k, and let $(x_i)_{i=1}^n \subseteq X$ be a finite ordered n-tuple of elements from X. Like any indefinite inner product, the kernel-induced form $[\cdot,\cdot]_k$ admits a quadratic form $Q(f) = [f,f]_k$ for all $f \in \mathcal{K}$, and a corresponding $n \times n$ *kernel Gram matrix* G for $(x_i)_{i=1}^n$ (through the Gram matrices, Definition 2.3, of the kernel sections of k) with entries

$$G_{ij} := \Big[k_{x_j}(\,\cdot\,), k_{x_i}(\,\cdot\,)\Big]_k = k(x_i, x_j). \tag{2.3.3}$$

Hence, $G = (G_{ij})_{i,j=1}^n = (k(x_i, x_j))_{i,j=1}^n$. Specifically, the reproducing kernel k is Hermitian if and only if, for every finite choice of $(x_i)_{i=1}^n$, the kernel Gram matrix $(k(x_i, x_j))_{i,j=1}^n$ is Hermitian. Similarly, k is positive semidefinite if and only if every such kernel Gram matrix G is positive semidefinite.

In addition, the supremum of the attainable numbers of positive and negative eigenvalues (squares of Q), counted with multiplicities, over all kernel Gram matrices (the quadratic form) coincides with the signature $(\kappa_-(\mathcal{K}), \kappa_+(\mathcal{K}))$ of the RKKS $\mathcal{K}$. Therefore, if the supremum of the number of negative eigenvalues over all kernel Gram matrices is finite, then $\mathcal{K}$ is an RKPS. If this supremum is zero, then $\mathcal{K}$ is an RKHS [Ghe13, ADRDS98].

3 Supervised Learning in Reproducing Kernel Kreĭn Spaces

This section introduces the main supervised machine learning setting of interest to this thesis, which sits at the intersection of reproducing kernel Kreĭn spaces (Section 2.3) and regularized empirical risk minimization. Solving such machine learning problems amounts to optimizing over a class of functions, which requires the existence (and uniqueness) of such a solution. In order to study the learning problem when functions belong to an RKKS, we first reinterpret objects relating to the Kreĭn space using the vernacular of the supervised learning literature in Section 3.1. We then state the general empirical risk minimization problem in an RKKS in Section 3.2. We then cover the first major representer theorem of learning (for stationary points) in an RKKS in Section 3.3. We then state a stronger version of the representer theorem (for minimizers) in Section 3.4. The overall scope of this thesis is supervised learning; consequently, unsupervised learning problems will not be covered. We assume a rudimentary familiarity with probability theory going forward.

3.1 RKKS in Statistical Learning Terminology

Here we reinterpret the RKKS framework for the purposes of formulating supervised learning problems. Let the *instance space* be a nonempty set $\mathcal{X}$, whose elements are called *training examples*, from which ordered input data $X = (x_1, \ldots, x_n)$ with $x_i \in \mathcal{X}$ may be drawn. Similarly, the *label space* is a nonempty set $\mathcal{Y}$ of *target* values (also called *labels*), from which ordered output data $Y = (y_1, \ldots, y_n)$ with $y_i \in \mathcal{Y}$ may be obtained. The availability of such a label space is what makes learning problems, which we introduce shortly, *supervised*. For this setting, we will only consider real-valued labels, i.e., $\mathcal{Y} \subseteq \mathbb{R}$, in line with the main references Ong et al. [OMCS04] and Oglic–Gärtner [OG18]. Finite ordered tuples $z = ((x_i, y_i))_{i=1}^n$ with elements from the *state space* $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ are called *ordered samples*. Note that the order of the samples will matter, but as a convenience we will simply refer to these as samples going forward. Later, when we define the learning objectives, the samples can be interpreted as realized outcomes of random variables (X_i, Y_i) defined on an underlying probability space.

Consider a Hermitian kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that admits a decomposition $k = k_+ - k_-$, where $k_\pm$ are positive semidefinite kernels on $\mathcal{X} \times \mathcal{X}$. Then, by

Theorem 2.18, there exists at least one RKKS $(\mathcal{K}, [\cdot,\cdot]_k)$ of real-valued functions on $\mathcal{X}$, that is $\mathcal{K} \subseteq \mathcal{F}(\mathcal{X};\mathcal{Y})$, here called the *feature space* (or the representation space of inputs), for which k is a reproducing kernel inducing the form $[\cdot,\cdot]_k$. Because we will work with the real scalar field $\mathbb{R}$ and real-valued kernels k, the conjugate symmetry condition $k(y,x) = \overline{k(x,y)}$, as we considered in Section 2, reduces to symmetry $k(x,y) = k(y,x)$ for all $x, y \in \mathcal{X}$. Consider any training example $x \in \mathcal{X}$. The same simplification applies to kernel sections at x. That is, $k_x(\,\cdot\,) = k(x,\cdot) = k(\cdot,x)$ as functions. Therefore, we may use all three notations interchangeably, but we will adopt $k(x,\cdot)$ and $k_x(\,\cdot\,)$ for the remainder of this survey to align with Ong et al. [OMCS04] and Oglic–Gärtner [OG18]. As a consequence, we obtain

$$[k_x(\,\cdot\,), k_y(\,\cdot\,)]_k = [k_y(\,\cdot\,), k_x(\,\cdot\,)]_k, \tag{3.1.1}$$

so either ordering may be used interchangeably in this section. We reiterate that in the complex-valued setting one only has $[k_x(\,\cdot\,), k_y(\,\cdot\,)]_k = \overline{[k_y(\,\cdot\,), k_x(\,\cdot\,)]_k}$ in general. The mapping $\phi : \mathcal{X} \to \mathcal{K}$ from the instance space into the RKKS is called the *feature map*. The image of x under ϕ is called the *feature* of x, and it coincides with the kernel section $\phi(x) = k_x(\,\cdot\,)$.

