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

the essence of knowledge
Boston — Delft

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 LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-093-4

Foundations and Trends® in Signal Processing
Volume 8, Issue 1-2, 2014

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

DOI: 10.1561/2000000050

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

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: 10.1561/2000000050.

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

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

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

$$f(\theta) = (r(\theta)\cos\theta, r(\theta)\sin\theta), \qquad r(\theta) = p_1\cos\theta + p_2\sin\theta. \tag{1.2}$$

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

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

The columns of $K(\theta)$ are

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

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:

$$f(L) = p_1\, k_1(L) + p_2\, k_2(L). \tag{1.5}$$

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:

$$\text{subspace} \to \text{kernel} \to \text{solution}. \tag{1.6}$$

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 $e_1, e_2, \cdots \in \mathcal{H}$. 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 $v_1, v_2, \cdots \in \mathcal{H}$ of vectors, each of which is a finite linear combination of the vectors $\{K(\cdot, t) \mid t \in T\}$, where $K\colon T \times T \to \mathbb{R}$ 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 tth 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 $\|v_k - v\| \to 0$, then v is fully determined by $v(t) = \lim_k v_k(t)$.

References

[1] P. Amblard. Versions récursives et adaptatives d'une mesure d'indépendance. In *GRETSI 2013, Brest, France*, 2013.

[2] P. O. Amblard and J. H. Manton. On data sparsification and a recursive algorithm for estimating a kernel-based measure of independence. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, 2013.

[3] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[4] F. R. Bach. Sharp analysis of low-rank kernel matrix approximations. *JMLR: Workshop and Conference Proceedings*, 30:1–25, 2013.

[5] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3(1–48), 2002.

[6] C. R. Baker. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289, 1973.

[7] R. J. Barton and H. V. Poor. An RKHS approach to robust l^2 estimation and signal detection. *IEEE Transactions on Information Theory*, 36(3):485–501, 1990.

[8] S. Bergman. *The Kernel Function and Conformal Mapping*, volume 5. American Mathematical Society, 1970.

[9] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, 2012.

[10] R. F. Brown. *A Topological Introduction to Nonlinear Analysis*. Birkhäuser Boston, Inc., Boston, MA, second edition, 2004.

[11] R. F. Curtain and H. Zwart. *An Introduction to Infinite-Dimensional Linear Systems Theory*, volume 21 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1995.

[12] G. Da Prato. *An Introduction to Infinite-Dimensional Analysis*. Universitext. Springer, 2006.

[13] L. Debnath and P. Mikusiński. *Introduction to Hilbert Spaces With Applications*. Elsevier Academic Press, third edition, 2005.

[14] M. F. Driscoll. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 26:309–316, 1973.

[15] D. L. Duttweiler and T. Kailath. RKHS approach to detection and estimation problems. IV. non-Gaussian detection. *IEEE Transactions on Information Theory*, 19(1):19–28, 1973.

[16] D. L. Duttweiler and T. Kailath. RKHS approach to detection and estimation problems. V. Parameter estimation. *IEEE Transactions on Information Theory*, 19(1):29–36, 1973.

[17] R. Elliott, L. Aggoun, and J. Moore. *Hidden Markov Models: Estimation and Control*. Applications of mathematics. Springer, 1995.

[18] Y. Engel, S. Mannor, and R. Meir. The kernel recursive least-squares algorithm. *IEEE Transactions on Signal Processing*, 52(8):2275–2284, 2004.

[19] S. Fine and K. Scheinberg. Efficient SVM training using low-rank kernel representations. *Journal of Machine Learning Research*, 2:243–264, 2001.

[20] R. Fortet. *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert (Random Vectors, Functions and Distributions in Hilbert Spaces)*. (in French), Hermès, 1995.

[21] K. Fukumizu, F. R. Bach, and M. I. Jordan. Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*, 4:73–99, 2004.

[22] K. Fukumizu, F. R. Bach, and M. I. Jordan. Kernel dimension reduction in regression. *The Annals of Statistics*, 37(4):1871–1905, 2009.

