

A Primer on Reproducing Kernel Hilbert Spaces

Jonathan H. Manton

The University of Melbourne
Victoria 3010 Australia
j.manton@ieee.org

Pierre-Olivier Amblard

CNRS
Grenoble 38402 France
pierre-olivier.amblard@gipsa-lab.grenoble-inp.fr

now

the essence of knowledge

Boston — Delft



Documents

Document Sharing Hub

Foundations and Trends® in Signal Processing

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

J. H. Manton and P.-O. Amblard. *A Primer on Reproducing Kernel Hilbert Spaces*.
Foundations and Trends® in Signal Processing, vol. 8, no. 1-2, pp. 1–126, 2014.

This Foundations and Trends® issue was typeset in L^AT_EX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-093-4

© 2015 J. H. Manton and P.-O. Amblard

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends® in Signal Processing

Volume 8, Issue 1-2, 2014

Editorial Board

Editor-in-Chief

Yonina Eldar

Technion - Israel Institute of Technology
Israel

Editors

Robert M. Gray
Founding Editor-in-Chief
Stanford University

Pao-Chi Chang
NCU, Taiwan

Pamela Cosman
UC San Diego

Michelle Effros
Caltech

Yariv Ephraim
GMU

Alfonso Farina
Selex ES

Sadaoki Furui
Tokyo Tech

Georgios Giannakis
University of Minnesota

Vivek Goyal
Boston University

Sinan Gunturk
Courant Institute

Christine Guillemot
INRIA

Robert W. Heath, Jr.
UT Austin

Sheila Hemami
Northeastern University

Lina Karam
Arizona State U

Nick Kingsbury
University of Cambridge

Alex Kot
NTU, Singapore

Jelena Kovacevic
CMU

Geert Leus
TU Delft

Jia Li
Penn State

Henrique Malvar
Microsoft Research

B.S. Manjunath
UC Santa Barbara

Urbashi Mitra
USC

Björn Ottersten
KTH Stockholm

Vincent Poor
Princeton University

Anna Scaglione
UC Davis

Mihaela van der Shaar
UCLA

Nicholas D. Sidiropoulos
TU Crete

Michael Unser
EPFL

P. P. Vaidyanathan
Caltech

Ami Wiesel
Hebrew U

Min Wu
University of Maryland

Josiane Zerubia
INRIA

Editorial Scope

Topics

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multirate signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing

Information for Librarians

Foundations and Trends® in Signal Processing, 2014, Volume 8, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Foundations and Trends® in Signal Processing
Vol. 8, No. 1-2 (2014) 1–126
© 2015 J. H. Manton and P.-O. Amblard
DOI: 10.1561/20000000050



A Primer on Reproducing Kernel Hilbert Spaces

Jonathan H. Manton
The University of Melbourne
Victoria 3010 Australia
j.manton@ieee.org

Pierre-Olivier Amblard
CNRS
Grenoble 38402 France
pierre-olivier.amblard@gipsa-lab.grenoble-inp.fr

Contents

1	Introduction	2
1.1	Assumed Knowledge	4
1.2	Extrinsic Geometry and a Motivating Example	5
1.3	Pointwise Coordinates and Canonical Coordinates	7
2	Finite-dimensional RKHSs	9
2.1	The Kernel of an Inner Product Subspace	10
2.2	Sequences of Inner Product Spaces	15
2.3	Extrinsic Geometry and Interpolation	16
2.4	Visualising RKHSs	18
2.5	RKHSs over the Complex Field	21
3	Function Spaces	23
3.1	Function Approximation	24
3.2	Topological Aspects of Inner Product Spaces	27
3.3	Evaluation Functionals	30
4	Infinite-dimensional RKHSs	32
4.1	Completions and Hilbert Spaces	33
4.2	Definition of a RKHS	36
4.3	Examples of RKHSs	37
4.4	Basic Properties	38

4.5	Completing a Function Space	40
4.6	Joint Properties of a RKHS and its Kernel	42
5	Geometry by Design	46
5.1	Embedding Points into a RKHS	46
5.2	The Gaussian Kernel	47
6	Applications to Linear Equations and Optimisation	49
6.1	Interpolation Problems	50
6.2	Solving Linear Equations	50
7	Applications to Stochastic Processes	56
7.1	Detection in Finite Dimensions	57
7.2	The RKHS Associated with a Stochastic Process	58
7.3	Signal Detection	59
8	Embeddings of Random Realisations	68
8.1	Finite Embeddings	69
8.2	Random elements in Hilbert spaces	71
8.3	Random elements in reproducing kernel Hilbert spaces	72
8.4	Universal and characteristic kernels	74
8.5	Empirical estimates of the mean elements and covariance operators	76
8.6	Generalisations and further considerations	77
9	Applications of Embeddings	80
9.1	Matrix representations of mean elements and covariance operators	81
9.2	Signal detection and the deflection criterion	83
9.3	Testing for independence	87
9.4	Conditional independence measures	93
9.5	Kernel Bayesian filtering	102
9.6	On-line filtering using RKHS	112
	References	121

Abstract

Reproducing kernel Hilbert spaces are elucidated without assuming prior familiarity with Hilbert spaces. Compared with extant pedagogic material, greater care is placed on motivating the definition of reproducing kernel Hilbert spaces and explaining when and why these spaces are efficacious. The novel viewpoint is that reproducing kernel Hilbert space theory studies extrinsic geometry, associating with each geometric configuration a canonical overdetermined coordinate system. This coordinate system varies continuously with changing geometric configurations, making it well-suited for studying problems whose solutions also vary continuously with changing geometry. This primer can also serve as an introduction to infinite-dimensional linear algebra because reproducing kernel Hilbert spaces have more properties in common with Euclidean spaces than do more general Hilbert spaces.

J. H. Manton and P.-O. Amblard. *A Primer on Reproducing Kernel Hilbert Spaces*. Foundations and Trends[®] in Signal Processing, vol. 8, no. 1-2, pp. 1–126, 2014.
DOI: 10.1561/20000000050.

1

Introduction

Hilbert space theory is a prime example in mathematics of a beautiful synergy between symbolic manipulation and visual reasoning. Two-dimensional and three-dimensional pictures can be used to reason about infinite-dimensional Hilbert spaces, with symbolic manipulations subsequently verifying the soundness of this reasoning, or suggesting modifications and refinements. Visualising a problem is especially beneficial because over half the human brain is involved to some extent with visual processing. Hilbert space theory is an invaluable tool in numerous signal processing and systems theory applications [61, 11, 9].

Hilbert spaces satisfying certain additional properties are known as Reproducing Kernel Hilbert Spaces (RKHSs), and RKHS theory is normally described as a transform theory between Reproducing Kernel Hilbert Spaces and positive semi-definite functions, called kernels: every RKHS has a unique kernel, and certain problems posed in RKHSs are more easily solved by involving the kernel. However, this description hides the crucial aspect that the kernel captures not just intrinsic properties of the Hilbert space but also how the Hilbert space is embedded in a larger function space, which is referred to here as its extrinsic geometry. A novel feature of this primer is drawing attention to this

extrinsic geometry, and using it to explain why certain problems can be solved more efficiently in terms of the kernel than the space itself.

Another novel feature of this primer is that it motivates and develops RKHS theory in finite dimensions before considering infinite dimensions. RKHS theory is ingenious; the underlying definitions are simple but powerful and broadly applicable. These aspects are best brought out in the finite-dimensional case, free from the distraction of infinite-dimensional technicalities. Essentially all of the finite-dimensional results carry over to the infinite-dimensional setting.

This primer ultimately aims to empower readers to recognise when and how RKHS theory can profit them in their own work. The following are three of the known uses of RKHS theory.

1. If a problem involves a subspace of a function space, and if the subspace (or its completion) is a RKHS, then the additional properties enjoyed by RKHSs may help solve the problem. (Explicitly computing limits of sequences in Hilbert spaces can be difficult, but in a RKHS the limit can be found pointwise.)
2. Certain classes of problems involving positive semi-definite functions can be solved by introducing an associated RKHS whose kernel is precisely the positive semi-definite function of interest. A classic example, due to Parzen, is associating a RKHS with a stochastic process, where the kernel of the RKHS is the covariance function of the stochastic process (see §7.2).
3. Given a set of points and a function specifying the desired distances between points, the points can be embedded in a RKHS with the distances between points being precisely as prescribed; see §5. (Support vector machines use this to convert certain non-linear problems into linear problems.)

In several contexts, RKHS methods have been described as providing a unified framework [76, 31, 45, 58]; although a subclass of problems was solved earlier by other techniques, a RKHS approach was found to be more elegant, have broader applicability, or offer new insight for obtaining actual solutions, either in closed form or numerically. Parzen

describes RKHS theory as facilitating a coordinate-free approach [45]. While the underlying Hilbert space certainly allows for coordinate-free expressions, the power of a RKHS beyond that of a Hilbert space is the presence of two coordinate systems: the pointwise coordinate system coming from the RKHS being a function space, and a canonical (but overdetermined) coordinate system coming from the kernel. The pointwise coordinate system facilitates taking limits while a number of geometric problems have solutions conveniently expressed in terms of what we define to be the canonical coordinate system. (Geometers may wish to think of a RKHS as a subspace $V \subset \mathbb{R}^X$ with pointwise coordinates being the extrinsic coordinates coming from $\mathbb{R}^X$ while the canonical coordinates are intrinsic coordinates on V relating directly to the inner product structure on V .)