By *hypothesis space*, we mean the space of functions (or *hypotheses*) $f : \mathcal{X} \to \mathcal{Y}$ over which we search when solving a supervised learning problem. In our RKKS setting, the hypothesis space is identified with the feature space $\mathcal{K}$ itself, or with a suitably restricted subset of it (e.g., $\operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$, which is dense in $\mathcal{K}$ by Lemma 2.16). By the reproducing kernel property 2.14, any function $f \in \mathcal{K}$ represents a prediction function that may be evaluated as $f(x) = [f, k_x(\,\cdot\,)]_k$. By our discussion of Gram matrices in Section 2.3, for any ordered sample $((x_i, y_i))_{i=1}^n$, we can form the *kernel Gram matrix* $G \in \mathbb{R}^{n\times n}$ by evaluating the kernel on all pairs of inputs, $G_{ij} = k(x_i, x_j) = [k_{x_i}(\,\cdot\,), k_{x_j}(\,\cdot\,)]_k$, using the symmetric notation. The kernel Gram matrix encodes all pairwise inner products between the features $k_{x_i}(\,\cdot\,)$ and $k_{x_j}(\,\cdot\,)$, and therefore contains all the information that the RKKS contributes to the learning task.

3.2 The Empirical Risk Minimization Problem in RKKS

This section specifies the general empirical risk minimization problem in the RKKS setting from which specific problem formulations will be developed in Sections 3.3 and 3.4. Supervised learning in reproducing kernel Kreĭn spaces (following Ong et al. [OMCS04] and Oglic–Gärtner [OG18]) requires a problem formulation that incorporates both a probabilistic framework of statistical learning and the functional-analytic geometry of an RKKS (already formulated in Section 2.3).

To develop the probabilistic framework, we draw on foundational concepts from measure theory, which are otherwise outside the scope of the thesis. Our goal is not to provide a rigorous measure-theoretic treatment of probability spaces, but simply to introduce the definitions needed to study the problems of Ong et al. [OMCS04] and Oglic–Gärtner [OG18] at the cursory level required here.

Let ρ be a probability measure on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. A sample $z = ((x_i, y_i))_{i=1}^n$ is obtained by drawing the pairs (x_i, y_i) i.i.d. from ρ. Given a measurable *loss function* $\ell : \mathbb{R} \times \mathcal{Y} \to [0, \infty)$, the *expected risk* R of a measurable hypothesis $f : \mathcal{X} \to \mathbb{R}$ is [OG18]

$$R(f) = \int_{\mathcal{X}\times\mathcal{Y}} \ell(f(x), y) d\rho(x, y), \tag{3.2.1}$$

where we assume that $\ell(f(x), y)$ is ρ-integrable. Statistical learning usually amounts to finding a function within the hypothesis space, in this setting an RKKS $(\mathcal{K}, [\cdot, \cdot]_k)$, that minimizes the expected risk (3.2.1). That is, we seek at least one (not necessarily unique) *minimizer* $f^* = \arg\min_{f\in\mathcal{K}} R(f)$, if such a minimizer exists. Here, $\min_{f\in\mathcal{K}} R(f)$ denotes the minimal value of R over $\mathcal{K}$ when it exists, and $\arg\min_{f\in\mathcal{K}} R(f)$ denotes the set of functions in $\mathcal{K}$ for which this minimum is attained. In applications, the underlying distribution ρ, and therefore the expected risk $R(f)$, is unknown. Thus, it becomes necessary to estimate the risk from a specific selection of data. For an i.i.d. ordered sample $z = ((x_i, y_i))_{i=1}^n \in \mathcal{Z}^n$, we define the *empirical risk* $\widehat{R}_n(f)$ of a fixed hypothesis f by

$$\widehat{R}_n(f) := \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i). \tag{3.2.2}$$

Hence, statistical learning amounts to *empirical risk minimization (ERM)*, i.e., finding an optimizer $\widehat{f} \in \arg\min_{f\in\mathcal{K}} \widehat{R}_n(f)$, which serves as an approximation of the true minimizer f^* in the RKKS $\mathcal{K}$.

Remark 3.1 (Statistical Properties of $\widehat{R}_n(f)$ [SSBD14, Chapter 4 & 6]). In passing, we record a few common common statistical properties of $\widehat{R}_n(f)$ under standard assumptions. The proofs of these are outside the scope of the thesis. Let $\mathbb{E}_\rho[\,\cdot\,]$ denote the expectation operator under the distribution ρ. Under the i.i.d. assumption, and for any fixed measurable hypothesis f, we have $\mathbb{E}_\rho[\widehat{R}_n(f)] = R(f)$. That is, the empirical risk $\widehat{R}_n(f)$ is an unbiased estimator of the risk $R(f)$ [SSBD14, p. 56].

In addition, by the strong law of large numbers, $\widehat{R}_n(f)$ converges to $R(f)$ almost surely for each fixed f, so $\widehat{R}_n$ is a consistent estimator of $R(f)$ [SSBD14, p. 56]. For the ERM procedure to be consistent uniformly over all $f \in \mathcal{K}$, we must require the additional constraint that the hypothesis space $\mathcal{K}$ satisfies the so-called *uniform convergence property*, the details to which are outside the scope of this thesis (for a statement and proof of this result, we refer to [SSBD14, Theorem 6.7]).

Because the underlying RKKS $\mathcal{K}$ may be large, even infinite-dimensional, minimizing the empirical risk on its own may lead to estimates $\widehat{f}$ that do not generalize well to new data. That is, although $\widehat{f}$ is constructed to attain the minimum empirical risk within an ordered sample $z_1 = ((x_i, y_i))_{i=1}^n$, the same function may perform poorly when the empirical risk is re-evaluated on a new, previously unseen, ordered sample z_2.

Remark 3.2 (Approximation-estimation decomposition [SSBD14, Chapter 5]). This issue of generalization is often described via the so-called canonical *approximation-estimation decomposition* in machine learning [SSBD14, Chapter 5]. Roughly, the excess risk $R(\widehat{f}) - R(f^*)$ of an ERM algorithm on new and unseen out-of-sample data can be separated into two contributions: *approximation error*, reflecting the bias introduced by restricting the class of admissible hypotheses $\mathcal{K}$, and *estimation error*, which arises from overfitting to the in-sample training data z_1 and depends on the complexity of the hypothesis class. Though we will not treat complexity formally, it generally refers to how rich or expressive the hypothesis space $\mathcal{K}$ is. Informally, complexity is how flexible the hypotheses are in fitting different patterns in the data. Introducing a broader, more flexible class of hypotheses, (e.g. allowing polynomials as opposed to only linear functions) tends to reduce approximation error but increase estimation error. When the loss ℓ is the squared error, this tradeoff is usually called the *bias-variance tradeoff.* We do not use or prove the formal decomposition in this thesis, but we keep this intuition in mind, for it motivates the use of regularization techniques adopted by Ong et al. [OMCS04] and Oglic–Gärtner [OG18].

