Shannon Sampling and Function Reconstruction from Point Values † *

Dedicated to the memory of René Thom

Steve Smale
Toyota Technological Institute at Chicago
1427 East 60th Street, Chicago, IL 60637, USA
E-mail: smale@math.berkeley.edu

Ding-Xuan Zhou
Department of Mathematics, City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong, CHINA
E-mail: mazhou@math.cityu.edu.hk

January 15, 2004

Preamble

I first met René at the 1956 well-known meeting in Topology at Mexico City. He then came to the University of Chicago where I was starting my job as instructor for the fall of 1956. He, Suzanne, Clara and I became good friends and saw much of each other for many decades, especially at IHES in Paris.

Thom's encouragement and support were important for me especially in my first years after my Ph.D. I studied his work in cobordism, singularities of maps, and transversality, gaining many insights. I also enjoyed listening to his provocations, for example his disparaging remarks on complex analysis, 19th century mathematics, and Bourbaki. There was also a stormy side in our relationship. Neither of us could hide the pain that our public conflicts over "catastrophe theory" caused.

René Thom was a great mathematician, leaving his impact on a wide part of mathematics. I will always treasure my memories of him.

Steve Smale

† The first author is partially supported by NSF grant 0325113. The second author is supported partially by the Research Grants Council of Hong Kong [Project No. CityU 103303] and by City University of Hong Kong [Project No. 7001442].

* For the Bull. of the Amer. Math. Soc.

§1. Introduction

This paper gives an account of sampling theory and interpolation, with some focus on the Shannon theorem. One goal is to deal with noise in the sampling data, from the point of view of exponential probability estimates. Our quantitative estimates give some guide as to how much resampling or regularization is required to balance noise in the form of a variance. A measure of the richness of the data is key in this development.

The theory evolves in a universe which is a Hilbert space of real valued functions on a (an "input") space X. In the Shannon case X is the space of real numbers. Other examples for X include a rectangle in the plane (image processing), a graph as in theoretical computer science, or a high dimensional space as in learning theory.

Our first generalization of the Shannon theorem centers around the case of rich data and the use of a Hilbert space and a kernel function, reminiscent of reproducing kernel Hilbert spaces derived from a Mercer kernel. Subsequently we see how poor data and general Hilbert function spaces fit into our analysis.

An objective is to integrate the theory with fast algorithms which work well in the presence of noise. Our main results are new general error estimates.

We have been inspired by the disciplines of learning theory, regression analysis, approximation theory, inverse problems, signal processing, and hope that in return this work can give some new insights to these subjects.

§2. Motivating Examples

To describe the general reconstruction of functions from their point values, we give some simple motivating examples.

Example 1 (exact polynomial interpolation) (a baby example). *Consider polynomials* $p_t : \mathbb{R} \to \mathbb{R}$, *for* $t \in \bar{t} := \{0, 1, \ldots, d\}$ *with* $p_t(x) = x^t$. *The polynomial interpolation problem is to find a polynomial* $f = \sum_{t \in \bar{t}} a_t p_t$ *of degree* d *such that* $f(x_i) = y_i$ *for* $i = 1, \ldots, d+1$. *Here* $(x_i, y_i)_{i=1}^{d+1}$ *is the data. The situation yields a system of equations:* $L(a_t)_{t \in \bar{t}} = (y_i)_{i=1}^{d+1}$ *with* $L = (p_t(x_i))_{i=1,\ldots,d+1, t \in \bar{t}}$ *being a* $(d+1) \times (d+1)$ *matrix. When* $\{x_i\}$ *are distinct, this system has a unique solution* $a_0, a_1, \ldots, a_d$, *which solves the problem.*

If we denote $\overline{x} = \{x_i\}_{i=1}^{d+1}$, then the "data" is given by the function on $\overline{x}$. Here $|\overline{x}| = |\overline{t}|$. Certainly the choice of p_t is quite naive. In Section 10 this kind of problem is studied.

The next two examples are from image processing. The first is borrowed from [7].

