Research Statement

Irina Holmes

October 2013

My current research interests lie in infinite-dimensional analysis and geometry, probability and statistics, and machine learning. The main focus of my graduate studies has been the development of the Gaussian Radon transform for Banach spaces, an infinite-dimensional generalization of the classical Radon transform. This transform and some of its properties are discussed in Sections 2 and 3.1 below. Most recently, I have been studying applications of the Gaussian Radon transform to machine learning, an aspect discussed in Section 4.

1 Background

The Radon transform was first developed by Johann Radon in 1917. For a function $f : \mathbb{R}^n \to \mathbb{R}$ the Radon transform is the function Rf defined on the set of all hyperplanes P in $\mathbb{R}^n$ given by:

$$Rf(P) \stackrel{\text{def}}{=} \int_P f\,dx,$$

where, for every P, integration is with respect to Lebesgue measure on P. If we think of the hyperplane P as a "ray" shooting through the support of f, the integral of f over P can be viewed as a way to measure the changes in the "density" of f as the ray passes through it. In other words, Rf may be used to reconstruct the density of an n-dimensional object from its $(n-1)$-dimensional cross-sections in different directions. Through this line of thinking, the Radon transform became the mathematical background for medical CT scans, tomography and other image reconstruction applications.

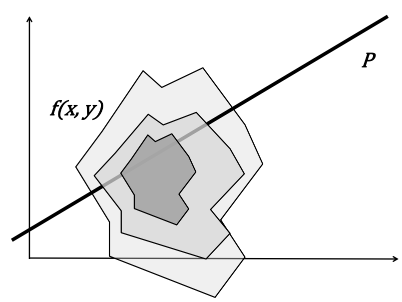

Figure 1: The Radon Transform

Besides the intrinsic mathematical and theoretical value, a practical motivation for our work is the ability to obtain information about a function defined on an infinite-dimensional space from its conditional expectations. Infinite-dimensional Radon transforms were developed in [4], [3] - within the framework of nuclear spaces - and in [17] - in the context of Hilbert spaces. However, the current standard framework in infinite-dimensional analysis and probability is that of *abstract Wiener spaces*, developed by L. Gross in [8]. A Radon transform theory in the setting of abstract Wiener spaces has potential to flourish in conjunction with the multitude of results already established in this field. This was the project proposed for my thesis research.

2 The Gaussian Radon Transform

The first obstacle was the usual one in infinite-dimensional analysis, namely the absence of a useful version of Lebesgue measure in infinite dimensions. However, Gaussian measures behave well in infinite dimensions. An abstract Wiener space is a triple (H, B, μ) where:

- H is a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$.
- B is the Banach space obtained by completing H with respect to a *measurable norm* $| \cdot |$. Specifically, a norm $| \cdot |$ on H is said to be measurable if for all $\epsilon > 0$ there is a finite-dimensional subspace F_ϵ of H such that:
$$\gamma_F \left[h \in F : |h| > \epsilon \right] < \epsilon \tag{1}$$
for all finite-dimensional subspaces F of H with $F \perp F_\epsilon$, where γ_F denotes standard Gaussian measure on F.
- μ is Wiener, or standard Gaussian, measure on B (see (4) below).

Any measurable norm $| \cdot |$ on H is weaker than the original Hilbert norm $\| \cdot \|$ (which is *not* measurable if H is infinite-dimensional), therefore the restriction to H of any $x^* \in B^*$ is continuous with respect to $\| \cdot \|$. We then have the map:

$$B^* \to H^*; \, x^* \mapsto h_{x^*}, \tag{2}$$

where for every $x^* \in B^*$, h_{x^*} is the unique element of H such that:

$$\langle h, h_{x^*} \rangle = (h, x^*) \,, \text{ for all } h \in H. \tag{3}$$

Thus B^* is continuously embedded as a dense subspace $H_{B^*} = \{h_{x^*} \in H : x^* \in B^*\}$ of H. With this notation, we may specify that Wiener measure μ is the unique Borel measure on B with:

$$\int_B e^{ix^*} \, d\mu = e^{-\frac{1}{2}\|h_{x^*}\|^2} \tag{4}$$

for all $x^* \in B^*$. Moreover, for any real separable infinite-dimensional Banach space B with a centered Gaussian measure μ, there is a Hilbert space H (called the *Cameron-Martin space*) such that (H, B, μ) is an abstract Wiener space.