The body of the primer elaborates on all of the points mentioned above and provides simple but illuminating examples to ruminate on. Parenthetical remarks are used to provide greater technical detail that some readers may welcome. They may be ignored without compromising the cohesion of the primer. Proofs are there for those wishing to gain experience at working with RKHSs; simple proofs are preferred to short, clever, but otherwise uninformative proofs. Italicised comments appearing in proofs provide intuition or orientation or both.

This primer is neither a review nor a historical survey, and as such, many classic works have not been discussed, including those by leading pioneers such as Wahba [71, 72].

Contributions This primer is effectively in two parts. The first part (§1–§7), written by the first author, gives a gentle and novel introduction to RKHS theory. It also presents several classical applications. The second part (§8–§9), with §8 written jointly and §9 written by the second author, focuses on recent developments in the machine learning literature concerning embeddings of random variables.

1.1 Assumed Knowledge

Basic familiarity with concepts from finite-dimensional linear algebra is assumed: vector space, norm, inner product, linear independence,

basis, orthonormal basis, matrix manipulations and so forth.

Given an inner product $\langle \cdot, \cdot \rangle$, the induced norm is $\|x\| = \sqrt{\langle x, x \rangle}$. Not every norm comes from an inner product, meaning some norms cannot be written in this form. If a norm does come from an inner product, the inner product can be uniquely determined from the norm by the polarisation identity $4\langle x, y \rangle = \|x + y\|^2 - \|x - y\|^2$. (A corresponding formula exists for complex-valued vector spaces.)

A metric $d(\cdot, \cdot)$ is a “distance function” describing the distance between two points in a metric space. To be a valid metric, it must satisfy several axioms, including the triangle inequality. A normed space is automatically a metric space by the correspondence $d(x, y) = \|x - y\|$.

1.2 Extrinsic Geometry and a Motivating Example

Differential geometry groups geometric properties into two kinds: intrinsic and extrinsic. Intrinsic properties depend only on the space itself, while extrinsic properties depend on precisely how the space is embedded in a larger space. A simple example in linear algebra is that the orientation of a straight line passing through the origin in $\mathbb{R}^2$ describes the *extrinsic* geometry of the line.

The following observation helps motivate the development of finite-dimensional RKHS theory in §2. Let

$$L(\theta) = \{(t \cos \theta, t \sin \theta) \mid t \in \mathbb{R}\} \subset \mathbb{R}^2 \quad (1.1)$$

denote a straight line in $\mathbb{R}^2$ passing through the origin and intersecting the horizontal axis at an angle of θ radians; it is a one-dimensional subspace of $\mathbb{R}^2$. Fix an arbitrary point $p = (p_1, p_2) \in \mathbb{R}^2$ and define $f(\theta)$ to be the point on $L(\theta)$ closest to p with respect to the Euclidean metric. It can be shown that

$$f(\theta) = (r(\theta) \cos \theta, r(\theta) \sin \theta), \quad r(\theta) = p_1 \cos \theta + p_2 \sin \theta. \quad (1.2)$$

Visualising $f(\theta)$ as the projection of p onto $L(\theta)$ shows that $f(\theta)$ depends *continuously* on the orientation of the line. While (1.2) verifies this continuous dependence, it resorted to introducing an *ad hoc* parametrisation θ , and different values of θ (e.g., θ , $\pi + \theta$ and $2\pi + \theta$) can describe the same line.

Is there a more natural way of representing $L(\theta)$ and $f(\theta)$, using linear algebra?

A first attempt might involve using an orthonormal basis vector to represent $L(\theta)$. However, there is no *continuous* map from the line $L(\theta)$ to an orthonormal basis vector $v(\theta) \in L(\theta)$. (This should be self-evident with some thought, and follows rigorously from the Borsuk-Ulam theorem.) Note that $\theta \mapsto (\cos \theta, \sin \theta)$ is not a well-defined map from $L(\theta)$ to $\mathbb{R}^2$ because $L(0)$ and $L(\pi)$ represent the same line yet $(\cos 0, \sin 0) \neq (\cos \pi, \sin \pi)$.

RKHS theory uses not one but two vectors to represent $L(\theta)$. Specifically, it turns out that the kernel of $L(\theta)$, in matrix form, is

$$K(\theta) = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix}. \quad (1.3)$$

The columns of $K(\theta)$ are

$$k_1(\theta) = \cos \theta \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad k_2(\theta) = \sin \theta \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}. \quad (1.4)$$

Note that $L(\theta)$ is spanned by $k_1(\theta)$ and $k_2(\theta)$, and moreover, both k_1 and k_2 are well-defined (and continuous) functions of L ; if $L(\theta) = L(\phi)$ then $k_1(\theta) = k_1(\phi)$ and $k_2(\theta) = k_2(\phi)$. To emphasise, although θ is used here for convenience to describe the construction, RKHS theory defines a map from L to k_1 and k_2 that does not depend on any *ad hoc* choice of parametrisation. It is valid to write $k_1(L)$ and $k_2(L)$ to show they are functions of L alone.

Interestingly, f has a simple representation in terms of the kernel:

$$f(L) = p_1 k_1(L) + p_2 k_2(L). \quad (1.5)$$

Compared with (1.2), this is both simple and natural, and does not depend on any *ad hoc* parametrisation θ of the line L . In summary,

- the kernel represents a vector subspace by a possibly overdetermined (i.e., linearly dependent) ordered set of vectors, and the correspondence from a subspace to this ordered set is continuous;

- this *continuous* correspondence cannot be achieved with an ordered set of basis (i.e., linearly independent) vectors;
- certain problems have solutions that depend continuously on the subspace and can be written elegantly in terms of the kernel:

$$\text{subspace} \rightarrow \text{kernel} \rightarrow \text{solution}. \quad (1.6)$$

The above will be described in greater detail in §2.

Remark The above example was chosen for its simplicity. Ironically, the general problem of projecting a point onto a subspace is not well-suited to the RKHS framework for several reasons, including that RKHS theory assumes there is a norm only on the subspace; if there is a norm on the larger space in which the subspace sits then it is ignored. A more typical optimisation problem benefitting from RKHS theory is finding the minimum-norm function passing through a finite number of given points; minimising the norm acts to regularise this interpolation problem; see §6.1.

1.3 Pointwise Coordinates and Canonical Coordinates

Aimed at readers already familiar with Hilbert space theory, this section motivates and defines two coordinate systems on a RKHS.

A separable Hilbert space $\mathcal{H}$ possesses an orthonormal basis $e_1, e_2, \dots \in \mathcal{H}$. An arbitrary element $v \in \mathcal{H}$ can be expressed as an infinite series $v = \sum_{i=0}^{\infty} \alpha_i e_i$ where the “coordinates” α_i are given by $\alpha_i = \langle v, e_i \rangle$. A classic example is using a Fourier series to represent a periodic function. The utility of such a construction is that an arbitrary element of $\mathcal{H}$ can be written as the limit of a linear combination of a manageable set of fixed vectors.

RKHS theory generalises this ability of writing an arbitrary element of a Hilbert space as the limit of a linear combination of a manageable set of fixed vectors. If $\mathcal{H} \subset \mathbb{R}^T$ is a (not necessarily separable) RKHS then an arbitrary element $v \in \mathcal{H}$ can be expressed as the limit of a sequence $v_1, v_2, \dots \in \mathcal{H}$ of vectors, each of which is a finite linear combination of the vectors $\{K(\cdot, t) \mid t \in T\}$, where $K: T \times T \rightarrow \mathbb{R}$ is

the kernel of $\mathcal{H}$. It is this ability to represent an arbitrary element of a RKHS $\mathcal{H}$ as the limit of a linear combination of the $K(\cdot, t)$ that, for brevity, we refer to as the presence of a canonical coordinate system. The utility of this canonical coordinate system was hinted at in §1.2.

There is another natural coordinate system: since an element v of a RKHS $\mathcal{H} \subset \mathbb{R}^T$ is a function from T to $\mathbb{R}$, its t th coordinate can be thought of as $v(t)$. The relationship between this pointwise coordinate system and the aforementioned canonical coordinates is that $v(t) = \langle v, K(\cdot, t) \rangle$. Note though that whereas an arbitrary linear combination of the $K(\cdot, t)$ is guaranteed to be an element of $\mathcal{H}$, assigning values arbitrarily to the $v(t)$, i.e., writing down an arbitrary function v , may not yield an element of $\mathcal{H}$; canonical coordinates are intrinsic whereas pointwise coordinates are extrinsic. The utility of the pointwise coordinate system is that limits in a RKHS can be determined pointwise: if v_k is a Cauchy sequence, implying there exists a v satisfying $\|v_k - v\| \rightarrow 0$, then v is fully determined by $v(t) = \lim_k v_k(t)$.

2

Finite-dimensional RKHSs

Pedagogic material on RKHSs generally considers infinite-dimensional spaces from the outset¹. Starting with finite-dimensional spaces though is advantageous because the remarkable aspects of RKHS theory are already evident in finite dimensions, where they are clearer to see and easier to study. The infinite-dimensional theory is a conceptually straightforward extension of the finite-dimensional theory.

Although elements of a RKHS must be functions, there is a canonical correspondence between $\mathbb{R}^n$ and real-valued functions on $\{1, 2, \dots, n\}$ given by the rule that $(x_1, \dots, x_n) \in \mathbb{R}^n$ is equivalent to the function $f: \{1, 2, \dots, n\} \rightarrow \mathbb{R}$ satisfying $f(i) = x_i$ for $i = 1, \dots, n$. To exemplify, the point $(5, 8, 4) \in \mathbb{R}^3$ is canonically equivalent to the function $f: \{1, 2, 3\} \rightarrow \mathbb{R}$ given by $f(1) = 5$, $f(2) = 8$ and $f(3) = 4$. For simplicity and clarity, this primer initially works with $\mathbb{R}^n$.