Example 2 (inpainting). *Consider a black white photograph as a function g from $\overline{t}$ to $[0,1]$ where $\overline{t}$ is a square of pixels (e.g. 512 by 512) and $g(t)$ represents a shade of grey of pixel t. Now suppose that the photograph has been partly masked as by some spilled ink or writing over it destroying g on the mask say $\hat{t}$ and leaving our function intact on $\overline{x} = \overline{t} \setminus \hat{t}$. The problem is to recover an approximation to g from its restriction to $\overline{x}$. Here the input or data is $(x, g(x))$ for $x \in \overline{x}$. Note that $|\overline{x}| < |\overline{t}|$. This is a case of what we call later "poor data".*

Example 3 (image compression). *Here $\overline{t}$ is a coarse pixel set and $\overline{x}$ is a fine pixel set. The original picture is represented by a function from $\overline{x}$ to the interval as in Example 2. The problem is to find a worse but reasonable representation (with small error) as a function from $\overline{t}$. The efficiency of a compression scheme is measured by the ratio $|\overline{x}|/|\overline{t}|$ (as large as possible, representing the richness of the data) and the error (within a threshold).*

§3. Learning and Sampling

The classical *Whittaker-Shannon-Nyquist Sampling Theorem* or simply Shannon Theorem gives conditions on a function on $\mathbb{R}$ (band-limited with band π) so that it can be reconstructed from its sampling values at integer points:

Theorem. *Let $\phi(x) = \frac{\sin \pi x}{\pi x}$ and $\phi_t(x) = \phi(x-t)$. If a function $f \in L^2(\mathbb{R})$ has its Fourier transform supported on $[-\pi, \pi]$, then*

$$f = \sum_{t \in \mathbb{Z}} f(t)\phi_t.$$

See [2, 31] for some background and some generalizations.

We proceed to state our own generalization.

Suppose X is a closed subset of $\mathbb{R}^n$ (a complete metric space is sufficient) and $\overline{t} \subset X$ is a discrete subset. In the Shannon special case, $X = \mathbb{R}, \overline{t} = \mathbb{Z}$. Another important case is when X is compact and (hence) $\overline{t}$ is finite.

Next consider a continuous symmetric map (a "kernel") $K : X \times X \to \mathbb{R}$ and use it to define a matrix (possibly infinite) $K_{\bar{t},\bar{t}} : \ell^2(\bar{t}) \to \ell^2(\bar{t})$ as

$$\left(K_{\bar{t},\bar{t}}a\right)_s = \sum_{t\in\bar{t}} K(s,t)a_t, \qquad s \in \bar{t}, a \in \ell^2(\bar{t}).$$

Here $\ell^2(\bar{t})$ is the set of sequences $a = (a_t)_{t\in\bar{t}} : \bar{t} \to \mathbb{R}$ with $< a, b >= \sum_{t\in\bar{t}} a_t b_t$ defining an inner product. For $t \in \bar{t}$, set $K_t : X \to \mathbb{R}$ to be the continuous function on X given by $K_t(x) = K(t, x)$. Unless said otherwise, we always assume the following.

Standing Hypothesis 1. *$K_{\bar{t},\bar{t}}$ is well-defined, bounded, and positive with bounded inverse $K^{-1}_{\bar{t},\bar{t}}$.*

In the Shannon case $K(t, s) = \phi(t-s)$, and it is seen that $K_{\bar{t},\bar{t}}$ is the identity, because $\phi(j) = 0$ for $j \in \mathbb{Z} \setminus \{0\}$ and $\phi(0) = \lim_{x\to 0} \phi(x) = 1$.

For Example 1, we can take $X = \mathbb{R}$, $\bar{t} = \{0, 1, \ldots, d\}$, and $K(t, s) = (1+t\cdot s)^d$. Then for $c \in \ell^2(\bar{t})$, there holds $< K_{\bar{t},\bar{t}}c, c >_{\ell^2(\bar{t})}= \sum_{k=0}^{d} \binom{d}{k} \left(\sum_{t\in\bar{t}} c_t t^k\right)^2$. Since the Vandermonde determinant $\det(t^k)_{t\in\bar{t},k=0,1,\ldots,d}$ is nonzero, Standing Hypothesis 1 is satisfied.

Next define a Hilbert space $\mathcal{H}_{K,\bar{t}}$ as follows. Consider the linear space of finite linear combinations of $K_t, t \in \bar{t}$, i.e., $\sum_{t\in\bar{t}} a_t K_t$ where only a finite number of a_t are nonzero. An inner product on this space is defined (from the positivity of $K_{\bar{t},\bar{t}}$) by linear extension from

$$< K_t, K_s >_K= K(t, s). \tag{3.1}$$

One takes the completion to obtain $\mathcal{H}_{K,\bar{t}}$.