By *regularization*, we mean techniques that restrict or control the level of complexity of the hypotheses in order to mitigate errors from overfitting. This is particularly important in the (potentially) very rich RKKS hypothesis spaces considered here. Let $X = (x_i)_{i=1}^n$ be a finite ordered sample with elements from a nonempty set $\mathcal{X}$, and let $\mathcal{K}$ be an RKKS of real-valued functions on $\mathcal{X}$ (i.e., the label space $\mathcal{Y}$

is a subset of $\mathbb{R}$). The map $\mathrm{Ev}_X : \mathcal{K} \to \mathcal{Y}^n$, defined as

$$\mathrm{Ev}_X(f) = (f(x_1), \ldots, f(x_n)), \tag{3.2.3}$$

is called the *evaluation vector* of the input data X under f. All pointwise evaluation functionals are bounded due to their continuity with respect to the strong topology on $\mathcal{K}$. It follows that Ev_X is also bounded in the same topology. In line with Ong et al. [OMCS04] and Oglic–Gärtner [OG18], we consider the general (constrained) *regularized empirical risk minimization* problem for supervised learning in the RKKS $\mathcal{K}$. We assume that the empirical risk $\widehat{R}_n$ and the constraint function $C : \mathbb{R}^n \to \mathbb{R}^m$ depend on a hypothesis f only through its evaluation vector $\mathrm{Ev}_X(f) \in \mathcal{Y}^n \subseteq \mathbb{R}^n$. We formulate the problem as

$$\begin{aligned} &\min_{f \in \mathcal{K}} \left\{ \widehat{R}_n(\mathrm{Ev}_X(f)) + \lambda\, \Phi\big(f\big) \right\} \\ &\text{subject to } \; C(\mathrm{Ev}_X(f)) \in S. \end{aligned} \tag{3.2.4}$$

Here $\Phi : \mathcal{K} \to \mathbb{R}$ is a given *penalty function*, and $\lambda\Phi(f)$ is called the *regularization term*, with a *hyperparameter* λ acting as a weight on the penalty. A hyperparameter is a model parameter that is set to some constant value before solving the problem. Typically, this constant is selected either using an informed prior or using a data-driven validation procedure on unseen data. This regularization term penalizes the complexity of the hypothesis. The evaluation-based constraint function C is also prespecified, and the condition is restricted to a selected feasible set $S \subseteq \mathbb{R}^m$. The existence of a solution to the ERM problem requires the existence of at least one function $f \in \mathcal{K}$ such that $C(\mathrm{Ev}_X(f)) \in S$. Reusing the notation for the empirical minimizer, a solution to (3.2.4) is denoted by $\widehat{f}$. This formulation connects the probabilistic framework of statistical learning with the functional-analytic geometry of reproducing kernel Kreĭn spaces. Posing different constrained regularized problems of the form (3.2.4) "instantiates" different theoretical settings in which the RKKS supervised learning task is performed. We will consider the two most canonical instantiations of the ERM problem, where one formulation involves finding stationary points and leads to a weak representer theorem (Section 3.3). The other is formulated in terms of minimizers and produces a strong representer theorem (Section 3.4).

3.3 The Weak Representer Theorem

This section introduces the Ong et al. [OMCS04] learning problem, which replaces the minimization in (3.2.4) with stabilization (defined shortly). Let $\mathcal{X}$ be a nonempty set and let $X = (x_i)_{i=1}^n$ be an ordered sample with elements in $\mathcal{X}$. Let $(\mathcal{K}, [\cdot,\cdot]_k)$ be an RKKS of real-valued functions on X, that is, a subspace of $\mathcal{F}(\mathcal{X};\mathcal{Y})$ with label space $\mathcal{Y} \subseteq \mathbb{R}$, with reproducing kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. We now introduce the notions of derivative and stationarity used in this setting, without pursuing their theory beyond the stated definitions. We restrict ourselves only to the concepts necessary to state the weak representer theorem. Recall that the strong topology on $\mathcal{K}$ is exactly the Hilbert space topology induced by an associated Hilbert norm (2.2.3). Notions of neighborhoods, convergence, and continuity below are taken with respect to this topology, and we assume the reader is familiar with these concepts.

Definition 3.3 (Gâteaux derivative [Lue69, Section 7.2])**.** Let $(\mathcal{K}, [\cdot,\cdot]_k)$ be an RKKS with associated Hilbert space $(\mathcal{H}_{\mathcal{K}}, \langle\cdot,\cdot\rangle_J)$. Let $F : \mathcal{K} \to \mathbb{R}$ be a functional. We say F is *Gâteaux differentiable* at $f \in \mathcal{K}$ if, for each $h \in \mathcal{K}$, the limit

$$DF(f)[h] \;:=\; \lim_{t\to 0} \frac{F(f+th) - F(f)}{t}$$

exists (as a finite real number), and the map $h \mapsto DF(f)[h]$ is linear and continuous on $\mathcal{K}$ in the strong topology. The linear functional $DF(f) : \mathcal{K} \to \mathbb{R}$ is called the *Gâteaux derivative* of F at f.

Definition 3.4 (Stationary point [Lue69, Section 7.4])**.** Let the functional $F : \mathcal{K} \to \mathbb{R}$ be Gâteaux differentiable on an RKKS $\mathcal{K}$. A point $g \in \mathcal{K}$ is called a *stationary point* of F if

$$DF(g)[h] \;=\; 0, \quad \text{for all } h \in \mathcal{K}.$$

Stationary points can be local minima, local maxima, or saddle points. Since we will not distinguish between these cases in this thesis, we do not pursue their finer classification further. For a functional $F : \mathcal{K} \to \mathbb{R}$ we write $\operatorname{stab}_{g\in\mathcal{K}} F(g)$ to denote the stationary points of F (if they exist), the functions in $\mathcal{K}$ for which the Gâteaux derivative of F vanishes. Let L be a convex loss function that is continuous with respect to the strong topology on an RKKS $\mathcal{K}$. We assume L depends on $f \in \mathcal{K}$ only through its evaluation vector $\mathrm{Ev}_X(f)$, and we will use the notation $L(\mathrm{Ev}_X(f))$ to denote its value at the evaluation vector of f on X. Recall the quadratic form $Q : f \mapsto [f,f]_k$. We assume the functional $f \mapsto \Omega(Q(f))$ is continuous on $\mathcal{K}$ with

$\Omega : \mathbb{R} \to \mathbb{R}$ a strictly monotonic function. As in the general ERM problem in (3.2.4), we assume a constraint functional C depends on f only through its evaluations on X, and we denote its evaluations as $C(\mathrm{Ev}_X(f))$. The feasible set is a coordinate-wise upper-bounded set of evaluations $S_{\leq d} = \{x \in \mathbb{R}^m : x_i \leq d_i \text{ componentwise}\}$ with $d \in \mathbb{R}^m$ [OMCS04].