Remark The term finite-dimensional RKHS is potentially ambiguous; is merely the RKHS itself finite-dimensional, or must the embedding space also be finite dimensional? For convenience, we adopt the latter

¹The tutorial [65] considers finite-dimensional spaces but differs markedly from the presentation here. An aim of [65] is explaining the types of regularisation achievable by minimising a RKHS norm.

interpretation. While this interpretation is not standard, it is consistent with our emphasis on the embedding space playing a significant role in RKHS theory. Precisely, we say a RKHS is finite dimensional if the set X in §4 has finite cardinality.

2.1 The Kernel of an Inner Product Subspace

Central to RKHS theory is the following question. Let $V \subset \mathbb{R}^n$ be endowed with an inner product. How can this configuration be described efficiently? The configuration involves three aspects: the vector space V , the orientation of V in $\mathbb{R}^n$, and the inner product on V . Importantly, the inner product is not defined on the whole of $\mathbb{R}^n$, unless $V = \mathbb{R}^n$. (An alternative viewpoint that will emerge later is that RKHS theory studies possibly degenerate inner products on $\mathbb{R}^n$, where V represents the largest subspace on which the inner product is not degenerate.)

One way of describing the configuration is by writing down a basis $\{v_1, \dots, v_r\} \subset V \subset \mathbb{R}^n$ for V and the corresponding Gram matrix G whose ij th element is $G_{ij} = \langle v_i, v_j \rangle$. Alternatively, the configuration is completely described by giving an orthonormal basis $\{u_1, \dots, u_r\} \subset V \subset \mathbb{R}^n$; the Gram matrix associated with an orthonormal basis is the identity matrix and does not need to be stated explicitly. However, these representations do not satisfy the following requirements because there is no unique choice for a basis, even an orthonormal basis.

One-to-one The relationship between a subspace V of $\mathbb{R}^n$ and its representation should be one-to-one.

Respect Topology If $(V_1, \langle \cdot, \cdot \rangle_1), (V_2, \langle \cdot, \cdot \rangle_2), \dots$ is a sequence of inner product spaces “converging” to $(V, \langle \cdot, \cdot \rangle)$ then the representations of $(V_1, \langle \cdot, \cdot \rangle_1), (V_2, \langle \cdot, \cdot \rangle_2), \dots$ should “converge” to the representation of $(V, \langle \cdot, \cdot \rangle)$, and *vice versa*.

Straightforward Inverse It should be easy to deduce V and its inner product from its representation.

The lack of a canonical basis for V can be overcome by considering spanning sets instead. Whereas a basis induces a coordinate system, a

spanning set that is not a basis induces an overdetermined coordinate system due to linear dependence.

RKHS theory produces a unique ordered spanning set $k_1, \dots, k_n$ for $V \subset \mathbb{R}^n$ by the rule that k_i is the unique vector in V satisfying $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$. Here and throughout, e_i denotes the vector whose elements are zero except for the i th which is unity; its dimension should always be clear from the context. In words, taking the inner product with k_i extracts the i th element of a vector. This representation looks *ad hoc* yet has proven to be remarkably useful, in part because it unites three different perspectives given by the three distinct but equivalent definitions below.

Definition 2.1. Let $V \subset \mathbb{R}^n$ be an inner product space. The *kernel* of V is the unique matrix $K = [k_1 \ k_2 \ \dots \ k_n] \in \mathbb{R}^{n \times n}$ determined by any of the following three equivalent definitions.

1. K is such that each k_i is in V and $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$.
2. $K = u_1 u_1^\top + \dots + u_r u_r^\top$ where $u_1, \dots, u_r$ is an orthonormal basis for V .
3. K is such that the k_i span V and $\langle k_j, k_i \rangle = K_{ij}$.

The third definition is remarkable despite following easily from the first definition because it shows that $K \in \mathbb{R}^{n \times n}$ is the Gram matrix corresponding to the vectors $k_1, \dots, k_n$. The kernel simultaneously encodes a set of vectors that span V and the corresponding Gram matrix for those vectors!

The name *reproducing kernel* can be attributed either to the kernel K reproducing itself in that its ij th element K_{ij} is the inner product of its j th and i th columns $\langle k_j, k_i \rangle$, or to the vectors k_i reproducing an arbitrary vector v by virtue of $v = (\langle v, k_1 \rangle, \dots, \langle v, k_n \rangle)$.

Example 2.1. Equip $V = \mathbb{R}^n$ with an inner product $\langle u, v \rangle = v^\top Q u$ where Q is symmetric (and positive definite). The equation $\langle v, k_i \rangle = e_i^\top v$ implies $k_i = Q^{-1} e_i$, that is, $K = Q^{-1}$. Alternatively, an eigendecomposition $Q = X D X^\top$ yields an orthonormal basis for V given by the scaled columns of X , namely, $\{X D^{-\frac{1}{2}} e_1, \dots, X D^{-\frac{1}{2}} e_n\}$. Therefore

$K = \sum_{i=1}^n (XD^{-\frac{1}{2}}e_i)(XD^{-\frac{1}{2}}e_i)^\top = XD^{-1}X^\top = Q^{-1}$. Observe that $\langle k_j, k_i \rangle = \langle Q^{-1}e_j, Q^{-1}e_i \rangle = e_i^\top Q^{-1}e_j = e_i^\top K e_j = K_{ij}$.

Example 2.2. Let $V \subset \mathbb{R}^2$ be the subspace spanned by the vector $(1, 1)$. Endow V with the unique inner product for which the vector $(1, 1)$ has unit norm. Then $\{(1, 1)\}$ is an orthonormal basis for V and therefore $K = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Example 2.3. The configuration having $K = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ as its kernel can be found as follows. Since K is two-by-two, V must sit inside $\mathbb{R}^2$. Since V is the span of $k_1 = [1 \ 0]^\top$ and $k_2 = [0 \ 0]^\top$, $V = \mathbb{R} \times \{0\}$. The vector k_1 has unit norm in V because $\langle k_1, k_1 \rangle = K_{11} = 1$.

Before proving the three definitions of K are equivalent, several remarks are made. Existence and uniqueness of K follows most easily from definition one because the defining equations are linear. From definition one alone it would seem remarkable that K is always positive semi-definite, denoted $K \geq 0$, yet this fact follows immediately from definition two. Recall that a positive semi-definite matrix is a symmetric matrix whose eigenvalues are greater than or equal to zero. That K is unique is not clear from definition two alone. Definition three gives perhaps the most important characterisation of K , yet from definition three alone it is not at all obvious that such a K exists. Definition three also implies that any element of V can be written as a linear combination $K\alpha$ of the columns of K , and $\langle K\alpha, K\beta \rangle = \beta^\top K\alpha$. The reader is invited to verify this by writing $K\beta$ as $\beta_1 k_1 + \cdots + \beta_n k_n$.

Lemma 2.1. Given an inner product space $V \subset \mathbb{R}^n$, there is precisely one $K = [k_1, \dots, k_n] \in \mathbb{R}^{n \times n}$ for which each k_i is in V and satisfies $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$.

Proof. This is an exercise in abstract linear algebra that is the generalisation of the linear equation $Ax = b$ having a unique solution if A is square and non-singular. Let $v_1, \dots, v_r$ be a basis for V . Linearity implies the constraints on k_i are equivalent to requiring $\langle v_j, k_i \rangle = e_i^\top v_j$ for $j = 1, \dots, r$. Let L denote the linear operator taking $k \in V$ to

$(\langle v_1, k \rangle, \langle v_2, k \rangle, \dots, \langle v_r, k \rangle)$. Both the domain V and the range $\mathbb{R}^r$ of L have dimension r . The linear operator L is equivalent to a square matrix. Now, L is injective because if $L(k) = L(\tilde{k})$ then $L(k - \tilde{k}) = 0$, that is, $\langle v, k - \tilde{k} \rangle = 0$ for all $v \in V$ (because $v_1, \dots, v_r$ is a basis), and in particular, $\langle k - \tilde{k}, k - \tilde{k} \rangle = 0$, implying $k = \tilde{k}$. The linear operator L is non-singular. Therefore, L is also surjective, and in particular, $L(k_i) = (e_i^\top v_1, \dots, e_i^\top v_r)$ has one and only one solution. $\square$

Definition three gives a non-linear characterisation of K requiring k_i to extract the i th element of the vectors k_j whereas definition one requires k_i to extract the i th element of any vector v . However, definition three also requires the k_j to span V , therefore, by writing an arbitrary vector v as a linear combination of the v_j , it becomes evident that definition three satisfies definition one.

Lemma 2.2. If K is such that the k_i span V and $\langle k_j, k_i \rangle = K_{ij}$ then $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$.

Proof. Fix a $v \in V$. Since v is in V and the k_i span V , there is a vector α such that $v = K\alpha$. Also, $k_i = Ke_i$. Therefore, $\langle v, k_i \rangle = \langle K\alpha, Ke_i \rangle = e_i^\top K\alpha = e_i^\top v$, as required. (Recall from earlier that $\langle k_j, k_i \rangle = K_{ij}$ implies $\langle K\alpha, K\beta \rangle = \beta^\top K\alpha$.) $\square$

The existence of a K satisfying definition three is implied by the existence of a K satisfying definition one.

Lemma 2.3. If K is such that each k_i is in V and $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$ then the k_i span V and $\langle k_j, k_i \rangle = K_{ij}$.