In the Shannon case, it can be shown (see Example 4 in Section 8) that $\mathcal{H}_{K,\bar{t}}$ is the space described, i.e., $f \in L^2(\mathbb{R})$ with $\mathrm{supp}\hat{f} \subseteq [-\pi, \pi]$. Here $\hat{f}$ denotes the Fourier transform of f. It is defined for an integrable function on $\mathbb{R}^n$ as $\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-i\xi\cdot x}dx$, and can be extended naturally to the space $L^2(\mathbb{R}^n)$.

In Example 1, with the kernel $K(t, s) = (1 + t \cdot s)^d$, we find that $\mathcal{H}_{K,\bar{t}}$ is exactly the space of polynomials of degree d.

If we define $\ell^2_K(\bar{t})$ as the Hilbert space consisting of sequences in $\ell^2(\bar{t})$ with the inner product $< a, b >_{\ell^2_K(\bar{t})}:=< K_{\bar{t},\bar{t}}a, b >_{\ell^2(\bar{t})}$, then the natural map from $\ell^2_K(\bar{t})$ to $\mathcal{H}_{K,\bar{t}}$, given by $a \to \sum_{t\in\bar{t}} a_t K_t$, is an isomorphism. Note that $\ell^2_K(\bar{t})$ does not depend on X, just $\bar{t}$ and K restricted to $\bar{t}\times\bar{t}$. Hence it discretizes the setting. Also, standing hypothesis 1 tells us that $\ell^2_K(\bar{t})$ is isomorphic to $\ell^2(\bar{t})$ under the isomorphism: $a \to K^{1/2}_{\bar{t},\bar{t}} a$.

If $\bar{s}\subset\bar{t}$ replaces $\bar{t}$, then the important invariants $\|K_{\bar{t},\bar{t}}\|$ and $\|K^{-1}_{\bar{t},\bar{t}}\|$ improve. That is, $\|K_{\bar{s},\bar{s}}\| \le \|K_{\bar{t},\bar{t}}\|$ and $\|K^{-1}_{\bar{s},\bar{s}}\| \le \|K^{-1}_{\bar{t},\bar{t}}\|$. Thus, if K is restricted to $X'\subset X$ and $\bar{s}=\bar{t}\cap X'$, then standing hypothesis 1 remains true.

If K is a Mercer kernel and $\mathcal{H}_K$ the corresponding reproducing kernel Hilbert space [3], then $\mathcal{H}_{K,\bar{t}}$ is the closed subspace generated by $\{K_t, t\in\bar{t}\}$ (with the induced inner product). This gives a class of spaces $\mathcal{H}_{K,\bar{t}}$ satisfying standing hypothesis 1 (besides the space generated by ϕ in the Shannon theorem). One such example is a Gaussian kernel $K(x,y)=e^{-|x-y|^2/\sigma^2}$ on any closed subset X of $\mathbb{R}^n$. See Section 8, and more examples and background in [8].

So far, we have a space $\mathcal{H}_{K,\bar{t}}$ which plays the role of a "representation space" in the Shannon theory. We now pass to the sampling side which we separate out. Moreover, noise is introduced into our model in this sampling, represented by a Borel measure ρ on $X\times\mathbb{R}$.

Let ρ_X be the marginal measure induced by ρ on X, i.e., the measure on X defined by $\rho_X(S)=\rho(\pi^{-1}(S))$ where $\pi: X\times\mathbb{R}\to X$ is the projection. It defines a space $L^2_{\rho_X}$ on X with L^2 norm $\|f\|=\|f\|_{L^2_{\rho_X}} := \left(\int_X |f(x)|^2 d\rho_X\right)^{1/2}$. It is not assumed that ρ_X is a probability measure as in the special case of learning theory. In fact in the Shannon case it is the Lebesgue measure.