The linear map $H_{B^*} \to L^2(B, \mu)$ sending $h_{x^*} \mapsto (\cdot, x^*)$ is continuous with respect to the Hilbert norm $\| \cdot \|$, and therefore has a unique extension to H which we denote by I:

$$I : H \to L^2(B, \mu); \, h \mapsto Ih. \tag{5}$$

This map is an isometry and Ih is centered Gaussian with mean $\|h\|^2$ for every $h \in H$.

These results were proved by L. Gross in the celebrated work [8]. For a comprehensive view on abstract Wiener spaces, see Kuo [16], and for an insightful summary see Stroock [25].

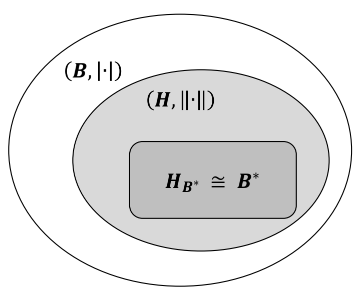

Figure 2: Abstract Wiener space: B^* is continuously embedded as a dense subspace of H.

My first work in this field was the paper [10], joint with my adviser, which appeared in the Journal of Functional Analysis in 2012. In Theorem 2.1 of this work we constructed Gaussian measures on B that are concentrated on B-closures of closed affine subspaces in H, that is sets of the form $M_p = p + M_0$ where M_0 is a closed subspace of H and $p \in M^{\perp}$.

Theorem 2.1 *Let (H, B, μ) be an abstract Wiener space and M_0 be a closed subspace of H. For every translate $M_p = p + M_0$, where $p \in M_0^{\perp}$, there is a unique Borel measure μ_{M_p} on B such that:*

$$\int_B e^{ix^*}\, d\mu_{M_p} = e^{i\langle p, h_{x^*}\rangle - \frac{1}{2}\|P_{M_0} h_{x^*}\|^2} \text{ for all } x^* \in B^*, \tag{6}$$

where P_{M_0} denotes orthogonal projection in H onto M_0. Moreover, μ_{M_p} is concentrated on the closure $\overline{M_p}$ of M_p in B.

While this result is more general, it has its origins in our observation that every hyperplane in B is the closure of a hyperplane in H, where by a *hyperplane* in a topological vector space we mean a translate of a closed subspace of codimension 1. However, this is not a one-to-one relationship, as Proposition 5.2 of [10] shows:

Proposition 2.2 *Let (H, B, μ) be an abstract Wiener space. Then:*

(i) If P is a hyperplane in B, then there is a unique hyperplane ξ in H such that $P = \overline{\xi}$, where we are taking the closure in B.

(ii) Let $\xi = pu + u^{\perp}$ be a hyperplane in H, where $p > 0$ and $u \in H$ is a unit vector. Then:

(a) If $u \in H_{B^*}$, then $\overline{\xi}$ is a hyperplane in B.

(b) If $u \notin H_{B^*}$, then $\overline{\xi} = B$.

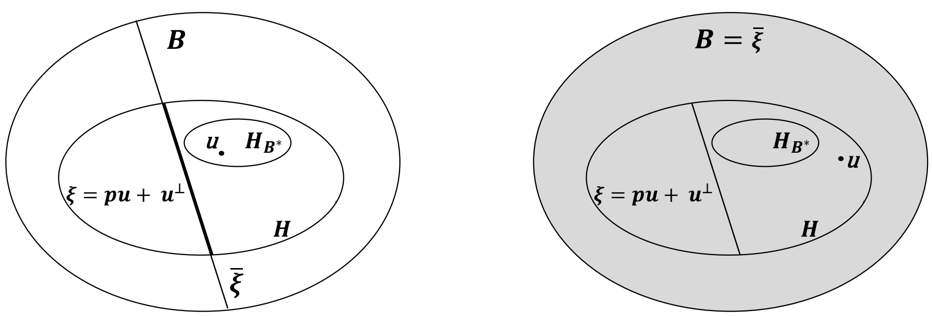

Figure 3: If $u \in H_{B^*}$ then $\overline{\xi}$ is a hyperplane in B; otherwise, $\overline{\xi} = B$.