[23] K. Fukumizu, A. Gretton, X. Sun, and B. Scholkopf. Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems 20*, pages 489–496, 2007.

[24] K. Fukumizu, L. Song, and A. Gretton. Kernel bayes' rules: Bayesian inference with positive definite kernels. *Journal of Machine Learning Research*, 14:3753–3783, 2013.

[25] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

[26] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

[27] D. Han. *Frames for Undergraduates.* Student mathematical library. American Mathematical Society, 2007.

[28] T. Hida and N. Ikeda. Analysis on Hilbert space with reproducing kernel arising from multiple Wiener integral. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, (Berkeley, Calif., 1965/66). Vol. II: Contributions to Probability Theory, Part 1*, pages 117–143. University of California Press, Berkeley, CA USA, 1967.

[29] E. Hille. Introduction to general theory of reproducing kernels. *The Rocky Mountain Journal of Mathematics*, 2(3):321–368, 1972.

[30] J. Hoffmann-Jørgensen and G. Pisier. The law of large numbers and the central limit theorem in Banach spaces. *The Annals of Probability*, 4(4):587–599, 1976.

[31] T. Kailath. RKHS approach to detection and estimation problems. I. deterministic signals in Gaussian noise. *IEEE Transactions on Information Theory*, 17(5):530–549, 1971.

[32] T. Kailath and D. Duttweiler. An RKHS approach to detection and estimation problems. III. generalized innovations representations and a likelihood-ratio formula. *IEEE Transactions on Information Theory*, 18(6):730–745, 1972.

[33] T. Kailath, R. Geesey, and H. Weinert. Some relations among RKHS norms, Fredholm equations, and innovations representations. *IEEE Transactions on Information Theory*, 18(3):341–348, May 1972.

[34] T. Kailath and H. L. Weinert. An RKHS approach to detection and estimation problems. II. Gaussian signal detection. *IEEE Transactions on Information Theory*, 21(1):15–23, 1975.

[35] G. Kallianpur. The role of reproducing kernel Hilbert spaces in the study of Gaussian processes. In *Advances in Probability and Related Topics*, volume 2, pages 49–83. Dekker, New York, 1970.

[36] G. S. Kimeldorf and G. Wahba. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495—502., 1970.

[37] F. M. Larkin. Optimal approximation in Hilbert spaces with reproducing kernel functions. *Mathematics of Computation*, 24(112):911–921, 1970.

[38] S. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

[39] W. Liu, P. Pokharel, and J. Principe. The kernel least mean squares algorithm. *IEEE Transactions on Signal Processing*, 56(2):543–554, 2008.

[40] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons Inc., New York, 1969.

[41] M. Lukic and J. Beder. Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Transactions of the American Mathematical Society*, 353(10):3945–3969, 2001.

[42] J. H. Manton. A primer on stochastic differential geometry for signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 7(4):681–699, 2013.

[43] J. Matoušek. *Thirty-three Miniatures: Mathematical and Algorithmic Applications of Linear Algebra*. Student mathematical library. American Mathematical Society, 2010.

[44] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 209:415–446, 1909.

[45] H. J. Newton. A conversation with Emanuel Parzen. *Statistical Science*, 17(3):357–378, 2002.

[46] W. D. Parreira, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret. Stochastic behavior analysis of the Gaussian kernel least mean square algorithm. *IEEE Transactions on Signal Processing*, 60(5):2208–2222, 2012.

[47] E. Parzen. Statistical inference on time series by Hilbert space methods. Technical Report 23, Applied Mathematics and Statistics Laboratory, Stanford University, 1959.

[48] E. Parzen. Extraction and detection problems and reproducing kernel Hilbert spaces. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control*, 1(1):35–62, 1962.

[49] E. Parzen. Probability density functionals and reproducing kernel Hilbert spaces. In *Proceedings of the Symposium on Time Series Analysis*, volume 196, pages 155–169. Wiley, New York, 1963.