The set for the sampling is a discrete set $\bar{x}\subset X$. The set $\bar{x}$ may be determined as in a net (Shannon, with $\bar{x}=\mathbb{Z}$) or have come from a random sample as in [8] or [4]. For $x\in X$, we denote the variance of the conditional measures ρ_x of ρ as σ^2_x. We assume that the conditional measures $\rho_x (x\in X)$ of ρ satisfy

Preliminary Version of Special Assumption. *For each $x\in X$, ρ_x is a probability measure with zero mean supported on $[-M_x, M_x]$ with $\mathcal{B} := \left(\sum_{x\in\bar{x}} M^2_x\right)^{1/2} < \infty$.*

To study the relationship between the discrete sets $\bar{t}$ and $\bar{x}$, we define the linear operator $K_{\overline{x},\overline{t}} : \ell^2(\overline{t}) \to \ell^2(\overline{x})$ and its adjoint $K_{\overline{t},\overline{x}} : \ell^2(\overline{x}) \to \ell^2(\overline{t})$ by the matrix

$$\left(K_{\overline{x},\overline{t}}a\right)_x = \sum_{t\in\overline{t}} K(x,t)a_t. \tag{3.2}$$

Standing Hypothesis 2. *$K_{\overline{x},\overline{t}}$ (and hence $K_{\overline{t},\overline{x}}$) is well-defined and bounded.*

The sampled values $y \in \ell^2(\overline{x})$ will have the form:

$$\text{For } f^* \in \mathcal{H}_{K,\overline{t}}, \text{ and each } x \in \overline{x},\ y_x = f^*(x) + \eta_x, \text{ where } \eta_x \text{ is drawn from } \rho_x. \tag{3.3}$$

Special Assumption implies that $\{\eta_x\} \in \ell^2(\overline{x})$ and $\|\{\eta_x\}\|_{\ell^2(\overline{x})} \le \mathcal{B} < \infty$.

Define the **sampling operator** $S_{\overline{x}} : \mathcal{H}_{K,\overline{t}} \to \ell^2(\overline{x})$ by $S_{\overline{x}}f = (f(x))_{x\in\overline{x}}$. That is, for a function f from $\mathcal{H}_{K,\overline{t}}$, $S_{\overline{x}}f$ is the restriction of f to $\overline{x} : f|_{\overline{x}}$. Then for $f = \sum_{t\in\overline{t}} c_tK_t$, we have $S_{\overline{x}}f = K_{\overline{x},\overline{t}}c$. It follows that $\sum_{x\in\overline{x}} f^*(x)^2 = \|S_{\overline{x}}f^*\|^2_{\ell^2(\overline{x})}$ can be bounded by $\|K_{\overline{x},\overline{t}}\|^2\|f^*\|^2_K/\|K^{-1}_{\overline{t},\overline{t}}\|$, and is finite according to (3.3), hence $y \in \ell^2(\overline{x})$.

In the Shannon case, $\overline{x} = \overline{t}$, ρ_x is trivial, so $\eta_x = 0$ for all $x \in \mathbb{Z}$.

Now our *sampling problem* is:

Reconstruct f^* (or an approximation of f^*) from $y \in \ell^2(\overline{x})$.

Towards its study, consider the minimization problem

$$\arg\min_{f\in\mathcal{H}_{K,\overline{t}}} \sum_{x\in\overline{x}} \big(f(x) - y_x\big)^2. \tag{3.4}$$

The solution of (3.4) is expressed using $K_{\overline{t},\overline{x}}$ and $K_{\overline{x},\overline{t}}$.

Definition 1. *We say that $\overline{x}$ provides* **rich data** *(with respect to $\overline{t}$) if*

$$\lambda_{\overline{x}} := \inf_{v\in\ell^2(\overline{t})} \|K_{\overline{x},\overline{t}}v\|_{\ell^2(\overline{x})}/\|v\|_{\ell^2(\overline{t})} \tag{3.5}$$

is positive. It provides **poor data** *if $\lambda_{\overline{x}} = 0$.*

One can easily see that $\overline{x}$ provides rich data if and only if the operator $K_{\overline{t},\overline{x}}K_{\overline{x},\overline{t}}$ has a bounded inverse, that is, its smallest eigenvalue $(\lambda_{\overline{x}})^2$ is positive.

Note that if $\overline{x} \subset \overline{\overline{x}}$, then $\lambda_{\overline{x}} \le \lambda_{\overline{\overline{x}}}$.