This led me to perform a more detailed analysis of the relationship between the closed subspaces of finite codimension in B and those in H, which will appear in my thesis and a future paper.

Definition 2.1 *Let f be a bounded Borel function on B. We define the* **Gaussian Radon transform** *Gf by:*

$$Gf(M_p) \stackrel{def}{=} \int_B f\, d\mu_{M_p} \tag{7}$$

for all closed affine subspaces $M_p = p + M_0$ in H, where μ_{M_p} is the Gaussian measure on B concentrated on $\overline{M_p}$ constructed in Theorem 2.1.

3 Properties of the Gaussian Radon Transform

3.1 The Support Theorem

One of the most important results about the classical Radon transform is the Helgason support theorem. In [10] we proved an analogue of this result for the Gaussian Radon transform:

Theorem 3.1 *Let (H, B, μ) be an abstract Wiener space and f be a bounded, continuous function on B. Let K be a closed, bounded and convex subset of H such that the Gaussian Radon transform Gf of f is 0 on all hyperplanes in H that do not intersect K. Then f is 0 on the complement of K in B.*

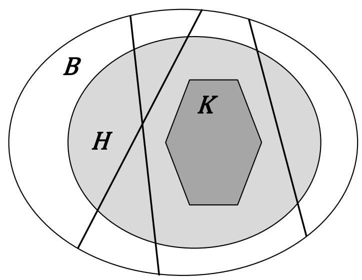

Figure 4: The Support Theorem

As in the classical case, this result provides information about the support of f, given knowledge about the support of Gf.

3.2 Disintegration Theorem and Conditional Expectation

Continuing the ideas I had been developing, in my next work [11] (which has been accepted for publication in the journal "Infinite Dimensional Analysis, Quantum Probability and Related Topics"), I proved a disintegration of Wiener measure by the measures constructed in Theorem 2.1.

Theorem 3.2 *Let (H, B, μ) be an abstract Wiener space and Q_0 be a closed subspace of finite codimension in H. Then the map:*

$$Q_0^{\perp} \ni p \mapsto Gf(Q_p) \tag{8}$$

is Borel measurable on $Q_0^{\perp}$ for all non-negative Borel functions f on B. Moreover:

$$\int_B f\, d\mu = \int_{Q_0^{\perp}} (Gf(Q_p))\, d\gamma_{Q_0^{\perp}}(p) \tag{9}$$

for all Borel functions $f : B \to \mathbb{C}$ for which the left side exists, where $\gamma_{Q_0^{\perp}}$ is standard Gaussian measure on $Q_0^{\perp}$.

This result led to a few interesting consequences, one of which (Corollary 3.2 of [11]) expresses the Gaussian Radon transform as a conditional expectation. Let me include here an immediate consequence of this result:

Proposition 3.3 *Let (H, B, μ) be an abstract Wiener space and linearly independent elements $h_1, h_2, \ldots, h_n$ of H. Then:*

$$\mathbb{E}\left[f | Ih_1 = y_1, \ldots, Ih_n = y_n\right] = Gf\left(\bigcap_{j=1}^{n}\left[\langle h_j, \cdot\rangle = y_j\right]\right). \tag{10}$$

for all $f \in L^2(B, \mu)$, where $I : H \to L^2(B, \mu)$ is the map discussed in (5).

3.3 An Inversion Procedure

In the work [11] I also obtained an inversion procedure for the Gaussian Radon transform using the Segal-Bargmann transform for abstract Wiener spaces. For background on the Segal-Bargmann transform, see [2], [6], [7], [9] and [20], [21].