Proof. That $\langle k_j, k_i \rangle = K_{ij}$ follows immediately from K_{ij} being the i th element of k_j , namely, $e_i^\top k_j$. If the k_i do not span V then there is a non-zero $k \in V$ which is orthogonal to each and every k_i , that is, $\langle k, k_i \rangle = 0$. Yet this implies $e_i^\top k = 0$ for all i , that is, $k = 0$, a contradiction. $\square$

If the columns of $U \in \mathbb{R}^{n \times r}$ form an orthonormal basis for $V \subset \mathbb{R}^n$ then an arbitrary vector in V can be written as $U\alpha$, and moreover, $\langle U\alpha, U\beta \rangle = \beta^\top \alpha$. Referring to definition two, if the u_i are the columns of U , then $u_1 u_1^\top + \dots + u_r u_r^\top = UU^\top$. Armed with these facts, showing definition two satisfies definition one becomes straightforward.

Lemma 2.4. If $u_1, \dots, u_r$ is an orthonormal basis for V then $K = u_1 u_1^\top + \dots + u_r u_r^\top$ satisfies $\langle v, k_i \rangle = e_i^\top v$ for all $v \in V$.

Proof. Let $U = [u_1, \dots, u_r]$, so that $K = UU^\top$ and $k_i = K e_i = UU^\top e_i$. An arbitrary $v \in V$ can be represented as $v = U\alpha$. Therefore $\langle v, k_i \rangle = \langle U\alpha, UU^\top e_i \rangle = e_i^\top U\alpha = e_i^\top v$, as required. $\square$

Since definition one produces a unique K , and definitions two and three both produce a K satisfying definition one, the equivalence of the three definitions has been proven.

The next pleasant surprise is that for every positive semi-definite matrix $K \in \mathbb{R}^{n \times n}$ there is an inner product space $V \subset \mathbb{R}^n$ whose kernel is K . The proof relies on the following.

Lemma 2.5. If $K \in \mathbb{R}^{n \times n}$ is positive semi-definite then $\alpha^\top K \alpha = 0$ implies $K\alpha = 0$, where α is a vector in $\mathbb{R}^n$.

Proof. Eigendecompose K as $K = UDU^\top$. Let $\beta = U^\top \alpha$. Then $\alpha^\top K \alpha = 0$ implies $\beta^\top D \beta = 0$. Since $D \geq 0$, $D\beta$ must be zero. Therefore $K\alpha = UDU^\top \alpha = UD\beta = 0$. $\square$

Lemma 2.6. Let $V = \text{span}\{k_1, \dots, k_n\}$ be the space spanned by the columns $k_1, \dots, k_n$ of a positive semi-definite matrix $K \in \mathbb{R}^{n \times n}$. There exists an inner product on V satisfying $\langle k_j, k_i \rangle = K_{ij}$.

Proof. Let $u, v \in V$. Then there exist vectors α, β such that $u = K\alpha$ and $v = K\beta$. Define $\langle u, v \rangle$ to be $\beta^\top K \alpha$. To show this is well-defined, let $u = K\tilde{\alpha}$ and $v = K\tilde{\beta}$ be possibly different representations. Then $\beta^\top K \alpha - \tilde{\beta}^\top K \tilde{\alpha} = (\beta - \tilde{\beta})^\top K \alpha + \tilde{\beta}^\top K (\alpha - \tilde{\alpha}) = 0$ because $K\alpha = K\tilde{\alpha}$, $K\beta = K\tilde{\beta}$ and $K = K^\top$. Clearly $\langle \cdot, \cdot \rangle$ so defined is bilinear. To prove $\langle \cdot, \cdot \rangle$ is positive definite, assume $\langle K\alpha, K\alpha \rangle = \alpha^\top K \alpha = 0$. By Lemma 2.5 this implies $K\alpha = 0$, as required. $\square$

Given K , there is a *unique* inner product space whose kernel is K .

Lemma 2.7. Let $V_1 \subset \mathbb{R}^n$ and $V_2 \subset \mathbb{R}^n$ be two inner product spaces having the same kernel K . Then V_1 and V_2 are identical spaces: $V_1 = V_2$ and their inner products are the same.

Proof. The columns of K span both V_1 and V_2 , hence $V_1 = V_2$. For the same reason, the inner products on V_1 and V_2 are uniquely determined from the Gram matrix K corresponding to $k_1, \dots, k_n$. Since V_1 and V_2 have the same Gram matrix, their inner products are identical too. (Indeed, the inner product must be given by $\langle K\alpha, K\beta \rangle = \beta^\top K\alpha$.) $\square$

To summarise, for a fixed n , there is a bijective correspondence between inner product spaces $V \subset \mathbb{R}^n$ and positive semi-definite matrices $K \in \mathbb{R}^{n \times n}$. Given $V \subset \mathbb{R}^n$ there is precisely one kernel $K \geq 0$ (Definition 2.1 and Lemma 2.1). Given a $K \geq 0$, there is precisely one inner product space $V \subset \mathbb{R}^n$ for which K is its kernel (Lemmata 2.6 and 2.7). Recalling the criteria listed earlier, this correspondence is One-to-one and has a Straightforward Inverse. That it Respects Topology is discussed next.

2.2 Sequences of Inner Product Spaces

A sequence of one-dimensional vector spaces in $\mathbb{R}^2$ can be visualised as a collection of lines passing through the origin and numbered $1, 2, \dots$. It can be recognised visually when such a sequence converges. (This corresponds to the topology of the Grassmann manifold.) The kernel representation of an inner product space induces a concept of convergence for the broader situation of a sequence of *inner product* subspaces, possibly of *differing dimensions*, by declaring that the limit $V_\infty \subset \mathbb{R}^n$ of a sequence of inner product spaces $V_1, V_2, \dots \subset \mathbb{R}^n$ is the space whose kernel is $K_\infty = \lim_{n \rightarrow \infty} K_n$, if the limit exists. This makes the kernel representation of a subspace Respect Topology. The pertinent question is whether the induced topology is sensible in practice.

Example 2.4. Define on $\mathbb{R}^2$ a sequence of inner products $\langle u, v \rangle_n = v^\top Q_n u$ where $Q_n = \text{diag}\{1, n^2\}$ is a diagonal matrix with entries 1 and n^2 . An orthonormal basis for the n th space is $\{(1, 0), (0, n^{-1})\}$. It seems acceptable, at least in certain situations, to agree that the limit of this sequence of two-dimensional spaces is the one-dimensional subspace of $\mathbb{R}^2$ spanned by $(1, 0)$. This accords with defining convergence via kernels. Let K_n be the kernel of the n th space: $K_n = \text{diag}\{1, n^{-2}\}$.

Then $K_\infty = \lim_{n \rightarrow \infty} K_n = \text{diag}\{1, 0\}$. The space V_∞ having kernel K_∞ is the subspace of $\mathbb{R}^2$ spanned by $(1, 0)$, where the inner product is such that $(1, 0)$ has unit norm.

The example above illustrates a general principle: since K can be defined in terms of an orthonormal basis as $K = u_1 u_1^\top + \cdots + u_r u_r^\top$, a sequence of higher-dimensional inner product spaces can converge to a lower-dimensional inner product space if one or more of the orthonormal basis vectors approaches the zero vector. This need not be the only sensible topology, but it is a topology that finds practical use, as later verified by examples.

Remark Recall from Example 2.1 that if $V = \mathbb{R}^n$ then $K = Q^{-1}$. If Q_n is a sequence of positive definite matrices becoming degenerate in the limit, meaning one or more eigenvalues goes to infinity, then Q_n does not have a limit. However, the corresponding eigenvalues of $K_n = Q_n^{-1}$ go to zero, hence the kernel K_n may have a well-defined limit. In this sense, RKHS theory encompasses degenerate inner products. A way of visualising this is presented in §2.4.

2.3 Extrinsic Geometry and Interpolation

Finite-dimensional RKHS theory studies subspaces V of $\mathbb{R}^n$ rather than abstract vector spaces. If it is only known that Z is an abstract vector space then there is no way of knowing what the elements of Z look like. The best that can be done is assert the existence of a basis $\{b_1, \dots, b_r\}$. By comparison, knowing $V \subset \mathbb{R}^n$ allows working with V using *extrinsic coordinates* by writing an element of V as a vector in $\mathbb{R}^n$.

Writing $V \subset \mathbb{R}^n$ may also signify the importance of the orientation of V inside $\mathbb{R}^n$. If V and W are r -dimensional linear subspaces of $\mathbb{R}^n$ then their *intrinsic geometry* is the same but their *extrinsic geometry* may differ; V and W are equivalent (precisely, isomorphic) as vector spaces but they lie inside $\mathbb{R}^n$ differently unless $V = W$.

The usefulness of extrinsic geometry is exemplified by considering the interpolation problem of finding a vector $x \in V \subset \mathbb{R}^n$ of smallest norm and some of whose coordinates $e_i^\top x$ are specified. This problem is posed using extrinsic geometry. It can be rewritten intrinsically, with-

out reference to $\mathbb{R}^n$, by using the linear operator $L_i: V \rightarrow \mathbb{R}$ defined to be the restriction of $x \mapsto e_i^\top x$. However, the extrinsic formulation is the more amenable to a unified approach.

Example 2.5. Endow $V \subset \mathbb{R}^n$ with an inner product. Fix an i and consider how to find $x \in V$ satisfying $e_i^\top x = 1$ and having the smallest norm. Geometrically, such an x must be orthogonal to any vector $v \in V$ satisfying $e_i^\top v = 0$, for otherwise its norm could be decreased. This approach allows x to be found by solving a system of linear equations, however, going further seems to require choosing a basis for V .