Definition 3.5 (Stabilization Problem [OMCS04, Theorem 11])**.** Let $\mathcal{K}$ be an RKKS with kernel k, and let $L, \Omega, C, S_{\leq d}, X$ be defined as above. The supervised *stabilization problem* over the RKKS is defined as

$$\begin{aligned} \operatorname*{stab}_{f \in \mathcal{K}} \; & L(\mathrm{Ev}_X(f)) \;+\; \Omega\big(Q(f)\big) \\ \text{subject to} \;\; & C(\mathrm{Ev}_X(f)) \in S_{\leq d}. \end{aligned} \tag{3.3.1}$$

Specifically, we seek stationary points of the functional $f \mapsto L(\mathrm{Ev}_X(f)) + \Omega(Q(f))$. In comparison to the general problem in (3.2.4), the empirical risk $\widehat{R}_n$ has been instantiated by $L(\mathrm{Ev}_X(f))$ and the hyperparameter λ in the regularization term $\lambda\Phi(f)$ has been embedded into Ω. If this constrained stabilization problem has at least one stationary point, then this solution will admit a (not necessarily unique) representation through a finite expansion in the *data span* $\operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$ built from the reproducing kernel k. This is referred to as the *weak* representer theorem, also called the representer theorem for stabilization. We will state this result, but the proof is omitted since it involves a more advanced treatment of Gâteaux derivatives and functional subdifferentials, which falls outside the scope of this thesis.

Theorem 3.6 (Weak Representer Theorem [OMCS04, Theorem 11])**.** *If the stabilized problem in Definition 3.5 admits a stationary point $\widehat{f} \in \mathcal{K}$, then there exists a real coefficient vector $\alpha \in \mathbb{R}^n$ for which the expansion $\widehat{f}(\,\cdot\,) = \sum_{i=1}^n \alpha_i \, k(x_i, \cdot)$ holds for $x_i \in X$.*

Proof. See [OMCS04, Theorem 11] for details. □

The weak representer theorem says that every stationary solution (be it a local minimum, local maximum, or saddle point) to the stabilization problem lies in the $\operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$. However, not every stationary point is a minima, and in an RKKS the regularized empirical risk $L(\mathrm{Ev}_X(f)) + \Omega(Q(f))$ can be unbounded below because the kernel-induced form $[\cdot,\cdot]_k$ is indefinite in general. Therefore, the theorem does not guarantee the existence of a minimum, even if there exist stationary solutions. But if a minimizer does exist, then it will admit an expansion of this form.

Moreover, even if $\widehat{f}$ is unique, the coefficient vector α in the expansion need not be unique [OMCS04].

Remark 3.7. As a comment, the strength of the theorem is that the (potentially) infinite-dimensional stabilization problem 3.5 can be reduced to a finite-dimensional stabilization problem over Euclidean space (the dual problem) [OMCS04]. We will not cover this in this stabilization setting, but rather we will show the analogous result in the minimization setting; see Corollary 3.13.

For an ordered training sample X and a reproducing kernel k, let $G \in \mathbb{R}^{n\times n}$ be the corresponding kernel Gram matrix (see Equation (2.3.3)). We now show three core identities in kernel Gram matrix algebra.

Corollary 3.8 (Kernelization Identities)**.** *Let $\mathcal{X}$ be a nonempty set and $\mathcal{Y} \subseteq \mathbb{R}$ be a set of real-valued labels, and let $X = (x_i)_{i=1}^n$ be a finite ordered training sample of elements in $\mathcal{X}$. Consider the RKKS $(\mathcal{K}, [\cdot,\cdot]_k)$ such that $\mathcal{K} \subseteq \mathcal{F}(\mathcal{X};\mathcal{Y})$ and k is its reproducing kernel. Let $G \in \mathbb{R}^{n\times n}$ be the kernel Gram matrix of X under k. Suppose the conditions of Theorem 3.6 hold, and let $f, g \in \operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$. Then there exist $\alpha, \beta \in \mathbb{R}^n$ such that*

$$\mathrm{Ev}_X(f) = G\alpha, \qquad \textit{and} \qquad [f,f]_k = \alpha^\top G\alpha, \qquad \textit{and} \qquad [f,g]_k = \alpha^\top G\beta.$$

In particular, these identities apply to any solution of the stabilization problem 3.5.

Proof. Because f, g lie in the data span $\operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$, there exist coefficient vectors $\alpha, \beta \in \mathbb{R}^n$ such that a representer expansion in the reproducing kernel can be constructed, $f(\,\cdot\,) = \sum_i \alpha_i k(x_i,\cdot)$ and $g(\,\cdot\,) = \sum_i \beta_i k(x_i,\cdot)$. Recall that in the real setting the Hermitian kernel is real and symmetric, so $k(x_i,\cdot) = k(\cdot,x_i)$ and Hermitian symmetry reduces to symmetry, so $[k_{x_i}(\,\cdot\,), k_{x_j}(\,\cdot\,)]_k = [k_{x_j}(\,\cdot\,), k_{x_i}(\,\cdot\,)]_k$. Similarly, sesquilinearity reduces to bilinearity. Consider the jth component of the evaluation vector $\mathrm{Ev}_X(f)$. Using $G_{ij} = k(x_i,x_j)$, we get $G_{ij} = [k_{x_i}(\,\cdot\,), k_{x_j}(\,\cdot\,)]_k$ by the symmetry and reproducing property of k. In addition, we also obtain

$$\mathrm{Ev}_X(f)_j = f(x_j) = \sum_{i=1}^n \alpha_i k(x_i,x_j) = (G\alpha)_j.$$

Because this holds for each component, $\mathrm{Ev}_X(f) = G\alpha$ follows. Furthermore,

$$[f,g]_k = \sum_{i,j} \alpha_i\beta_j\,[k_{x_i}(\,\cdot\,), k_{x_j}(\,\cdot\,)]_k = \sum_{i,j} \alpha_i\beta_j\,G_{ij} = \alpha^\top G\beta,$$

where $\sum_{i,j}$ denotes the double sum over indices i and j. Finally, assuming $g = f$ gives $Q(f) = [f, f]_k = \alpha^\top G\alpha$. Moreover, because solutions to problem 3.5 lie in the data span, by Theorem 3.6, these identities apply to such solutions as well. □

The *kernelization identities* show that, once f and g lie in the data span, evaluations and inner products (and hence the quadratic form) reduce to finite-dimensional Gram-matrix algebra: $\mathrm{Ev}_X(f) = G\alpha$ and $[f, g]_k = \alpha^\top G\beta$. Consequently, any evaluation-based quantity (for example, the regularized risk and constraints) depends on the data X only through the kernel Gram matrix G and the coefficient vectors.