Theorem 3.4 *Let (H, B, μ) be an abstract Wiener space, $f \in L^2(B, \mu)$ and Q_0 be a closed subspace of finite codimension in H. Then:*

$$S_{Q_0^\perp} \circ Gf(Q_p) = (S_B f)|_{(Q_0^\perp)_{\mathbb{C}}} \tag{11}$$

where $S_{Q_0^\perp}$ and S_B are the Segal-Bargmann transforms on $L^2(Q_0, \gamma_{Q_0^\perp})$ and $L^2(B, \mu)$, respectively, and $(Q_0^\perp)_{\mathbb{C}}$ denotes the complexification of $Q_0^\perp$.

In particular, (11) implies that for all non-zero $h \in H$:

$$S\left[Gf\left(P(\cdot, h)\right)\right](\|h\|) = S_B f(h) \tag{12}$$

where S is the Segal-Bargmann transform on $\mathbb{R}$ and $P(t, h)$ is the hyperplane in H given by:

$$P(t, h) = \frac{t}{\|h\|} h + h^\perp$$

for all $t > 0$. In other words, if we know $Gf(P)$ for all hyperplanes in H, then we know $S_B f|_H$; taking the holomorphic extension to the complexification $H_{\mathbb{C}}$ of H, we know $S_B f$ and we can then obtain f from the inverse Segal-Bargmann transform.

4 Machine Learning and the Gaussian Radon Transform

The Gaussian Radon transform has recently led us in the direction of machine learning. Roughly, one may think of machine learning as the study of predicting the future from known snapshots from the past.

Support vector machine (SVM) methods have been extremely popular in machine learning for quite some time, but recent literature has seen an increasing interest in probabilistic interpretations of kernel-based methods (see [1, 13, 18, 19, 22, 28] for a few examples). SVM's involve projecting the data into an (often infinite-dimensional) reproducing kernel Hilbert space (RKHS) H. When H is finite-dimensional, there is a clear Bayesian interpretation of kernel methods, but when H is infinite-dimensional, the absence of Lebesgue measure on H presents an obstacle. Since the concept of abstract Wiener space was motivated exactly by this lack of a "standard Gaussian measure" on infinite-dimensional Hilbert spaces, we decided to explore some kernel-based methods from this perspective. Moreover, the fact that the Gaussian Radon transform is really a conditional expectation gives it potential to be applied in this setting. Our results so far, the most important of which I will briefly discuss below, appear in [12], submitted in October 2013.

Let $\mathcal{X}$ be a non-empty set, the *input space*, and:

$$D = \{(p_1, y_1), \dots, (p_n, y_n)\}$$

be a finite collection of *input values* p_j together with their corresponding real *outputs* y_j. Our goal is to predict the output y corresponding to a (yet unobserved) input value $p \in \mathcal{X}$ by finding a suitable *decision function* $f : \mathcal{X} \to \mathbb{R}$. The central assumption of kernel-based methods is that the decision function belongs to a *reproducing kernel Hilbert space* H over $\mathcal{X}$ with a certain *reproducing kernel* K. This means that H is a Hilbert space whose elements are functions $f : \mathcal{X} \to \mathbb{R}$ and there is a positive definite map $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that

$$f(p) = \langle f, K_p \rangle, \text{ for all } f \in H,\ p \in \mathcal{X},$$

where $K_p \in H$ denotes the function $K_p(q) = K(p, q)$ for all $p, q \in \mathcal{X}$. For more details about reproducing kernel Hilbert spaces, see Chapter 4 of [24].

