Separability of reproducing kernel Hilbert spaces

Houman Owhadi and Clint Scovel
California Institute of Technology

June 16, 2015

Abstract

We demonstrate that a reproducing kernel Hilbert space of functions on a separable absolute Borel space or an analytic subset of a Polish space is separable if it possesses a Borel measurable feature map.

1 Introduction

Reproducing kernel Hilbert spaces (RKHS), see e.g. Berlinet and Thomas-Agnan [1] and Steinwart and Christmann [23, Sec. 4], are important in Statistics and Learning Theory. Moreover, when using these spaces in Probability and Statistics, separability has powerful effects. For example, for separable metrizable spaces we have: $\mathcal{B}(X\times X)=\mathcal{B}(X)\times\mathcal{B}(X)$ for the Borel σ-algebras, the Ky-Fan metric can be defined so as to metrize convergence in probability, convergence in probability implies convergence in law, convergence in law is metrized by the Prokhorov metric, the space of probability measures with the weak topology is separable and metrizable, and the Kantorovich-Rubinstein and Strassen theorems have sharp forms. Moreover, separable Hilbert spaces are Polish so that we have all the machinery of descriptive set theory available, regular conditional probabilities exist, Bochner integration is simple, and all probability measures on them are tight. Most importantly, by a classical result, see e.g. Halmos [13, Prob. 17], all separable Hilbert spaces are isomorphic with ℓ^2.

According to Montgomery [18], "Separability is a property which greatly facilitates work in metric spaces, but it may be of some interest to point out that this property has been unnecessarily assumed in the proofs of certain theorems concerning such spaces and concerning functions defined on them." Indeed, many works do assume separability of the RKHS. For example, Steinwart and Christmann's [23, Thm. 7.22] oracle inequality for SVMs, Christmann and Steinwart [6, Thms. 7 & 12], [5], [4], Steinwart and Christmann [24], [25], De Vito, Rosasco and Toigo [7], Hable and Christmann [12, Thm. 3.2], Lukić and Beder [17], Steinwart [22] and Vovk [28, Thm. 3]. De Vito, Umanità and Villa [8] assume it in their generalization of Mercer's theorem to matrix valued kernels, and Christmann, Van Messem and Steinwart [6] assert that Support Vector Machines (SVMs) are known to be consistent and robust for classification and regression if they are based on a Lipschitz continuous loss function and on a bounded kernel with a dense and separable reproducing kernel Hilbert space. Cambanis [2] proves that a positive definite function is the reproducing kernel corresponding to the autocorrelation function of a

stochastic processes if and only if the corresponding RKHS is separable, and proves that a second order stochastic process is oscillatory if and only its corresponding RKHS space is separable. Nashed and Walter [19] require a separable RKHS in their development of sampling theorems for functions in reproducing kernel Hilbert spaces, and Zhang and Zhang [31] in reproducing kernel Banach spaces. Hein and Bousquet [15] require it and give some sufficient conditions for it. For an example of a non-separable RKHS, see Canu, Mary, and Rakotomamonjy [3, Ex. 8.1.6].

It is the purpose of this paper to establish separability for both RKHSs and reproducing kernel Banach spaces when the domain is a separable absolute Borel space or an analytic subset of a Polish space, in particular when it is a Borel subset of a Polish space, under the simple assumption that the reproducing kernel space possesses a Borel measurable feature map.

2 Main Results

Before our main result, we review some existing results regarding the separability of RKHSs. Steinwart and Christmann [23, Lem. 4.33] asserts that if X is separable and the kernel k corresponding to the RKHS H is continuous, then H is separable. More generally, Steinwart and Scovel [26, Cor. 3.6] show that if there exists a finite and strictly positive Borel measure on X, then every bounded and separately continuous kernel k has a separable RKHS. Also, Berlinet and Thomas-Agnan [1, Thm. 15, pg. 33] shows that a RKHS H is separable if there is a countable subset $X_0 \subset X$ such that if $f \in H$ and $f(x) = 0, x \in X_0$ implies that $f = 0$. A result of Fortet [9, Thm. 1.2] asserts that a RKHS with kernel k is separable if and only if for all $\varepsilon > 0$ there exists a countable partition $B_j, j \in \mathbb{N}$ of X such that for all $j \in \mathbb{N}$ and all $x_1, x_2 \in B_j$ we have