RKHS theory replaces an *ad hoc* choice of basis for $V \subset \mathbb{R}^n$ by the particular choice $k_1, \dots, k_n$ of spanning vectors for V . Because the kernel representation Respects Topology, the vectors $k_1, \dots, k_n$ vary continuously with changing V . The solution to the interpolation problem should also vary continuously as V changes. There is thus a chance that the solution can be written elegantly in terms of $k_1, \dots, k_n$.

Example 2.6. Continuing the example above, let $K = [k_1, \dots, k_n]$ be the kernel of V . Let $x = K\alpha$ and $v = K\beta$. The constraints $e_i^\top x = 1$ and $\langle x, v \rangle = 0$ whenever $e_i^\top v = 0$ become $k_i^\top \alpha = 1$ and $\beta^\top K\alpha = 0$ whenever $\beta^\top K e_i = 0$. (Recall that i is fixed.) The latter requirement is satisfied by $\alpha = c e_i$ where $c \in \mathbb{R}$ is a scalar. Solving $k_i^\top \alpha = 1$ implies $c = K_{ii}^{-1}$. Therefore, $x = K_{ii}^{-1} k_i$. Note $\|x\|^2 = \langle x, x \rangle = K_{ii}^{-1}$.

In the above example, as V changes, both the kernel K and the solution x change. Yet the relationship between x and K remains constant: $x = K_{ii}^{-1} k_i$. This is further evidence that the representation K of the extrinsic geometry is a useful representation.

There is a geometric explanation for the columns of K solving the single-point interpolation problem. Let $L_i: V \rightarrow \mathbb{R}$ denote the i th coordinate function: $L_i(v) = e_i^\top v$. That $\langle v, k_i \rangle = L_i(v)$ means k_i is the gradient of L_i . In particular, the line determined by k_i meets the level set $\{v \mid L_i(v) = 1\}$ at right angles, showing that k_i meets the orthogonality conditions for optimality. Turning this around leads to yet another definition of K .

Lemma 2.8. The following is a geometric definition of the k_i that is equivalent to Definition 2.1. Let $H_i = \{z \in \mathbb{R}^n \mid e_i^\top z = 1\}$ be the hyperplane consisting of all vectors whose i th coordinate is unity. If $V \cap H_i$ is empty then define $k_i = 0$. Otherwise, let $\tilde{k}_i$ be the point in the intersection $V \cap H_i$ that is closest to the origin. Define k_i to be $k_i = c\tilde{k}_i$ where $c = \langle \tilde{k}_i, \tilde{k}_i \rangle^{-1}$.

Proof. Since K is unique (Lemma 2.1), it suffices to prove the k_i defined here satisfy definition one of Definition 2.1. If $V \cap H_i$ is empty then $e_i^\top v = 0$ for all $v \in V$ and $k_i = 0$ satisfies definition one. Assume then that $V \cap H_i$ is non-empty. It has a closest point, $\tilde{k}_i$, to the origin. As $e_i^\top \tilde{k}_i = 1$, $\tilde{k}_i$ is non-zero and c is well-defined. Also, $\langle w, \tilde{k}_i \rangle = 0$ whenever $w \in V$ satisfies $e_i^\top w = 0$, for otherwise $\tilde{k}_i$ would not have the smallest norm. Let $v \in V$ be arbitrary and define $w = v - a\tilde{k}_i$ where $a = e_i^\top v$. Then $e_i^\top w = e_i^\top v - a = 0$. Thus $\langle v, k_i \rangle = \langle w + a\tilde{k}_i, c\tilde{k}_i \rangle = ac\langle \tilde{k}_i, \tilde{k}_i \rangle = a = e_i^\top v$, as required. $\square$

The c in the lemma scales $\tilde{k}_i$ so that $\langle k_i, k_i \rangle = e_i^\top k_i$. It is also clear that if $k_i \neq 0$ then $K_{ii} \neq 0$, or in other words, if $K_{ii} = 0$ then the i th coordinate of every vector v in V is zero and the interpolation problem has no solution.

2.4 Visualising RKHSs

It is more expedient to understand RKHS theory by focusing on the columns k_i of the kernel K rather than on the matrix K itself; K being a positive semi-definite matrix corresponding to the Gram matrix of the k_i is a wonderful bonus. The indexed set of vectors $k_1, \dots, k_n$ changes in a desirable way as $V \subset \mathbb{R}^n$ changes. This can be visualised explicitly for low-dimensional examples.

The inner product on a two-dimensional vector space V can be depicted by drawing the ellipse $\{v \in V \mid \langle v, v \rangle = 1\}$ on V because the inner product is uniquely determined from the norm, and the ellipse uniquely determines the norm. Based on Lemma 2.8, the columns k_1 and k_2 of the kernel K corresponding to $V \subset \mathbb{R}^2$ can be found geometrically, as explained in the caption of Figure 2.1.

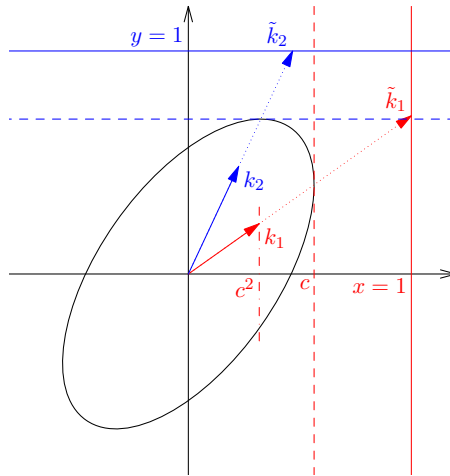


Figure 2.1: The ellipse comprises all points one unit from the origin. It determines the chosen inner product on $\mathbb{R}^2$. The vector $\tilde{k}_1$ is the closest point to the origin on the vertical line $x = 1$. It can be found by enlarging the ellipse until it first touches the line $x = 1$, or equivalently, as illustrated, it can be found by shifting the line $x = 1$ horizontally until it meets the ellipse tangentially, represented by the dashed vertical line, then travelling radially outwards from the point of intersection until reaching the line $x = 1$. The vector k_1 is a scaled version of $\tilde{k}_1$. If the dashed vertical line intersects the x -axis at c then $k_1 = c^2 \tilde{k}_1$. Equivalently, k_1 is such that its tip intersects the line $x = c^2$. The determination of k_2 is analogous but with respect to the horizontal line $y = 1$.

Figures 2.2 and 2.3 reveal how k_1 and k_2 vary as the inner product on $V = \mathbb{R}^2$ changes. Figure 2.3 shows that k_1 and k_2 vary smoothly as the inner product changes. By comparison, an orthonormal basis formed from the principal and minor axes of the ellipse must have a discontinuity somewhere, because after a 180 degree rotation, the ellipse returns to its original shape yet, for example, $\{e_1, \frac{1}{2}e_2\}$ rotated 180 degrees is $\{-e_1, -\frac{1}{2}e_2\}$, which differs from $\{e_1, \frac{1}{2}e_2\}$.

Figure 2.4 illustrates the kernel of various one-dimensional subspaces in $\mathbb{R}^2$. As V is one-dimensional, k_1 and k_2 must be linearly dependent since they span V . Together, k_1 and k_2 describe not only V and its inner product, but also the orientation of V in $\mathbb{R}^2$.

Figure 2.5 portrays how a sequence of two-dimensional subspaces can converge to a one-dimensional subspace. From the perspective of

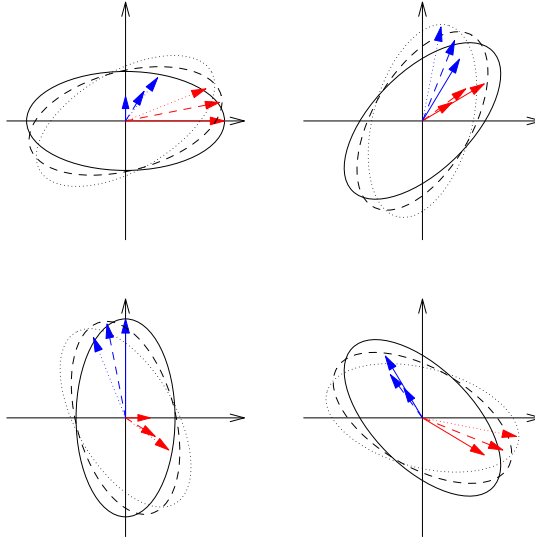


Figure 2.2: Shown are the vectors k_1 (red) and k_2 (blue) corresponding to rotated versions of the inner product $\langle u, v \rangle = v^\top Q u$ where $Q = \text{diag}\{1, 4\}$. The magnitude and angle of k_1 and k_2 are plotted in Figure 2.3.

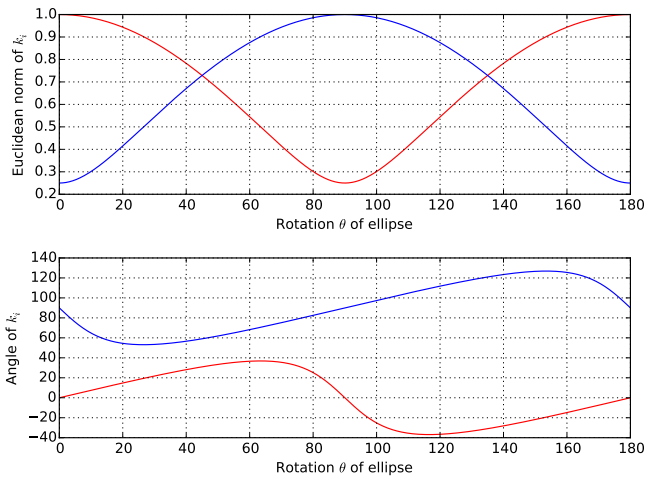


Figure 2.3: Plotted are the magnitude and angle of k_1 (red) and k_2 (blue) corresponding to rotated versions of the inner product $\langle u, v \rangle = v^\top Q u$ where $Q = \text{diag}\{1, 4\}$, as in Figure 2.2.

the inner product, the convergence is visualised by an ellipse degenerating to a line segment. From the perspective of the kernel $K = [k_1 \ k_2]$, the convergence is visualised by k_1 and k_2 individually converging to vectors lying in the same subspace.