In *ridge regression*, the decision function f_λ is given by:

$$\hat{f}_\lambda = \arg\min_{f \in H} \left(\sum_{j=1}^{n} (y_j - f(p_j))^2 + \lambda \|f\|^2 \right), \tag{13}$$

where $\|f\|$ denotes the norm of f in H and $\lambda > 0$ is a *regularization parameter*, whose role is to penalize functions that 'overfit' the training data. The solution exists and is unique, given by:

$$\hat{f}_\lambda = \sum_{j=1}^{n} c_j K_{p_j}, \tag{14}$$

where $c \in \mathbb{R}^n$ is $c = (K_D + \lambda I_n)^{-1} y$, with K_D the $n \times n$ matrix with entries $[K_D]_{ij} = K(p_i, p_j)$, I_n the identity matrix of size n and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ the vector of inputs. This is a classic result, but we include a geometrical proof of this fact in Theorem 5.1 of [12], for completeness.

From a Bayesian perspective, assume that the data arises as $y_j = f(p_j) + \epsilon_j$ where f is random, $\epsilon_1, \dots, \epsilon_n$ are independent standard Gaussian with mean 0 and variance $\lambda > 0$, and are independent of f. This yields a *prior distribution* which, combined with standard Gaussian measure on H if H is finite-dimensional, gives a *posterior distribution* proportional to $\exp(-\frac{1}{2\lambda} \sum_{j=1}^{n} (y_j - f(p_j))^2 - \frac{1}{2}\|f\|^2)$. The maximum a posteriori (MAP) estimator then yields the same function as in (14).

This approach does not work if H is infinite-dimensional, even though the solution in (14) still exists in this case. However, another way to think about the finite-dimensional case is that the data arises as a sample path of a centered Gaussian process $\{f(p) : p \in \mathcal{X}\}$, indexed by $\mathcal{X}$, perturbed by Gaussian measurement error, i.e. a centered Gaussian process $\{\epsilon(p) : p \in \mathcal{X}\}$, independent of f and with covariance $\mathrm{Cov}(\epsilon(p), \epsilon(q)) = \sqrt{\lambda}\delta_{p,q}$:

$$y(p) = f(p) + \epsilon(p).$$

Let K denote the covariance of the process f: $K(p,q) = \mathrm{Cov}(f(p), f(q))$. If we choose to predict $\hat{y}$, the output of some unknown input $p \in \mathcal{X}$, by the conditional expectation $\mathbb{E}[f(p)|y(p_1) = y_1, \ldots, y(p_n) = y_n]$, then

$$\hat{y} = \hat{f}_\lambda(p)$$

where $\hat{f}_\lambda$ is the SVM solution in (14). Although in the infinite-dimensional case we cannot have the Gaussian process $\{f(p) : p \in \mathcal{X}\}$ on H itself, we have shown that this stochastic approach does go through in infinite dimensions by working within the the abstract Wiener space framework and constructing the Gaussian process on B instead of H, and, moreover, that the prediction function is again $\hat{f}_\lambda$.

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a positive definite map and H be the RKHS over X with reproducing kernel K (such an H exists by the Moore-Aronszajn theorem - see [24]). Suppose further that $\mathcal{X}$ is a separable topological space, in which case H is a real separable Hilbert space. We complete H with respect to a measurable norm $|\cdot|$ and obtain an abstract Wiener space (H, B, μ). Then for every $p \in \mathcal{X}$ let $\tilde{K}_p$ denote $IK_p \in L^2(B, \mu)$, where $I : H \to L^2(B, \mu)$ is the map in (5) and $K_p(q) = K(p,q)$ for all $q \in \mathcal{X}$:

$$\tilde{K}_p = IK_p.$$

Given the training data $D = \{(p_j, y_j)\ 1 \leq j \leq n\} \subset \mathcal{X} \times \mathbb{R}$ let $p \in \mathcal{X}$ and $\{e_1, \ldots, e_n\}$ be an orthonormal set in H, contained in $\mathrm{span}\{K_{p_1}, \ldots, K_{p_n}, K_p\}^\perp$, and $\tilde{e}_j = Ie_j$ for all $1 \leq j \leq n$. In Theorem 4.1 of [12] we showed that:

Theorem 4.1 *With the notation above:*

$$\mathbb{E}\left[\tilde{K}_p | \tilde{K}_{p_1} + \sqrt{\lambda}\tilde{e}_1 = y_1, \ldots, \tilde{K}_{p_n} + \sqrt{\lambda}\tilde{e}_n = y_n\right] = \hat{f}_\lambda(p) \tag{15}$$

where for every $\lambda > 0$, $\hat{f}_\lambda$ is the ridge regression solution in (14). *This may be expressed as:*

$$\hat{f}_\lambda(p) = G\tilde{K}_p\left(\bigcap_{j=1}^{n}\left[\left\langle K_{p_j} + \sqrt{\lambda}e_j, \cdot\right\rangle = y_j\right]\right), \tag{16}$$

where $G\tilde{K}_p$ is the Gaussian Radon transform of $\tilde{K}_p$.

We also showed that, by essentially taking $\lambda = 0$ above, we obtain the solution in the traditional *spline* setting (see [14, 15]), where the goal is to find $f \in H$ of minimal norm such that $f(p_j) = y_j$ for all $1 \leq j \leq n$. The solution is then given by:

$$\hat{f}_0(p) = \mathbb{E}[\tilde{K}_p | \tilde{K}_{p_1} = y_1, \ldots, \tilde{K}_{p_n} = y_n] = G\tilde{K}_p\left(\bigcap_{j=1}^{n}[\langle K_{p_j}, \cdot\rangle = y_n]\right). \tag{17}$$

Finally, learning in Banach spaces is another area that has recently been of interest in the literature (see [5, 23, 27]). Such an approach would be useful in situations where the norm is not given by an inner product, and Banach spaces also offer a richer geometrical diversity. The concept of *reproducing kernel Banach space*, introduced in [27], is a natural extension of reproducing kernel Hilbert spaces. However, these are Banach spaces whose elements are functions, and the Banach space B we used so far in Theorem 4.1 does not necessarily consist of functions. We explore this idea in Section 6 of [12], where we show that if K is continuous and $\mathcal{X}$ is a separable topological space, there exists a measurable norm $|\cdot|_1$ on H such that the Banach space B obtained by completing H with respect to $|\cdot|_1$ is a space of functions on a dense countable subspace D of $\mathcal{X}$.

5 Future Directions

As noted in Becnel and Sengupta [4], the theory of infinite-dimensional Radon transforms is still in its infancy. However, the foundations for this theory within the framework of abstract Wiener spaces have been set. We hope that this will lead to many more developments in the future.

One avenue I am pursuing is to build upon the results we have already found. For instance, I am exploring stronger forms of the disintegration theorem and the inversion procedure in (11) for closed subspaces of infinite codimension. Moreover, Helgason used the support theorem for the Radon transform to prove many other interesting results, so I am also examining possible applications of our support theorem for the Gaussian Radon transform - such as the *exterior problem*, or reconstructing f from values of Gf on hyperplanes outside a closed, bounded convex subset of H.

On the machine learning side, in [12] we propose some future work where the Gaussian Radon transform could be used for a much broader class of prediction problems, such as predicting the maximum value of an unknown function over some future interval based on the training data. Specifically, we showed so far that the problem of simply predicting a future value at an unknown input $p \in \mathcal{X}$ can be solved by

$$G\tilde{K}_p(L),$$

where L is the closed subspace of H reflecting the training data, as in Theorem 4.1, (16). But suppose now that the goal was to predict the maximum value attained over some particular set $S \subset \mathcal{X}$ of "future" input values. Then

$$GF(L) \tag{18}$$

could be used instead, where

$$F = \sup_{p \in S} \tilde{K}_p.$$

Of course, there are any number of functions that could be used for F, the supremum is just an illustrative example. What is important to note about the prediction proposed in (18) is that it is not the same thing as finding a decision function and then taking the supremum of that decision function. In other words, this is not the same as

$$\sup_{p \in S} G\tilde{K}_p(L).$$

I am currently exploring this line of thought and analyzing the "accuracy" of the prediction proposed in (18).