[50] E. Parzen. Statistical inference on time series by RKHS methods. In *Proceedings of the 12th Biennial Seminar of the Canadian Mathematical Congress on Time Series and Stochastic Processes, Convexity and Combinatorics, Vancouver, 1969*, pages 1–37. Canadian Mathematical Congress, Montreal, 1970.

[51] V. I. Paulsen. An introduction to the theory of reproducing kernel Hilbert spaces. `http://www.math.uh.edu/~vern/rkhs.pdf`, 2009.

[52] B. Picinbono and P. Duvaut. Optimal linear-quadratic systems for detection and estimation. *IEEE Transactions on Information Theory*, 34(2):304–311, 1988.

[53] B. Picinbono and P. Duvaut. Geometrical properties of optimal Volterra filters for signal detection. *IEEE Transactions on Information Theory*, 36(5):1061–1068, 1990.

[54] T. S. Pitcher. Likelihood ratios of Gaussian processes. *Arkiv för Matematik*, 4:35–44, 1960.

[55] H. Poor. *An Introduction to Signal Detection and Estimation*. Springer, 1994.

[56] A. Rényi. On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungarica*, 10(2):441–451, 1959.

[57] C. Richard, J. C. M. Bermudez, and P. Honeine. Online prediction of time series data with kernels. *IEEE Transactions on Signal Processing*, 57(3):1058–1067, 2009.

[58] S. Saitoh. *Theory of Reproducing Kernels and Its Applications*, volume 189 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1988.

[59] B. Schölkopf and A. J. Smola. *Learning With Kernels*. MIT Press, Cambridge, MA, USA, 2002.

[60] K. Slavakis, P. Bouboulis, and S. Theodoridis. *Academic Press Library in Signal Processing: Volume 1, Signal Processing Theory and Machine Learning*, chapter 17, Online Learning in Reproducing Kernel Hilbert Spaces, pages 883–987. Elsevier, 2014.

[61] C. G. Small and D. L. McLeish. *Hilbert Space Methods in Probability and Statistical Inference*. John Wiley & Sons, 2011.

[62] L. Song, K. Fukumizu, and A. Gretton. Kernel embeddings of conditional distributions. *IEEE Signal Processing Magazine*, 30(4):98–111, 2013.

[63] L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13:1393–1434, 2012.

[64] B. K. Sriperumbudur, K. Fukumizu, and G. R. G. Lanckriet. Universality, characteristic kernels and RKHS embeddings of measures. *Journal of Machine Learning Research*, 12:2389–2410, 2011.

[65] F. Steinke and B. Schölkopf. Kernels, regularization and differential equations. *Pattern Recognition*, 41(11):3271–3286, 2008.

[66] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of Machine Learning Research*, 2:67–93, 2001.

[67] F. Stulajter. RKHS approach to nonlinear estimation of random variables. In *Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1978), Vol. B*, pages 239–246. Reidel, Dordrecht, 1978.

[68] Y. Sun, F. Gomez, and J. Schmidhuber. On the size of the online kernel sparsification dictionary. In *Proceedings of the International Conference on Machine Learning, Edinburgh, Scotland*, 2012.

[69] H. J. Sussmann and J. C. Willems. 300 years of optimal control: From the brachystochrone to the maximum principle. *IEEE Control Systems Magazine*, 17(3):32–44, 1997.

[70] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. John Wiley&Sons, 1977.

[71] G. G. Wahba. *Interpolating Spline Methods for Density Estimation*. Department of Statistics, University of Wisconsin, 1973.

[72] G. G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981.

[73] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley&Sons, 1989.

[74] D. Williams. *Probability with Martingales*. Cambridge mathematical textbooks. Cambridge University Press, 1991.

[75] E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems*. Springer Texts in Electrical Engineering. Springer New York, 2011.

[76] K. Yao. Applications of reproducing kernel Hilbert spaces — bandlimited signal models. *Information and Control*, 11(4):429–444, 1967.