$$k(x_1, x_1) + k(x_2, x_2) - k(x_1, x_2) - k(x_2, x_1) < \varepsilon\,.$$

Reproducing Kernel Banach Spaces (RKBS), introduced by Zhang, Xu, and Zhang in [30] are Banach spaces of real valued functions for which point evaluation is continuous. If and only if characterization of separability is obtained through a generalization of Fortet's Theorem from RKHSs to RKBSs. We suspect the proof of our version, Lemma 2.2, is similar to Fortet's [9, Thm. 2.1] for RKHSs, but it is not written down there. Indeed Fortet's, result mentioned above, is a regularity condition on the pullback (pseudo) metric

$$d_H(x_1, x_2) := \|\Phi(x_1) - \Phi(x_2)\|_{H_1} = \sqrt{k(x_1, x_1) + k(x_2, x_2) - k(x_1, x_2) - k(x_2, x_1)}$$

to X determined by a feature map $\Phi : X \to H_1$. In particular, Fortet's condition then becomes: for all $j \in \mathbb{N}$ and all $x_1, x_2 \in B_j$ we have

$$d_H(x_1, x_2) < \sqrt{\varepsilon}\,.$$

We refer to [30] for the foundational facts and terminology regarding RKBSs. We begin with a preparatory lemma asserting that the separability of the image of the feature map implies the separability of the corresponding RKHS or RKBS.

Lemma 2.1. *Consider a (RKBS) RKHS $\mathcal{K}$ with feature (Banach) Hilbert space $\mathcal{W}$ and (primary) feature map $\Phi : X \to \mathcal{W}$. If $\Phi(X) \subset \mathcal{W}$ is a separable subspace, then $\mathcal{K}$ is separable.*

Since separability is preserved under continuous maps, see e.g. [29, Thm. 16.4], we conclude the RKBS version of Steinwart and Christmann [23, Lem. 4.33] when combined with [23, Lem. 4.29]: A RKBS of functions on a separable space X is separable if it has a continuous feature map.

Lemma 2.2 (Fortet)**.** *A RKBS B is separable if and only if there exists a feature Banach space $\mathcal{W}$ and feature map $\Phi : X \to \mathcal{W}$ for B such that for all $\varepsilon > 0$ there is a countable partition $B_j \subset X, j \in \mathbb{N}$ with $\cup_{j\in\mathbb{N}} B_j = X$ such that for all $j \in \mathbb{N}$ and all $x_1, x_2 \in B_j$, we have*

$$\|\Phi(x_1) - \Phi(x_2)\|_{\mathcal{W}} < \varepsilon\,. \tag{2.1}$$

Finally, our main tool to derive separability comes from theorems of Stone [27, Thm. 16, pg. 32], when X is separable absolute Borel, and Srivastava's [21, Thm. 4.3.8] version of Simpson [20] when X is an analytic subset of a Polish space, both of which apply when X is a Borel subset of a Polish space. Following Frolik [11], a metrizable space X is said to be absolute Borel if $X \subset Z$ is a Borel subset for all metrizable Z for which it is a subspace. In particular, Frolik [11, Thm. 1] asserts that a Borel subset of Polish space is separable absolute Borel. It follows from the definition, see e.g. Srivastava [21, Pg. 128], that a Borel subset of a Polish space is analytic. Moreover, Frolik [10] introduces *bianalytic spaces* as analytic spaces such that their complement in their Čech compactification is also analytic and, in Frolik [10, Thm. 12], shows that a metrizable space is separable absolute Borel if and only if it is bianalytic.

Theorem 2.3. *Let X be separable absolute Borel or an analytic subset of a Polish space and let Y be a metric space, and suppose that $f : X \to Y$ is Borel measurable. Then $f(X) \subset Y$ is separable.*

Steinwart and Christmann [23, Lem. 4.25] shows that separate measurability of the kernel combined with separability of the corresponding RKHS implies that the canonical feature map is measurable. Our main result is a kind of converse when X is separable absolute Borel or an analytic subset of a Polish space.

Theorem 2.4. *Let X be separable absolute Borel or an analytic subset of a Polish space and let $\mathcal{K}$ be a RKHS with measurable feature map, or a RKBS with measurable primary feature map, of real-valued functions on X. Then $\mathcal{K}$ is separable.*

3 Proofs

3.1 Proof of Lemma 2.1

For RKHS this assertion is contained in the proof of Steinwart and Christmann [23, Lem. 4.33] and is essentially the same for the RKBS case. Roughly, the argument is that rational linear combinations are dense in the linear span of $\Phi(X)$ and the linear span is dense in the closed linear span in the metric defined in the proof of [23, Thm. 4.21].