In light of its possible applications to the highly computational field of machine learning, I am also working on finite-dimensional approximations to the Gaussian Radon transform.

I plan to continue exploring the idea of realizing B as a space of functions initiated in [12]. For instance, a "ready-made" example where B is a space of functions already exists - namely the classical Wiener space, where the Banach space is $C[0,1]$, the space of continuous functions starting at 0 on $[0,1]$. One major obstacle in mirroring the RKHS theory through reproducing kernel Banach spaces is the absence of an inner-product, which some authors circumvent by using semi-inner-products. But when a Banach space is viewed through the lens of abstract Wiener spaces, it is really the underlying Hilbert space that dictates the geometry of the Gaussian measure on B, a potentially useful idea for RKBS theory.

I have a rich list of problems and ideas to explore in this exciting new field, and I am looking forward to developing new methods and applications.

References

[1] Aleksandr Y Aravkin, Bradley M. Bell, James V. Burke, and Gianluigi Pillonetto, *The connection between Bayesian estimation of a Gaussian random field and RKHS*, Submitted to IEEE Transactions on Neural Networks and Learning Systems, available at `http://arxiv.org/pdf/1301.5288v3.pdf`.

[2] V. Bargmann, *On a Hilbert space of analytic functions and an associated integral transform*, Comm. Pure Appl. Math. **14** (1961), 187–214. MR0157250 (28 #486)

[3] Jeremy J. Becnel, *The support theorem for the Gauss-Radon transform*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **15** (2012), no. 2, 1250006, 21, DOI 10.1142/S0219025712500063. MR2957133

[4] Jeremy J. Becnel and Ambar N. Sengupta, *A support theorem for a Gaussian Radon transform in infinite dimensions*, Trans. Amer. Math. Soc. **364** (2012), no. 3, 1281–1291, DOI 10.1090/S0002-9947-2011-05365-1, available at `https://www.math.lsu.edu/~sengupta/papers/RGBecSenDec09.pdf`. MR2869177

[5] Ricky Der and Daniel Lee, *Large-Margin Classification in Banach Spaces*, Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS-07), 2007, pp. 91-98, available at `http://jmlr.csail.mit.edu/proceedings/papers/v2/der07a/der07a.pdf`.

[6] Bruce K. Driver and Maria Gordina, *Square integrable holomorphic functions on infinite-dimensional Heisenberg type groups*, Probab. Theory Related Fields **147** (2010), no. 3-4, 481–528, DOI 10.1007/s00440-009-0213-y. MR2639713 (2011k:58052)

[7] Bruce K. Driver and Brian C. Hall, *The energy representation has no non-zero fixed vectors*, (Leipzig, 1999), CMS Conf. Proc., vol. 29, Amer. Math. Soc., Providence, RI, 2000, pp. 143–155. MR1803410 (2002f:22034)

[8] Leonard Gross, *Abstract Wiener spaces*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1, Univ. California Press, Berkeley, Calif., 1967, pp. 31–42. MR0212152 (35 #3027)

[9] Brian C. Hall, *The heat operator in infinite dimensions*, Infinite dimensional stochastic analysis, QP–PQ: Quantum Probab. White Noise Anal., vol. 22, World Sci. Publ., Hackensack, NJ, 2008, pp. 161–174. MR2412888 (2010a:58044)

[10] Irina Holmes and Ambar N. Sengupta, *A Gaussian Radon transform for Banach spaces*, J. Funct. Anal. **263** (2012), no. 11, 3689–3706, DOI 10.1016/j.jfa.2012.09.005, available at `http://arxiv.org/pdf/1208.5743v2.pdf`. MR2984081

[11] Irina Holmes, *An inversion formula for the Gaussian Radon transform for Banach spaces*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **16** (2013), no. 4, available at `http://arxiv.org/pdf/1308.1392.pdf`.

[12] Irina Holmes and Ambar N. Sengupta, *The Gaussian Radon Transform and Machine Learning*, Submitted.