Remark 3.9 (Kernel trick). Informally, any learning algorithm that uses data X only via inner products of feature maps is said to use the **kernel trick**. In our setting the feature map is given by $\phi(x) = k_x(\,\cdot\,)$, and the inner products of features satisfy

$$[\phi(x), \phi(y)]_k = [k_x(\,\cdot\,), k_y(\,\cdot\,)]_k = k(x, y), \quad \text{for all } x, y \in X,$$

by the reproducing property and the fact that the reproducing kernel k is real-valued (hence symmetric). This so-called kernel trick allows us to replace the inner products of feature maps with their corresponding kernel evaluations $k(x, y)$. This obviates the need to construct the (possibly infinite-dimensional) feature maps $\phi(x)$ explicitly [SS02, Chapter 2.2.2]. In the stabilization problem 3.5, the kernel trick is fully available. By the weak representer theorem 3.6 and Corollary 3.8, any stationary solution $\widehat{f}$ to this problem belongs to the data span, so all its evaluations can be computed via the kernelization identities without ever constructing a feature map. Solving for stationary points thus becomes a finite-dimensional computation in (G, α), with no explicit feature map required [OMCS04]. We do not treat the dual of the stabilization problem in this thesis, but instead present the analogous result for the minimization setting in Corollary 3.13.

3.4 The Strong Representer Theorem

This section introduces an alternative minimizer-oriented instantiation of the regularized ERM problem (3.2.4), following Oglic–Gärtner [OG18]. Let $\mathcal{X}$ be a nonempty set and let $(\mathcal{K}, [\cdot, \cdot]_k)$ be an RKKS of real-valued functions $f : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{Y} \subseteq \mathbb{R}$, with fundamental decomposition $\mathcal{K} = \mathcal{H}_+ [\oplus] \mathcal{H}_-$ and a reproducing kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For an ordered sample $z = ((x_i, y_i))_{i=1}^n \in (\mathcal{X} \times \mathcal{Y})^n$, the risk is taken as the expected squared error under a probability measure ρ on $\mathcal{X} \times \mathcal{Y}$, i.e., $R(f) = \int_{\mathcal{X}\times\mathcal{Y}} (f(x) - y)^2 \, d\rho(x, y)$. The empirical counterpart is the *mean squared error* $\widehat{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2$, which is convex in f and depends on f only through its evaluations $\mathrm{Ev}_{x_i}(f) = f(x_i)$. For any hypothesis $f = f_+ + f_-$, with $f_\pm \in \mathcal{H}_\pm$, we penalize the complexity of f in the strong topology on $\mathcal{K}$ using a so-called *Tikhonov regularization term*, which is of the form $\lambda_+\|f_+\|^2_{\mathcal{H}_+} + \lambda_-\|f_-\|^2_{\mathcal{H}_-}$, with hyperparameters $\lambda_\pm > 0$. The regularized ERM is further constrained by fixing the empirical *variance* of the evaluations to a prescribed hyperparameter r^2.

Definition 3.10 (Minimization Problem [OG18, KREĬN Problem, Equation 3])**.** Let $\mathcal{X}$ be a nonempty set and $\mathcal{K} = \mathcal{H}_+ [\oplus] \mathcal{H}_-$ be an RKKS with reproducing kernel k and associated RKHS norms $\|\cdot\|_{\mathcal{H}_\pm}$. Given ordered training data $X = (x_i)_{i=1}^n \in \mathcal{X}^n$, ordered real-valued labels $y_i \in \mathcal{Y} \subseteq \mathbb{R}$, and fixed constants $\lambda_\pm > 0$ and $r > 0$, the *Kreĭn minimization problem* is

$$
\begin{aligned}
\min_{f\in\mathcal{K}} \quad & \frac{1}{n}\sum_{i=1}^{n}\Big(f(x_i) - y_i\Big)^2 \;+\; \lambda_+\|f_+\|^2_{\mathcal{H}_+} \;+\; \lambda_-\|f_-\|^2_{\mathcal{H}_-} \\
\text{subject to} \quad & \frac{1}{n}\sum_{i=1}^{n}\Big(f(x_i) - \bar{f}\Big)^2 \;=\; r^2.
\end{aligned}
\tag{3.4.1}
$$

Here $f = f_+ + f_-$, for $f_\pm \in \mathcal{H}_\pm$, and $\bar{f} := \frac{1}{n}\sum_{i=1}^{n} f(x_i)$ is the average of evaluations.

Because the evaluation vector operator $\mathrm{Ev}_X : \mathcal{K} \to \mathcal{Y}^n$ with $\mathcal{Y} \subseteq \mathbb{R}$ is linear and bounded in an RKKS (with respect to the strong topology), and the mean-squared error mapping $\omega \mapsto \frac{1}{n}\sum_{i=1}^{n}(\omega_i - y_i)^2$ is convex and continuous, the composed loss $\widehat{R}_n(f)$ is also convex and continuous (hence lower semi-continuous) on $\mathcal{K}$. The penalty term corresponds to the square of a weighted fundamental norm, equivalent to the fundamental norm (2.2.5) when $\lambda_\pm > 0$. In contrast, "naive Tikhonov" regularization would penalize the indefinite Kreĭn form, $[f, f]_k = \|f_+\|^2_{\mathcal{H}_+} - \|f_-\|^2_{\mathcal{H}_-}$, which is not a valid norm and can render the regularized risk unbounded below.