3.2 Proof of Lemma 2.2

Let us begin with "if". To that end, let us show that condition 2.1 implies that $\Phi(X)$ is separable. Indeed, fix $\varepsilon > 0$ and for each $\frac{\varepsilon}{2^k}, k \in \mathbb{N}$ let $B_j^k, j \in \mathbb{N}$ denote

the corresponding partition and let $x_j^k \in B_j^k$ denote a selection. Then the set $\Phi(x_j^k), k \in \mathcal{N}, j \in \mathcal{N}$ is countable, and it is easy to show it is dense in $\Phi(X)$. That is, $\Phi(X)$ is separable, and the separability of $\mathcal{K}$ follows from Lemma 2.1. Now for the "only if", suppose that B is separable. Then the canonical feature space $\mathcal{W} := B$ is separable, and since B is metric, by e.g. [29, Thm. 16.8] it is second countable. Therefore, since second countability is inherited by subspaces, see e.g. [29, Thm. 16.2], it follows for the corresponding canonical feature map $\Phi : X \to B$, that $\Phi(X) \subset B$ is second countable, and therefore, by e.g. [29, Thm. 16.9], it is separable. Therefore there exists a countable dense set $\Phi(x_j) \in \Phi(X), j \in \mathbb{N}$. Therefore, if for each $\varepsilon > 0$ and for each $j \in \mathbb{N}$ we define $B_j = \{x \in X : \|\Phi(x_j) - \Phi(x)\|_B < \frac{\varepsilon}{2}\}$, it follows that $\cup_{j\in\mathbb{N}} B_j = X$ and $\|\Phi(x_1) - \Phi(x_2)\|_B < \varepsilon$ for all $x_1, x_2 \in B_j$.

3.3 Proof of Theorem 2.3

The case when X is an analytic subset of a Polish space follows directly from Srivastava [21, Thm. 4.3.8]. When X is separable absolute Borel, it follows from Stone's Theorem [27, Thm. 16, pg. 32] that when Y is a metric space and $\Phi : X \to Y$ is a measurable bijection, that the image Y is separable. However, when Φ is not surjective, since $\Phi(X) \subset Y$ is a metric space, the assertion that the metric subspace $\Phi(X) \subset Y$ is separable follows assuming that Φ is a measurable injection. Moreover, injectivity is also unnecessary. To see this, extend to the injective map $\hat{\Phi} : X \to X \times Y$ defined by $\hat{\Phi}(x) := \big(x, \Phi(x)\big)$. Then it follows from Hansell's [14, Thm. 1] generalization of Kuratowski [16, Thm. 1, Sec. 31, VI] to the nonseparable case, that $\hat{\Phi}$ is measurable. To see how it is obtained, since X is assumed to be separable and metrizable, it is second countable, see e.g. [29, Thm. 16.11], so that it has a countable base $\{G_n, n \in \mathbb{N}\}$ of open sets generating its topology. Let $W \subset X \times Y$ be open and define

$$V_n = \cup\{V : V \text{ open}, G_n \times V \subset W\}\,.$$

Then

$$W = \underset{n\in\mathbb{N}}{\cup} G_n \times V_n$$

and therefore

$$\hat{\Phi}^{-1}(W) = \underset{n\in\mathbb{N}}{\cup} G_n \cap \Phi^{-1}(V_n)\,.$$

Since G_n and V_n are open and therefore measurable and Φ is measurable it follows that $\hat{\Phi}^{-1}(W)$ is measurable. Consequently, since the open sets generate the Borel σ-algebra, it follows that $\hat{\Phi}$ is Borel measurable. Moreover, since $\hat{\Phi}$ is injective the above discussion shows that $\hat{\Phi}(X) \subset X \times Y$ is separable. Since $\Phi(X) = P_Y \hat{\Phi}(X)$ where P_Y is the projection onto the second component and P_Y is continuous, and separability is preserved under continuous maps, see e.g. [29, Thm. 16.4], it follows that $\Phi(X) \subset Y$ is separable.

3.4 Proof of Theorem 2.4

Since the feature space is metric, Theorem 2.3 implies that the image $\Phi(X)$ is separable for any measurable feature map Φ. The assertion then follows from Lemma 2.1.

Acknowledgments

The authors gratefully acknowledge this work supported by the Air Force Office of Scientific Research under Award Number FA9550-12-1-0389 (Scientific Computation of Optimal Statistical Estimators).