[13] Ferenc Huszár and Simon Lacoste-Julien, *A Kernel Approach to Tractable Bayesian Nonparametrics*, arXiv preprint (2011), available at `http://arxiv.org/pdf/1103.1761v3.pdf`.

[14] George S. Kimeldorf and Grace Wahba, *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Statist. **41** (1970), 495–502, available at `http://www.stat.wisc.edu/~wahba/ftp1/oldie/kw70bayes.pdf`. MR0254999 (40 #8206)

[15] ———, *Some Results on Tchebycheffian Spline Functions*, Journal of Mathematical Analysis and Applications **33** (1971), no. 1, 82–95, available at `http://www.stat.wisc.edu/~wahba/ftp1/oldie/kw71.pdf`.

[16] Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin, 1975. MR0461643 (57 #1628)

[17] Vochita Mihai and Ambar N. Sengupta, *The Radon-Gauss transform*, Soochow J. Math. **33** (2007), no. 3, 415–433. MR2344371 (2009a:44006)

[18] Umut Ozertem and Deniz Erdogmus, *RKHS Bayes Discriminant: A Subspace Constrained Nonlinear Feature Projection for Signal Detection*, Neural Networks, IEEE Transactions on **20** (2009), no. 7, 1195-1203, DOI 10.1109/TNN.2009.2021473.

[19] Natesh S. Pillai, Qiang Wu, Feng Liang, Sayan Mukherjee, and Robert L. Wolpert, *Characterizing the function space for Bayesian kernel models*, J. Mach. Learn. Res. **8** (2007), 1769–1797, available at `http://jmlr.org/papers/volume8/pillai07a/pillai07a.pdf`. MR2332448

[20] I. E. Segal, *Mathematical characterization of the physical vacuum for a linear Bose-Einstein field. (Foundations of the dynamics of infinite systems. III)*, Illinois J. Math. **6** (1962), 500–523. MR0143519 (26 #1075)

[21] ———, *The complex-wave representation of the free boson field*, Topics in functional analysis (essays dedicated to M. G. Kreĭn on the occasion of his 70th birthday), Adv. in Math. Suppl. Stud., vol. 3, Academic Press, New York, 1978, pp. 321–343. MR538026 (82d:81069)

[22] Peter Sollich, *Bayesian Methods for Support Vector Machines: Evidence and Predictive Class Probabilities*, Mach. Learn. **46** (2002), no. 1-3, 21–52, DOI 10.1023/A:1012489924661, available at `http://dx.doi.org/10.1023/A:1012489924661`.

[23] Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet, *Learning in Hilbert vs. Banach Spaces: A Measure Embedding Viewpoint.*, NIPS, 2011, pp. 1773-1781, available at `http://cosmal.ucsd.edu/~gert/papers/nips_rkbs_11.pdf`.

[24] Ingo Steinwart and Andreas Christmann, *Support vector machines*, Information Science and Statistics, Springer, New York, 2008. MR2450103 (2010f:62002)

[25] Daniel W. Stroock, *Abstract Wiener space, revisited*, Commun. Stoch. Anal. **2** (2008), no. 1, 145–151. MR2446996 (2010d:60095)

[26] Haizhang Zhang, Yuesheng Xu, and Jun Zhang, *Reproducing kernel Banach spaces for machine learning*, J. Mach. Learn. Res. **10** (2009), 2741–2775, DOI 10.1109/IJCNN.2009.5179093. MR2579912 (2011c:62219)

[27] ———, *Reproducing kernel Banach spaces for machine learning*, J. Mach. Learn. Res. **10** (2009), 2741–2775, DOI 10.1109/IJCNN.2009.5179093, available at `http://machinelearning.wustl.edu/mlpapers/paper_files/jmlr10_zhang09b.pdf`. MR2579912 (2011c:62219)

[28] Zhihua Zhang, Guang Dai, Donghui Wang, and Michael I. Jordan, *Bayesian Generalized Kernel Models*, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10), 2010, pp. 972-979, available at `http://www.cs.berkeley.edu/~jordan/papers/zhang-etal-aistats10.pdf`.