Our *generalized Shannon Sampling Theorem* (for rich data) can be stated as follows (the proof will be given in Section 7). Define the **variance** of the system $(\rho, \overline{x}, \overline{t}, K)$ as

$$\sigma^2 := \sum_{x\in\overline{x}} \sigma_x^2 \sum_{t\in\overline{t}} K(t,x)^2 = \sum_{x\in\overline{x}} \sigma_x^2 \|K_{\overline{t},\overline{x}} e_x\|^2_{\ell^2(\overline{t})}, \tag{3.6}$$

where e_x is the delta sequence supported at x. It represents how the variance on $\overline{x}$ is transferred to $\overline{t}$ by the operator $K_{\overline{t},\overline{x}} : \ell^2(\overline{x}) \to \ell^2(\overline{t})$. Standing hypothesis 2 and special assumption tell us that σ^2 is finite.

Theorem 1. *Assume $f^* \in \mathcal{H}_{K,\overline{t}}$ with $X, K, \overline{t}, \rho$ as above, y as in (3.3) together with the special assumption, and that $\overline{x}$ provides rich data. Then the problem (3.4) can be solved:*

$$f_{\mathbf{z}} = \sum_{t\in\overline{t}} a_t K_t, \quad a = Ly \quad \text{and} \quad L = \left(K_{\overline{t},\overline{x}} K_{\overline{x},\overline{t}}\right)^{-1} K_{\overline{t},\overline{x}},$$

and its solution approximately reconstructs f^* from its values at $\overline{x}$ in the following sense.

For every $\varepsilon > 0$, $\|f_{\mathbf{z}} - f^*\|_K^2 \le \kappa\sigma^2 + \varepsilon$ with probability $1-\delta$ where
$$\kappa := \frac{\|K_{\overline{t},\overline{t}}\|}{\lambda_{\overline{x}}^4}, \qquad \delta = \exp\left\{-\frac{\varepsilon\lambda_{\overline{x}}^2}{2\|K_{\overline{t},\overline{t}}\|\mathcal{B}^2} \log\left(1 + \frac{\varepsilon}{\kappa\sigma^2}\right)\right\}.$$

Remark. *Since $\overline{x}$ provides rich data, we see from Definition 1 that the operator $K_{\overline{x},\overline{t}}$ is injective. The operator L defined in Theorem 1 is exactly the Moore-Penrose inverse of $K_{\overline{x},\overline{t}}$. See e.g. [11, 13].*

When the richness of the data increases such that $\lambda_{\overline{x}} \to \infty$ (see Proposition 1 below), we have $\kappa \to 0$. If moreover $\mathcal{B}^2/\lambda_{\overline{x}}^2$ is kept bounded, then from Theorem 1 we see that for the error bound $\kappa\sigma^2 + \varepsilon$ with any $\varepsilon > 0$ the confidence tends to 1. This yields the convergence with confidence if $\sigma^2/\lambda_{\overline{x}}^4 \to 0$. Also, we find that for any $\lambda_{\overline{x}}$ when the variance vanishes, $f_{\mathbf{z}} = f^*$ with probability one by taking $\sigma^2 \to 0$ in Theorem 1; thus we cover the classical Shannon theorem.

When the data is resampled k times over $\overline{x}$, the richness increases to $\sqrt{k}\lambda_{\overline{x}}$, $\kappa\sigma^2$ is reduced to $\kappa\sigma^2/k$, while the bound $\mathcal{B}^2$ of the system becomes $k\mathcal{B}^2$. Then $c := \frac{\varepsilon\lambda_{\overline{x}}^2}{2\|K_{\overline{t},\overline{t}}\|\mathcal{B}^2}$ is unchanged. We see from Theorem 1 that for the better error bound $\kappa\sigma^2/k + \varepsilon$ with the same ε, the confidence $1 - (1 + \varepsilon/(\kappa\sigma^2))^{-c}$ is improved to $1 - (1 + k\varepsilon/(\kappa\sigma^2))^{-c}$.

Corollary 1. *Under the assumption of Theorem 1, if the data is resampled k times over $\overline{x}$, then for every $\varepsilon > 0$, $\|f_{\mathbf{z}} - f^*\|_K^2 \leq \kappa\sigma^2/k + \varepsilon$ with probability $1 - (1 + k\varepsilon/(\kappa\sigma^2))^{-c}$ while the probability given in Theorem 1 is $1 - (1 + \varepsilon/(\kappa\sigma^2))^{-c}$.*

Corollary 1 convinces us that resampling improves the error when one takes the same probability as in Theorem 1. See also Proposition 3 in Section 6.

The constant κ is the infimum of error bounds for positive probability in Theorem 1. This threshold quantity relates the key variables. The case of exact interpolation corresponds to $|\overline{t}| = |\overline{x}|, \lambda_{\overline{x}} > 0$.