References

[1] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, Boston, MA, 2004.

[2] S. Cambanis. The measurability of a stochastic process of second order and its linear space. *Proceedings of the American Mathematical Society*, 47(2):467–475, 1975.

[3] S. Canu, X. Mary, and A. Rakotomamonjy. Functional learning through kernels. In *Advances in Learning Theory: Methods, Models and Applications*, volume 190 of *NATO Science Series*, pages 89–110. IOS Press, 2003.

[4] A. Christmann and I. Steinwart. Consistency of kernel-based quantile regression. *Applied Stochastic Models in Business and Industry*, 24(2):171–183, 2008.

[5] A. Christmann and I. Steinwart. Universal kernels on non-standard input spaces. In *Advances in neural information processing systems*, pages 406–414, 2010.

[6] A. Christmann, A. Van Messem, and I. Steinwart. On consistency and robustness properties of support vector machines for heavy-tailed distributions. *Statistics and Its Interface*, 2:311–327, 2009.

[7] E. De Vito, L. Rosasco, and A. Toigo. Learning sets with separating kernels. *Applied and Computational Harmonic Analysis*, 37(2):185–217, 2014.

[8] E. De Vito, V. Umanità, and S. Villa. An extension of Mercer theorem to matrix-valued measurable kernels. *Applied and Computational Harmonic Analysis*, 34(3):339–351, 2013.

[9] R. Fortet. Espaces á noyau reproduisant et lois de probabilités des fonctions aléatoires. *Annales de l'I. H. P., Sec. B*, 9(1):41–58, 1973.

[10] Z. Frolik. On bianalytic spaces. *Czechoslovak Math. J.*, 13 (88):561–573, 1963.

[11] Z. Frolik. Absolute Borel and Souslin sets. *Pacific J. Math.*, 32:663–683, 1970.

[12] R. Hable and A. Christmann. Estimation of scale functions to model heteroscedasticity by regularised kernel-based quantile methods. *Journal of Nonparametric Statistics*, 26(2):219–239, 2014.

[13] P. R. Halmos. *A Hilbert Space Problem Book*, volume 19. Springer-Verlag, 1982.

[14] R. W. Hansell. Sums, products and continuity of Borel maps in nonseparable metric spaces. *Proceedings of the American Mathematical Society*, 104(2):465–471, 1988.

[15] M. Hein and O. Bousquet. Kernels, associated structures and generalizations. *Max-Planck-Institut fuer biologische Kybernetik, Technical Report*, 2004.

[16] K. Kuratowski. Topology, vol. 1. *Acad. Press, Warsaw*, 1966.

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

[18] D. Montgomery. Non-separable metric spaces. *Fundamenta Mathematicae*, 1(25):527–533, 1935.

[19] M. Z. Nashed and G. G. Walter. General sampling theorems for functions in reproducing kernel Hilbert spaces. *Mathematics of Control, Signals and Systems*, 4(4):363–390, 1991.

[20] S. Simpson. BQO theory and Fraïssé's conjecture. *Recursive Aspects of Descriptive Set Theory, Oxford University Press, Oxford*, 1985.

[21] S. M. Srivastava. *A Course on Borel Sets*, volume 180. Springer, 1998.

[22] I. Steinwart. Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *arXiv preprint arXiv:1403.1040*, 2014.

[23] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, New York, 2008.

[24] I. Steinwart and A. Christmann. Sparsity of SVMs that use the ϵ-insensitive loss. In *Advances in Neural Information Processing Systems*, pages 1569–1576, 2009.

[25] I. Steinwart and A. Christmann. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17(1):211–225, 2011.

[26] I. Steinwart and C. Scovel. Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.*, 35(3):363–417, 2012.

[27] A. H. Stone. Non-separable Borel sets. *Rozprawy Mat.(=Dissertations Math.)*, 28:1–40, 1962.

[28] V. Vovk. On-line regression competitive with reproducing kernel Hilbert spaces. In *Theory and applications of models of computation*, pages 452–463. Springer, 2006.

[29] S. Willard. *General Topology*. Addison-Wesley, 1970.

[30] H. Zhang, Y. Xu, and J. Zhang. Reproducing kernel Banach spaces for machine learning. *J. Mach. Learn. Res.*, 10:2741–2775, 2009.

[31] H. Zhang and J. Zhang. Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products. *Applied and Computational Harmonic Analysis*, 31(1):1–25, 2011.