Remark 3.11 (On convexity [OG18])*.* Because the empirical risk $\widehat{R}_n(f)$ is convex and continuous, and the regularizer $\lambda_+\|f_+\|^2_{\mathcal{H}_+} + \lambda_-\|f_-\|^2_{\mathcal{H}_-}$ is convex, the unconstrained problem is also convex. The variance constraint introduces a quadratic equality, whose level sets are generally nonconvex, so the constrained problem in Definition 3.10 is typically nonconvex. We include this remark only for context. A detailed analysis of convexity and nonconvex constraints would require additional optimization theory and lies beyond the scope of this thesis.

The variance constraint fixes the scale of the centered evaluations $f(x_i) - \bar{f}$ and rephrases the problem as minimizing a quadratic objective over a hyperellipsoid of radius r [OG18]. We can now state the *strong representer theorem*, also called the representer theorem for minimization.

Theorem 3.12 (Strong Representer Theorem [OG18, Theorem 2])**.** *If $\widehat{f} \in \mathcal{K}$ is an optimal solution to the minimization problem in Definition 3.10, then $\widehat{f}$ admits an expansion $\widehat{f} = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$ with $\alpha_i \in \mathbb{R}$.*

Proof. As $\mathcal{K}$ is an RKKS, its reproducing kernel is decomposable $k = k_+ - k_-$, where $k_\pm$ are the reproducing kernels for the RKHS $\mathcal{H}_\pm$. Any hypothesis $f \in \mathcal{K}$ can be written as $f_+ + f_-$ for $f_\pm \in \mathcal{H}_\pm$. Define the data spans $\mathcal{H}_\pm(X)$ of an ordered training sample X inside each RKHS by

$$\mathcal{H}_+(X) := \operatorname{span}\{k_+(x, \cdot\,) : x \in X\} \quad \text{and} \quad \mathcal{H}_-(X) := \operatorname{span}\{k_-(x, \cdot\,) : x \in X\},$$

where $\mathcal{H}_+(X) \subseteq \mathcal{H}_+$ and $\mathcal{H}_-(X) \subseteq \mathcal{H}_-$. We denote by $\mathcal{H}_+^\perp(X)$ the orthogonal companion (see equation (2.1.1)) of the data span $\mathcal{H}_+(X)$ in $\mathcal{H}_+$, and $\mathcal{H}_-^\perp(X)$ is defined analogously in $\mathcal{H}_-$. Note that this construction yields $\mathcal{H}_+ = \mathcal{H}_+(X)\,[\oplus]\,\mathcal{H}_+^\perp(X)$ and $\mathcal{H}_- = \mathcal{H}_-(X)\,[\oplus]\,\mathcal{H}_-^\perp(X)$, respectively. As a result, we can decompose $f_\pm \in \mathcal{H}_\pm$ as $f_+ = u_+ + v_+$ and $f_- = u_- + v_-$ for $u_\pm \in \mathcal{H}_\pm(X)$ and $v_\pm \in \mathcal{H}_\pm^\perp(X)$. Because any $v_\pm$ belongs to the orthogonal companion of the data span and $k_\pm(x, \cdot) \in \mathcal{H}_\pm(X)$, it holds that $v_\pm(x) = \langle v_\pm, k_\pm(x, \cdot)\rangle_{\mathcal{H}_\pm} = 0$ for all $x \in X$ by the reproducing property 2.14 in each RKHS. We will use the bilinearity of the form $[\cdot, \cdot]_k$, and the symmetry $k(x, \cdot) = k(\cdot, x)$ recovered from this real-valued setting. We start from the reproducing property on the RKKS, which asserts that $f(x) = [f, k(x, \cdot)]_k$ for all $f \in \mathcal{K}$. Then by writing out the kernel sections $k(x, \cdot) = k_+(x, \cdot) + k_-(x, \cdot)$ as an element of $\mathcal{H}_+[\oplus]\mathcal{H}_-$, and by expanding out the form $[\cdot, \cdot]_k$ with bilinearity we get

$$\begin{aligned} f(x) &= [f, k(x, \cdot)]_k \\ &= [(u_+ + v_+) + (u_- + v_-), k_+(x, \cdot) + k_-(x, \cdot)]_k \\ &= \langle u_+, k_+(x, \cdot)\rangle_{\mathcal{H}_+} + \langle v_+, k_+(x, \cdot)\rangle_{\mathcal{H}_+} - \langle u_-, k_-(x, \cdot)\rangle_{\mathcal{H}_-} - \langle v_-, k_-(x, \cdot)\rangle_{\mathcal{H}_-} \\ &= \langle u_+, k_+(x, \cdot)\rangle_{\mathcal{H}_+} - \langle u_-, k_-(x, \cdot)\rangle_{\mathcal{H}_-} \\ &= u_+(x) - u_-(x). \end{aligned}$$

In the last step we used the reproducing properties of the reproducing kernels $k_\pm$ in $\mathcal{H}_\pm$. By defining $u(x) := u_+(x) - u_-(x)$ we have that $f(x) = u(x)$ for all $x \in X$. By orthogonality, we have

$$\|f_\pm\|^2 = \|u_\pm + v_\pm\|^2 = \|u_\pm\|^2_{\mathcal{H}_\pm} + \|v_\pm\|^2_{\mathcal{H}_\pm}.$$

Hence, the minimization problem in Definition 3.10 may be restated as

$$\begin{aligned} &\min_{f\in\mathcal{K}} \frac{1}{n}\sum_{i=1}^{n}\Big(u(x_i) - y_i\Big)^2 + \lambda_+ \left(\|u_+\|^2_{\mathcal{H}_+} + \|v_+\|^2_{\mathcal{H}_+}\right) + \lambda_- \left(\|u_-\|^2_{\mathcal{H}_-} + \|v_-\|^2_{\mathcal{H}_-}\right) \\ &\text{subject to } \frac{1}{n}\sum_{i=1}^{n}\Big(u(x_i) - \bar{u}\Big)^2 = r^2. \end{aligned}$$