The figures convey the message that the k_i are a spanning set for V that vary continuously with changing V . Precisely how they change is of more algebraic than geometric importance; the property $\langle k_j, k_i \rangle = K_{ij}$ is very convenient to work with algebraically.

2.5 RKHSs over the Complex Field

RKHS theory extends naturally to complex-valued vector spaces. In finite dimensions, this means considering subspaces $V \subset \mathbb{C}^n$ where V is equipped with a complex-valued inner product $\langle \cdot, \cdot \rangle: V \times V \rightarrow \mathbb{C}$. The only change to Definition 2.1 is replacing $K = u_1 u_1^\top + \cdots + u_r u_r^\top$ by $K = u_1 u_1^H + \cdots + u_r u_r^H$ where H denotes Hermitian transpose.

The kernel K will always be Hermitian ($K = K^H$) and positive semi-definite. Swapping the order of an inner product introduces a conjugation: $\langle u, v \rangle = \overline{\langle v, u \rangle}$. Beyond such minor details, the real-valued and complex-valued theories are essentially the same.

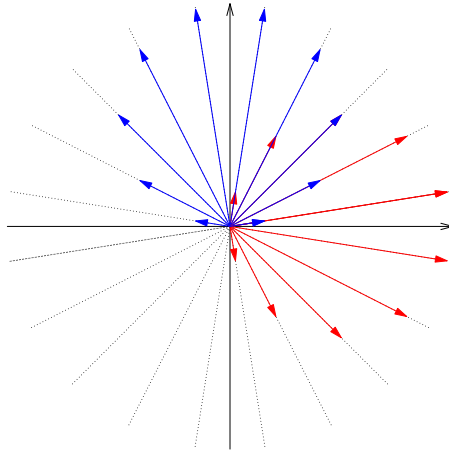


Figure 2.4: Shown are k_1 (red) and k_2 (blue) for various one-dimensional subspaces (black) of $\mathbb{R}^2$. In all cases, the inner product is the standard Euclidean inner product. Although not shown, k_2 is zero when V is horizontal, and k_1 is zero when V is vertical. Therefore, the magnitude of the red vectors increases from zero to a maximum then decreases back to zero. The same occurs for the blue vectors.

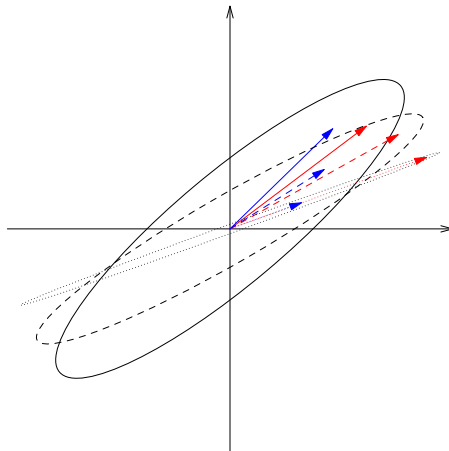


Figure 2.5: Illustration of how, as the ellipse gets narrower, the two-dimensional inner product space $V = \mathbb{R}^2$ converges to a one-dimensional inner product space. The kernels of the subspaces are represented by red (k_1) and blue (k_2) vectors.

3

Function Spaces

Typifying infinite-dimensional vector spaces are function spaces. Given an arbitrary set X , let $\mathbb{R}^X = \{f: X \rightarrow \mathbb{R}\}$ denote the set of all functions from X to $\mathbb{R}$. It is usually given a vector space structure whose simplicity belies its usefulness: vector-space operations are defined *pointwise*, meaning scalar multiplication $\alpha \cdot f$ sends the function f to the function g given by $g(x) = \alpha f(x)$, while vector addition sends $f + g$ to the function h defined by $h(x) = f(x) + g(x)$.

The space $\mathbb{R}^X$ is often too large to be useful on its own, but it contains useful subspaces, such as the spaces of all continuous, smooth or analytic functions. (This necessitates X having an appropriate structure; for example, it does not make sense for a function $f: X \rightarrow \mathbb{R}$ to be continuous unless X is a topological space.) Note that to be a subspace, as opposed to merely a subset, the restrictions placed on the functions must be closed with respect to vector-space operations: if f and g are in the subset and α is an arbitrary scalar then $f + g$ and αf must also be in the subset. While the set of all continuous functions $f: [0, 1] \rightarrow \mathbb{R}$ on the interval $[0, 1] \subset \mathbb{R}$ is a vector subspace of $\mathbb{R}^{[0,1]}$, the set of all functions $f: [0, 1] \rightarrow \mathbb{R}$ satisfying $f(1) \leq 1$ is not.

3.1 Function Approximation

Function approximation is an insightful example of how function spaces facilitate geometric reasoning for solving algebraic problems [40].

Example 3.1. Denote by $C_2[0, 1]$ the space of continuous functions $f: [0, 1] \rightarrow \mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x) dx$. The square of the induced norm is $\|f\|^2 = \langle f, f \rangle = \int_0^1 [f(x)]^2 dx$. Let V denote the subspace of $C_2[0, 1]$ spanned by the functions $g_0(x) = 1$, $g_1(x) = x$ and $g_2(x) = x^2$. In words, V is the space of all polynomials of degree at most two. Fix an f , say, $f(x) = e^x - 1$. Consider finding $g \in V$ that minimises $\|f - g\|^2$. Figure 3.1 visualises this in two ways. Graphing f does not help in finding g , whereas treating f and g as elements of a function space suggests g can be found by solving the linear equations $\langle f - g, v \rangle = 0$ for all $v \in V$.

An orthogonal basis for V is $p_0(x) = 1$, $p_1(x) = 2x - 1$ and $p_2(x) = 6x^2 - 6x + 1$. (These are the shifted Legendre polynomials.) Note $\|p_i\|^2 = (2i + 1)^{-1}$. Write g as $g = \alpha_0 p_0 + \alpha_1 p_1 + \alpha_2 p_2$. Attempting to solve $\langle f - g, v \rangle = 0$ for $v \in \{p_0, p_1, p_2\}$ leads to the equations $\langle \alpha_i p_i, p_i \rangle = \langle f, p_i \rangle$. Since $\langle f, p_0 \rangle = e - 2$, $\langle f, p_1 \rangle = 3 - e$ and $\langle f, p_2 \rangle = 7e - 19$, the coefficients α_i are found to be $\alpha_0 = e - 2$, $\alpha_1 = 9 - 3e$ and $\alpha_2 = 35e - 95$.

The simple optimisation problem in Example 3.1 could have been solved using calculus by setting to zero the derivatives of $\|f - \sum_i \alpha_i p_i\|^2$ with respect to the α_i . Importantly though, setting the derivatives to zero in general does not guarantee finding the optimal solution because the optimal solution need not exist¹. When applicable, vector space methods go beyond calculus in that they can guarantee the optimal solution has been found. (Calculus examines local properties whereas being a global optimum is a global property.)

Infinite-dimensional inner product spaces can exhibit behaviour not found in finite-dimensional spaces. The projection of f onto V shown in Figure 3.1 suggesting every minimum-norm problem has a solution is not necessarily true in infinite dimensions, as Example 3.2 will exemplify. Conditions must be imposed before the geometric picture in

¹See [69] for a historical account relating to existence of optimal solutions.

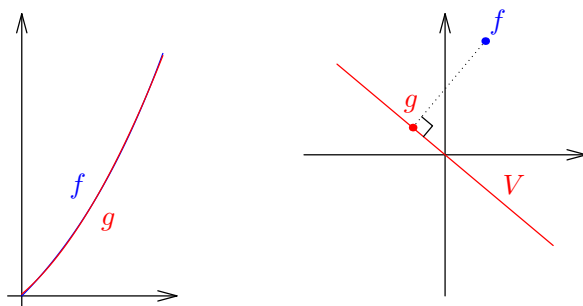


Figure 3.1: On the left is the direct way of visualising the function approximation problem: given a function f (blue), find a function g (red) that best approximates f , where g is constrained to a subclass of functions. Shown here is the second-order polynomial approximation of $f(x) = e^x - 1$ found in Example 3.1. It approximates f well, in that the graphs of f and g overlap. On the right is a geometric visualisation of the function approximation problem, made possible by representing functions not by their graphs (left) but as elements of a vector space (right). The subspace V is the subclass of functions in which g must lie. Minimising the square of the norm $\|f - g\|$ means finding the point on V that is closest to f . As the norm comes from an inner product, the point g must be such that $f - g$ is perpendicular to every function in V .

Figure 3.1 accurately describes the infinite-dimensional function approximation problem.

When coming to terms with infinite-dimensional spaces, one of the simplest (Hilbert) spaces to contemplate is l^2 , the space of square-summable sequences. Elements of l^2 are those sequences of real numbers for which the sum of the squares of each term of the sequence is finite. In symbols, $x = (x_1, x_2, \dots)$ is in l^2 if and only if $\sum_{i=1}^{\infty} x_i^2 < \infty$. For example, $(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots)$ is in l^2 . The inner product on l^2 is implicitly taken to be $\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i$. Note that the square of the induced norm is $\|x\|^2 = \sum_{i=1}^{\infty} x_i^2$ and hence the condition for x to be in l^2 is precisely the condition that it have finite l^2 -norm.

Approximating an element of l^2 by an element of a subspace $V \subset l^2$ can be considered a function approximation problem because there is a canonical correspondence between l^2 and a subclass of functions from $\{1, 2, \dots\}$ to $\mathbb{R}$, just as there is a canonical correspondence between R^n and functions from $\{1, \dots, n\}$ to $\mathbb{R}$. (The subclass comprises those functions f for which $\sum_{i=1}^{\infty} f(i)^2 < \infty$.)

Example 3.2. Let $V \subset l^2$ be the vector space of sequences with only a finite number of non-zero terms, thought of as a subspace of l^2 . There is no element g of V that minimises $\|f - g\|^2$ when $f = (1, \frac{1}{2}, \frac{1}{4}, \dots) \in l^2$. Given any approximation g of f , it can always be improved by changing a zero term in g to make it match the corresponding term in f . A rigorous proof goes as follows. Assume to the contrary there exists a $g \in V$ such that $\|f - \tilde{g}\| \geq \|f - g\|$ for all $\tilde{g} \in V$. Let i be the smallest integer such that $g_i = 0$. It exists because only a finite number of terms of g are non-zero. Let $\tilde{g}$ be a replica of g except for setting its i th term to $\tilde{g}_i = f_i$. Let e be the zero sequence except for the i th term which is unity. Then $f - g = (f - \tilde{g}) + f_i e$. Since $\langle f - \tilde{g}, e \rangle = 0$, $\|f - g\|^2 = \|f - \tilde{g}\|^2 + f_i^2 \|e\|^2$, implying $\|f - \tilde{g}\| < \|f - g\|$ and contradicting the existence of an optimal element g .

Reconciling the *algebra* in Example 3.2 proving the non-existence of an optimal approximation, with the *geometry* depicted in Figure 3.1 suggesting otherwise, requires understanding the interplay between vector space structures and topological structures induced by the norm.

Before delving into that topic, it is penetrating to ascertain whether RKHS theory may be beneficially applied to the function approximation problem. A novel feature of this primer is considering RKHS theory from the “dynamic” perspective of how K changes in response to changing V . The solution of the interpolation problem in Examples 2.5 and 2.6 is continuous in K , and in particular, a solution can be found even for a lower-dimensional $V \subset \mathbb{R}^n$ by working in $\mathbb{R}^n$ and taking limits as the inner product becomes degenerate, much like in Figure 2.5. The function approximation problem lacks this property because enlarging V to include f will mean the optimal solution is $g = f$ no matter how close the inner product is to being degenerate. Since the solution of the function approximation problem does not Respect Topology it is unlikely RKHS theory will be of assistance. (It is irrelevant that the space used in Example 3.1 is not a RKHS; the function approximation problem could have been posed equally well in a genuine RKHS.)

No theory is a panacea for all problems. RKHS theory has broad but not universal applicability.

3.2 Topological Aspects of Inner Product Spaces

Reproducing kernel Hilbert spaces have additional properties that more general Hilbert spaces and other inner product spaces do not necessarily enjoy. These properties were chosen to make RKHSs behave more like finite-dimensional spaces. Learning what can “go wrong” in infinite dimensions helps put RKHS theory in perspective.

Some subspaces should not be drawn as sharp subspaces, as in Figure 3.1, but as blurred subspaces denoting the existence of points not lying in the subspace yet no more than zero distance away from it. The blur represents an infinitesimally small region as measured by the norm, but need not be small in terms of the vector space structure. Every point in l^2 is contained in the blur around the V in Example 3.2 despite V being a very small subspace of l^2 in terms of cardinality of basis vectors. (A Hamel basis for l^2 is uncountable while a Hamel basis for V is countable. Note that in an infinite-dimensional Hilbert space, every Hamel basis is uncountable.)

Since the blur is there to represent the geometry coming from the norm, the blur should be drawn as an infinitesimally small region around V . An accurate way of depicting $V \subset l^2$ in Example 3.2 is by using a horizontal plane for V and adding a blur of infinitesimal height to capture the rest of l^2 . This visually implies correctly that there is no vector in l^2 that is orthogonal to every vector in V . Any $f \in l^2$ that is not in V makes an infinitesimally small angle with V , for otherwise the vectors $f, 2f, 3f, \dots$ would move further and further away from V .

As these facts are likely to generate more questions than answers, it is prudent to return to the beginning and work slowly towards deriving these facts from first principles.

An inner product induces a norm, and a norm induces a topology. It suffices here to understand a topology as a rule for determining which sequences converge to which points. A norm coming from an inner product determines the inner product uniquely. It is pragmatic to think of the norm as the more dominant structure. The presence of an inner product means the square of the norm is “quadratic”, a very convenient property.

Inner product spaces therefore have two structures: a vector space structure and a norm whose square is “quadratic”. The axioms of a normed space ensure these two structures are compatible. For example, the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$ involves both the norm and vector addition. Nevertheless, the two structures capture different aspects that are conflated in finite dimensions.

A norm induces a topology, and in particular, a norm determines what sequences have what limit points. Topology is a weaker concept than that of norm because different norms may give rise to the same topology. In fact, in *finite* dimensions, every norm gives rise to the same topology as every other norm: if $x_k \rightarrow x$ with respect to one norm $\|\cdot\|$, meaning $\|x_k - x\| \rightarrow 0$, then $x_k \rightarrow x$ with respect to any other norm.

The first difference then in infinite dimensions is that different norms can induce different topologies.

Example 3.3. Let $\|\cdot\|_2$ denote the standard norm on l^2 and let $\|x\|_\infty = \sup_i |x_i|$ be an alternative norm. Let $x^{(i)}$ denote the sequence $(\frac{1}{i}, \dots, \frac{1}{i}, 0, 0, \dots)$ whose first i^2 terms are $\frac{1}{i}$ and the rest zero. Then $x^{(1)}, x^{(2)}, \dots$ is a sequence of elements of l^2 . Since $\|x^{(i)}\|_\infty = \frac{1}{i} \rightarrow 0$, the sequence converges to the zero vector with respect to the norm $\|\cdot\|_\infty$. However, $\|x^{(i)}\|_2 = 1$ does not converge to zero and hence $x^{(i)}$ does not converge to the zero vector with respect to the norm $\|\cdot\|_2$.

A second difference is that an infinite dimensional subspace need not be topologically *closed*. A subspace V of a normed space W is closed in W if, for any convergent sequence $v_1, v_2, \dots \rightarrow w$, where the v_i are in V but w need only be in W , it is nevertheless the case that w is in V . Simply put, V is closed if every limit point of every sequence in V is also in V . The subspace V in Example 3.2 is not closed because the sequence $(1, 0, \dots), (1, \frac{1}{2}, 0, \dots), (1, \frac{1}{2}, \frac{1}{4}, 0, \dots)$ is in V but its limit $(1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots) \in l^2$ is not in V . A non-closed subspace should have an infinitesimal blur around it denoting the existence of elements not in the subspace but which are infinitesimally close to it, a consequence of $v_1, v_2, \dots \rightarrow w$ meaning $\|v_i - w\| \rightarrow 0$.

In finite dimensions, if $V \subset \mathbb{R}^n$ and $v_1, v_2, \dots$ is a sequence of elements of V converging to a point $v \in \mathbb{R}^n$, it is visually clear from a diagram that the limit point v must lie in V .

Lemma 3.1. Every finite-dimensional subspace $V \subset W$ of an inner product space W is closed.

Proof. Let $v_1, v_2, \dots \rightarrow w$ be a convergent sequence, with v_i in V and w in W . Let $u_1, \dots, u_r$ be an orthonormal basis for V . Let $\tilde{w} = w - \langle w, u_1 \rangle u_1 - \dots - \langle w, u_r \rangle u_r$. Here, $w - \tilde{w}$ is the projection of w onto V . In finite dimensions, the projection is always well-defined. Then $\tilde{w}$ is orthogonal to $u_1, \dots, u_r$. By Pythagoras' theorem, $\|\tilde{w}\| \leq \|w - v_i\|$. By definition, $v_i \rightarrow w$ implies $\|w - v_i\| \rightarrow 0$, therefore, $\|\tilde{w}\|$ must equal zero, implying w lies in V . $\square$

A key ingredient in the above proof is that a vector $w \notin V$ can be projected onto V and hence is a non-zero distance away from V . In Example 3.2 though, there is no orthogonal projection of f onto V .

Lemma 3.2. Define f and V as in Example 3.2. There is no vector $\tilde{f} \in V$ such that $\langle f - \tilde{f}, v \rangle = 0$ for all $v \in V$.

Proof. Let $\tilde{f} \in V$ be arbitrary. Let i be the smallest integer such that the i th term of $\tilde{f}$ is zero. Then the i th term of $f - \tilde{f}$ is non-zero and the inner product of $f - \tilde{f}$ with the element $(0, \dots, 0, 1, 0, \dots, 0) \in V$ having 1 for its i th term, is non-zero. $\square$

The space l^2 , as a vector space, is larger than it looks. The sequences $x^{(i)} = (0, \dots, 0, 1, 0, \dots, 0)$, with the unit appearing in the i th term, do not form a (Hamel) basis for l^2 . The $x^{(i)}$ span V in Example 3.2 but do not come close to spanning l^2 . The $x^{(i)}$ are countable in number whereas a basis for l^2 must necessarily be uncountable. This stems from the span of a set of vectors being the vectors representable by a *finite* linear combination of vectors in the set. Elements in l^2 having infinitely many non-zero terms cannot be written as finite linear combinations of the $x^{(i)}$. (Since infinite summations involve taking limits, only finite linear combinations are available in spaces with only a vector space structure. A prime mathematical reason for introducing a topology, or *a fortiori* a norm, is to permit the use of infinite summations.)