Here we define $\bar{u} := \frac{1}{n}\sum_{i=1}^{n} u(x_i)$. Note that only the regularization term depends on $v_\pm$. Therefore, by the nonnegativity of the RKHS norms $\|\cdot\|_{\mathcal{H}_\pm}$, any minimizer will satisfy $v_\pm = 0$. Thus, the entire minimization problem is stated only in terms of components u_+ and u_-. Specifically, the solution is of the form $\widehat{f} = \widehat{u}_+ + \widehat{u}_-$ and so $\widehat{f} \in \mathcal{H}_+(X)[\oplus]\mathcal{H}_-(X)$, a finite-dimensional subspace of $\mathcal{K}$. In particular, Oglic and Gärtner assert (see [OG18, Theorem 2 and Appendix A]) that for this problem there exists such an optimal solution which admits a representation

$$\widehat{f} = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot),$$

for some $\alpha \in \mathbb{R}^n$, though we refer to Oglic and Gärtner for this final step. This completes the proof. □

By the strong representer theorem 3.12, any minimizer $\widehat{f}$ of the empirical minimization problem in Definition 3.10 lies in $\operatorname{span}\{k_x(\,\cdot\,) : x \in X\}$. As a result, every term in the objective and constraint becomes finite-dimensional. In other words, the representer theorem turns the infinite-dimensional search over the RKKS $\mathcal{K}$ into an algebraic problem in α alone, which we now formulate.

Corollary 3.13 (Coefficient–Space Dual Problem [OG18, Equation 4])**.** *Consider the minimization problem in Definition 3.10 and let G be the Gram matrix of a sample $X = (x_i)_{i=1}^n$ under the reproducing kernel $k = k_+ - k_-$ and $G_\pm$ be the Gram matrices under $k_\pm$. Then the minimization problem is equivalent to the finite-dimensional problem*

$$\begin{aligned} &\min_{\alpha\in\mathbb{R}^n} \ \|G\alpha - y\|_2^2 \ + \ n\,\alpha^\top\Big(\lambda_+ G_+ + \lambda_- G_-\Big)\alpha \\ &\textit{subject to} \ \ \alpha^\top G^2 \alpha = nr^2. \end{aligned} \tag{3.4.2}$$

Here $\|\cdot\|_2$ denotes the standard Euclidean norm.

Proof. Let $X = (x_i)_{i=1}^n$ be an ordered sample with elements in a nonempty set $\mathcal{X}$ and $\mathcal{K} = \mathcal{H}_+[\oplus]\mathcal{H}_-$ be an RKKS of real-valued functions on $\mathcal{X}$ with reproducing kernel $k = k_+ - k_-$, where $k_\pm : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are the reproducing kernels of the RKHS $\mathcal{H}_\pm$. By the strong representer theorem 3.12, any minimizer of problem 3.10 admits an expansion $f(\cdot) = \sum_{j=1}^n \alpha_j k(x_j, \cdot)$ for some $\alpha \in \mathbb{R}^n$. We start from the kernelization identity $\mathrm{Ev}_X(f) = G\alpha$ (from Corollary 3.8), which means $f(x_i) = (G\alpha)_i$. Plugging this into the empirical loss term $\widehat{R}_n$ gives

$$\widehat{R}_n(f) = \tfrac{1}{n} \sum_i ((G\alpha)_i - y_i)^2 = \tfrac{1}{n}\|G\alpha - y\|_2^2.$$

We decompose the minimizer $f = f_+ + f_-$ for which $f_\pm \in \mathcal{H}_\pm$. In the proof of Theorem 3.12 we showed that f_+ belongs to the data span $\mathcal{H}_+(X) = \mathrm{span}\{k_+(x, \cdot) : x \in X\} \subseteq \mathcal{H}_+$ and f_- belongs to the data span $\mathcal{H}_-(X) = \mathrm{span}\{k_-(x, \cdot) : x \in X\} \subseteq \mathcal{H}_-$. Since each kernel section in $\mathcal{K}$ decomposes according to the fundamental decomposition, $k(x_j, \cdot) = k_+(x_j, \cdot) + k_-(x_j, \cdot)$, we may write $f_\pm(\cdot) = \sum_{j=1}^n \alpha_j k_\pm(x_j, \cdot)$. Because $\mathcal{H}_\pm$ are both RKHS, $\|f_\pm\|_{\mathcal{H}_\pm}^2 = \langle f_\pm, f_\pm \rangle_{\mathcal{H}_\pm}$, and by the Kernelization identities 3.8 we have $\|f_+\|_{\mathcal{H}_+}^2 = \alpha^\top G_+ \alpha$ and $\|f_-\|_{\mathcal{H}_-}^2 = \alpha^\top G_- \alpha$. Therefore, the strong-topology penalty corresponds to

$$\lambda_+ \, \alpha^\top G_+ \alpha + \lambda_- \, \alpha^\top G_- \alpha = \alpha^\top \left(\lambda_+ G_+ + \lambda_- G_-\right) \alpha.$$

Because the Gram matrix is assumed to be centered, $\mathbf{1}^\top G = 0$, and because $\mathrm{Ev}_X(f) = G\alpha$, we have $\bar{f} = \frac{1}{n}\mathbf{1}^\top G\alpha = 0$. Therefore, the variance constraint becomes

$$r^2 = \frac{1}{n} \sum_{i=1}^n \left(f(x_i) - \bar{f}\right)^2 \;=\; \frac{1}{n}\,\|G\alpha\|_2^2 = \frac{1}{n}\,\alpha^\top G^2 \alpha.$$

Finally, after collecting the terms, we obtain

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^n} \quad & \frac{1}{n}\,\|G\alpha - y\|_2^2 + \alpha^\top \Big(\lambda_+ G_+ + \lambda_- G_-\Big)\alpha \\ \text{subject to} \quad & \frac{1}{n}\alpha^\top G^2 \alpha = r^2. \end{aligned}$$

Multiplying the objective and the equality constraint by the sample size n (which does not change the minimizers or the feasible set) yields the desired result. □

The dual problem formulation of Corollary 3.13 matters because it turns the function space problem in an RKKS $\mathcal{K}$ into a simple finite-dimensional problem, where only the real coefficients and the Gram matrices $(\alpha, G, G_\pm)$ matter for learning. Therefore, similar to the stabilization problem in Definition 3.5, the kernel trick is operational in this setting, and no explicit feature map is needed for learning. Once α is found, prediction for a new unseen data point $\tilde{x} \in \mathcal{X}$ is performed using $f(\tilde{x}) = \sum_{i=1}^{n} \alpha_i k(\tilde{x}, x_i)$. Because the RKKS allows for such a reformulation, this is what makes supervised learning in an RKKS tractable. Practitioners may fine-tune the hyperparameters $(\lambda_\pm, r)$ using standard validation procedures.