From the norm's perspective though, l^2 is not large: any point in l^2 is reachable as a limit of a sequence of points in the span of the $x^{(i)}$. Precisely, $(f_1, f_2, \dots) \in l^2$ is the limit of $(f_1, 0, \dots)$, $(f_1, f_2, 0, \dots)$, $\dots$.

This shows *the norm is coarser than the vector space structure*. The vector space structure in Example 3.2 places f away from V yet the norm places f no more than zero distance from V .

To demonstrate the phenomenon in Example 3.2 is not esoteric, the following is a more familiar manifestation of it. Once again, the chosen subspace V is not closed.

Example 3.4. Define $C_2[0, 1]$ as in Example 3.1, but this time take V to be the subspace of all polynomials on the unit interval $[0, 1]$. Then there is no polynomial $g \in V$ that is closest to $f(x) = \sin(x)$ because f can be approximated arbitrarily accurately on $[0, 1]$ by including more terms of the Taylor series approximation $\sin(x) = 1 - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$.

To summarise, it is sometimes possible to extend an inner product on an infinite dimensional space V to an inner product on $V \oplus \text{span}\{f\}$ so that there is a sequence in V converging to f . Here, f is a new vector being added to form a larger space. The affine spaces $cf + V$ for $c \in \mathbb{R}$ would then be infinitesimally close to each other, crammed one on top of the other. Similarly, l^2 is obtained from the V in Example 3.2 by adding basis vectors that are all crammed flat against V , hence the earlier suggestion of visualising l^2 as an infinitesimal blur around V .

3.3 Evaluation Functionals

Evaluation functionals play a crucial role in RKHS theory. In finite dimensions, they went unnoticed as the linear functionals $v \mapsto e_i^\top v$.

Let V be a subspace of $\mathbb{R}^X$ where X is an arbitrary set. An element f of V is therefore a function $f: X \rightarrow \mathbb{R}$ and can be evaluated pointwise. This means mathematically that for any $x \in X$, there is a function $l_x: V \rightarrow \mathbb{R}$ given by $l_x(f) = f(x)$. To become familiar with this definition, the reader is invited to verify l_x is linear.

If there is a norm on V then it can be asked whether the norm “controls” the pointwise values of elements of V . The following examples are intended to convey the concept prior to a rigorous definition.

Example 3.5. Define $C_2[0, 1]$ as in Example 3.1. Let $f(x)$ be the tent function given by $f(x) = 8x + 4$ for $x \in [-\frac{1}{2}, 0]$, $f(x) = 4 - 8x$ for

$x \in [0, \frac{1}{2}]$ and $f(x) = 0$ otherwise. Let $f_i(x) = i f(i^2(x - \frac{1}{2}))$ for $x \in [0, 1]$. Then $\|f_i\| = \frac{1}{i}$ decreases to zero yet $f_i(\frac{1}{2})$ diverges to infinity. No matter how small the norm is, there are elements of $C_2[0, 1]$ taking on arbitrarily large values pointwise.

The next example uses a version of the Cauchy-Schwarz inequality:

$$\left(\int_a^b g(t) dt \right)^2 \leq (b-a) \int_a^b g(t)^2 dt. \quad (3.1)$$

It is valid for any real-valued a and b , not just when $a < b$.

Example 3.6. Let $C^1[0, 1]$ be the space of continuously differentiable functions. (If $f \in C^1[0, 1]$ then $f'(x)$ exists and is continuous.) Define a norm by $\|f\|^2 = \int_0^1 f(x)^2 dx + \int_0^1 f'(x)^2 dx$. It is shown below that $\max_x |f(x)| \leq 2\|f\|$. The norm controls the values of $f(x)$ for all x .

Let $c = \max_x |f(x)|$. By considering $-f$ instead of f if necessary, it may be assumed that $c = \max_x f(x)$. If $f(x) \geq \frac{c}{2}$ for all x then $\|f\| \geq \frac{c}{2}$ and thus $\max_x |f(x)| \leq 2\|f\|$. Alternatively, there must exist a $y \in [0, 1]$ such that $f(y) = \frac{c}{2}$. Let $x \in [0, 1]$ be such that $f(x) = c$. Then $2 \int_y^x f'(t) dt = c$ because $f(x) = f(y) + \int_y^x f'(t) dt$. Applying (3.1) with $g(t) = f'(t)$ yields $c^2 \leq 4(x-y) \int_y^x f'(t)^2 dt$, from which it follows that $c^2 \leq 4 \int_0^1 f'(t)^2 dt$, that is, $\|f\| \geq \frac{c}{2}$. Thus, $\max_x |f(x)| \leq 2\|f\|$.

Linear functionals, such as l_x , fall into two categories, bounded and unbounded, based on whether $c_x = \sup_{f \in V} |l_x(f)| \cdot \|f\|^{-1}$ is finite or infinite. If l_x is bounded then $|l_x(f)| \leq c_x \|f\|$ for all f , demonstrating that the norm being small implies $f(x) = l_x(f)$ is small in magnitude.

A key requirement for $V \subset \mathbb{R}^X$ to be a RKHS is for the l_x to be bounded for all x . This implies that if $f_n \in V$ is a sequence converging in norm, meaning $\|f_n - f\| \rightarrow 0$, then the sequence also converges pointwise, meaning $f_n(x) \rightarrow f(x)$. This requirement was not mentioned earlier because l_x is automatically bounded when V is finite dimensional.

4

Infinite-dimensional RKHSs

The finite-dimensional RKHS theory introduced earlier suggests the general theory should take the following form. Fix an arbitrary set X . Endow a subspace V of the function space $\mathbb{R}^X$ with an inner product. RKHS theory should associate with the tuple $(V, \langle \cdot, \cdot \rangle, X)$ a set of vectors in $\mathbb{R}^X$ that span V , that vary continuously as $(V, \langle \cdot, \cdot \rangle)$ changes, and that encode the inner product on V . Furthermore, by analogy with Definition 2.1, an appropriate set should be the vectors k_x satisfying $\langle f, k_x \rangle = f(x)$ for all $f \in V$ and $x \in X$.

The above can be achieved with only several minor adjustments. Requiring the span of the k_x to be V is too demanding. For example, a basis for l^2 would necessarily be uncountable and cannot even be constructed explicitly, whereas the countable set of vectors $(1, 0, \dots)$, $(0, 1, 0, \dots)$, $\dots$ suffices for working with l^2 if taking limits is acceptable. Although seemingly wanting to reconstruct V by taking the topological closure of the span of the k_x , this would necessitate endowing $\mathbb{R}^X$ with a topology, and there might not be a topology on $\mathbb{R}^X$ compatible with all possible choices of $(V, \langle \cdot, \cdot \rangle)$. Instead, the mathematical process of completion (§4.1) can be used to achieve the same aim of adding to the span of the k_x any additional vectors that “should be”

infinitesimally close to the k_x . A Hilbert space is an inner product space that is complete, hence the above can be summarised by saying RKHS theory requires $(V, \langle \cdot, \cdot \rangle)$ to be a Hilbert space.

There may not exist a k_x such that $\langle f, k_x \rangle = f(x)$ for all $f \in V$ because the linear functional $f \mapsto \langle f, k_x \rangle$ is always bounded whereas Example 3.5 showed $f \mapsto f(x) = l_x(f)$ need not be bounded. Another advantage of requiring V to be a Hilbert space is that any bounded linear functional can be written as an inner product. (This is the Riesz representation theorem.) Therefore, a necessary and sufficient condition for $\langle f, k_x \rangle = f(x)$ to have a solution k_x is for l_x to be bounded. Requiring l_x to be bounded for all $x \in X$ is the second requirement RKHS theory places on $(V, \langle \cdot, \cdot \rangle)$.

RKHSs are similar to Euclidean spaces. A feature of Euclidean space is the presence of coordinate functions $\pi_i: \mathbb{R}^n \rightarrow \mathbb{R}$ sending $(x_1, \dots, x_n)$ to x_i . These coordinate functions are continuous. A RKHS replicates this: if $V \subset \mathbb{R}^X$ is a RKHS, the coordinate functions $\pi_x: V \rightarrow \mathbb{R}$ sending f to $f(x)$ are continuous by definition. (Recall that a linear functional is continuous if and only if it is bounded.)

4.1 Completions and Hilbert Spaces

If $V \subset W$ is not closed (§3.2) then there are points in W not in V but infinitesimally close to V . The closure $\bar{V}$ of V in W is the union of these infinitesimally close points and V . Perhaps a new W can be found though for which $\bar{V} \subset W$ is not closed?

Given a normed space V , there is a unique (up to an isometric isomorphism) normed space $\hat{V} \supset V$, called the *completion* of V , such that every point in $\hat{V}$ is infinitesimally close to V and there is no normed space W for which $\hat{V} \subset W$ is not closed. (Here, the norm on W must agree with the norm on $\hat{V}$ which must agree with the norm on V .) This means there is one and only one way to enlarge V maximally by adding infinitesimally close points. Once this is done, no more points can be added infinitesimally close to the enlarged space $\hat{V}$.

This fact permits a refinement of the previous visualisation of a non-closed subspace as being blurred. If $V \subset W$ is not closed then there are