To make a comparison, the stabilization setup of Ong et al. [OMCS04], presented in Section 3.3, frames learning as finding stationary points in an RKKS where the regularizer depends on the indefinite Kreĭn form $[f, f]_k$. This yields a weak representer theorem in which stationary solutions lie in the data span. However, the objective function can be unbounded from below, so minimizers may not exist. In contrast, the minimization setup of Oglic–Gärtner [OG18] uses a Tikhonov-style penalty built from the fundamental norm together with a variance constraint. This leads to a strong representer theorem in which minimizers lie in the data span. In both these settings, the learning problems can be reformulated as finite-dimensional dual problems in coefficient space, which lend themselves neatly to kernel Gram matrix algebra.

4 Conclusion

The contribution of this thesis is not the derivation of new results, but rather a systematic exposition that connects Kreĭn space geometry with kernel methods in supervised learning. The journey begins with the foundational axioms of an indefinite inner product space, and the final destination is the two core representer theorems of supervised learning problems in reproducing kernel Kreĭn spaces. The emphasis on indefiniteness of the inner product makes the functional-geometric structure more general and more mathematically rich than the typical Hilbert space setting. In particular, it requires a specific set of concepts to make the learning problem better posed, e.g., the fundamental decomposition, the fundamental symmetry, the fundamental norm, and the strong topology induced by the associated Hilbert space. However, the crucial ingredient to make learning possible is that of a reproducing kernel, which admits a unique reproducing kernel Kreĭn space (RKKS), where all evaluation functionals are continuous by construction. The RKKS inner product equals the kernel-induced form, so the geometry and the kernel coincide, allowing us to study the geometry indirectly through the kernel. Crucially, the reproducing property of the kernel allows us to represent functions in the data span using kernel sections.

After reinterpreting the RKKS in a supervised learning setting, we instantiate two well-established regularized supervised learning problems in an RKKS. First we consider a stabilization problem, and then we consider a minimization problem. We also show that the problem formulations have to be done with care, the indefiniteness of the inner product can lead to minimization problems that are unbounded below. We show how the solution to the stabilization problem is subject to a weak representer theorem and how the solution to the minimization problem is subject to a strong representer theorem. The representer theorems allow us to kernelize computations involving the inner product and all evaluations, rendering feature maps redundant in the representer-theorem settings, introducing the "kernel trick" to learning in an RKKS. Furthermore, although the stabilization and minimization problems are, generally, infinite-dimensional problems, the representer theorems allow us in both cases to reformulate the relevant optimality conditions as finite-dimensional problems in which the search is over coefficient vectors in Euclidean space. This dual/finite-dimensional formulation is much more tractable; learning becomes fully expressible in terms of linear algebra through the Gram matrix and a vector of real-valued coefficients, rather than a search over abstract functions in an RKKS. At the same time, several limitations remain. The weak representer theorem does not ensure the existence of minimizers, and the strong formulation introduces variance constraint to prevent ill-posedness caused by the indefinite directions of the Kreĭn structure. These difficulties indicate that some RKKS learning setups require further (and potentially restrictive) assumptions or regularization strategies.

By synthesizing results from Kreĭn geometry, reproducing kernels, and empirical risk minimization, the thesis provides a unified view of *machine learning with kernels in RKKS*. Overall, the results illustrate that, when paired with appropriate regularization, RKKS offer a viable and mathematically rich alternative to the corresponding Hilbert space setting. RKKS offer rich opportunities to apply indefinite kernels in machine learning.

References

[ADRDS98] Daniel Alpay, Aad Dijksma, James Rovnyak, and Hendrik S. V. De Snoo. *Reproducing Kernel Pontryagin Spaces*, page 425–444. Mathematical Sciences Research Institute Publications. Cambridge University Press, 1998.

[Fri82] Avril Friedman. *Foundations of Modern Analysis.* Dover Books on Mathematics Series. Dover Publications, Garden City, NY, New York, 1982.

[Ghe13] Aurelian Gheondea. A Survey on Reproducing Kernel Krein Spaces, 2013. arXiv.

[Ghe22] Aurelian Gheondea. *An Indefinite Excursion in Operator Theory: Geometric and Spectral Treks in Kreĭn Spaces*, volume 476 of *London Mathematical Society Lecture Note Series.* Cambridge University Press, 2022.

[Lue69] David G. Luenberger. *Optimization by vector space methods.* John Wiley & Sons Inc., New York, 1969. reprint 1998.

[OG18] Dino Oglic and Thomas Gärtner. Learning in reproducing kernel Kreın spaces. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3859–3867. PMLR, 2018.

[OMCS04] Cheng Soon Ong, Xavier Mary, Stéphane Canu, and Alexander J. Smola. Learning with non-positive kernels. In *Proceedings of the Twenty-First International Conference on Machine Learning*, volume 69 of *ACM International Conference Proceeding Series.* ACM, 2004.

[SS02] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, MA, USA, 2002.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press, 2014.

AI Statement

In preparing this thesis, I have used artificial intelligence (AI) tools under the conditions and limitations described below, and in accordance with the applicable guidelines and regulations Stockholm University:

AI tools used

The assistance provided by AI tools (in practice exclusively ChatGPT) was limited to the following supporting roles:

- **Language and style:** ChatGPT was used to suggest improvements in wording, propose rephrasings of sentences for improved clarity, provide feedback on organizational structure and sentence flow, and to check English grammar. All suggestions were critically reviewed individually and manually accepted, modified, and adapted; or rejected by me.
- **Technical assistance:** ChatGPT was used as a supporting tool for LaTeX formatting. This involved assistance with formatting isolated mathematical environments, suggesting improvements in the use of LaTeX notation, and acting as a code debugging assistant when compilation errors occurred. The final LaTeX code and implementation decisions were made at my own discretion.
- **Mathematical content:** ChatGPT was used as a supplementary pedagogical tool for the purpose of improving my understanding of the core references in the bibliography, strictly within the "Study and learn" setting mode offered by OpenAi. This usage involved walkthroughs of terminology and specialized vocabulary, rephrasing and simplifying mathematical definitions, and supplemental explanations and exemplification of results. This usage was at all times ancillary to, and only done in conjunction with, my overall study of the core references stated in the bibliography. AI tools were *not* used in the selection of the topic, the design of the methodology, or in the interpretation of results.

Place and date: 17/11/2025 Signature: Andreas Karlsson