Note that error bounds less than κ may be studied by the introduction of a regularization parameter $\gamma > 0$ (see below).

Theorem 1 will be extended to include the case of poor data.

The regularized version of the problem (3.4) takes the form

$$\tilde{f}_{\mathbf{z},\gamma} := \arg \min_{f \in \mathcal{H}_{K,\overline{t}}} \sum_{x \in \overline{x}} \big(f(x) - y_x\big)^2 + \gamma \|f\|_K^2, \tag{3.7}$$

where $\gamma \geq 0$ and the case $\gamma = 0$ includes the setting of Theorem 1.

As in Theorem 1, the problem (3.7) can be solved by means of a linear operator: $\tilde{f}_{\mathbf{z},\gamma} = \sum_{t \in \overline{t}} a_t K_t$, where $a = Ly$ and $L = \big(K_{\overline{t},\overline{x}} K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\big)^{-1} K_{\overline{t},\overline{x}}$.

We expand the setting a bit by introducing a weighting w on $\overline{x}$. A weighting is necessary to expand beyond the special case of $\overline{x}$ defined by a uniform grid on X.

So we let $w := \{w_x\}_{x \in \overline{x}}$ be a weighting with $w_x > 0$. One example is to take w as the ρ_X-volume of the Voronoi [28] associated with $\overline{x}$. Another example is $w \equiv 1$ or if $|\overline{x}| = m < \infty, w \equiv \frac{1}{m}$.

We require $\|w\|_\infty = \sup_{x \in \overline{x}} w_x < \infty$. Denote $D_w : \ell^2(\overline{x}) \to \ell^2(\overline{x})$ as the diagonal matrix (multiplication operator on $\ell^2(\overline{x})$) with main diagonal entries $\{w_x\}_{x \in \overline{x}}$. Then $\|D_w\| \leq \|w\|_\infty$. The square root $D_w^{\frac{1}{2}}$ is the diagonal matrix with main diagonal entries $\{\sqrt{w_x}\}_{x \in \overline{x}}$.

Definition 2. *The regularization scheme for the sampling problem in the space $\mathcal{H}_{K,\overline{t}}$ takes the form:*

$$f_{\mathbf{z},\gamma} := \arg \min_{f \in \mathcal{H}_{K,\overline{t}}} \left\{ \sum_{x \in \overline{x}} w_x \big(f(x) - y_x\big)^2 + \gamma \|f\|_K^2 \right\}. \tag{3.8}$$

Theorem 2. *Assume $f^* \in \mathcal{H}_{K,\bar{t}}$ and the standing hypotheses with $X, K, \bar{t}, \rho$ as above, y as in (3.3). Suppose $K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}$ is invertible. Define L to be the linear operator $L = \left(K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}\right)^{-1} K_{\bar{t},\bar{x}} D_w$. Then the problem (3.8) has a unique solution:*

$$f_{\mathbf{z},\gamma} = \sum_{t \in \bar{t}} \big(Ly\big)_t K_t. \tag{3.9}$$

The corresponding errors will be analyzed in the next sections (Theorems 4 and 5). The error analysis will generalize Theorem 1 with general bound M, weighting w and $\gamma \geq 0$. It also extends to the poor data setting. Observe that under the standing hypotheses, $K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}$ is invertible, if $\gamma > 0$ or $\lambda_{\bar{x}} > 0$.

Consider the case when K is a "convolution kernel" $K(s,u) = \psi(s-u)$. Let $\psi \in L^2(\mathbb{R}^n)$ whose Fourier transform $\hat{\psi}$ satisfies

$$\hat{\psi}(\xi) \geq c_0 > 0, \qquad \forall \xi \in [-\pi, \pi]^n \tag{3.10}$$

and the following decay condition for some $C_0 > 0, \alpha > n$:

$$0 \leq \hat{\psi}(\xi) \leq C_0 (1 + |\xi|)^{-\alpha} \qquad \forall \xi \in \mathbb{R}^n. \tag{3.11}$$

Definition 3. *We say that $\bar{x}$ is Δ-dense in X if for each $y \in X$ there is some $x \in \bar{x}$ satisfying $\|x - y\|_{\ell^\infty(\mathbb{R}^n)} \leq \Delta$.*

Proposition 1. *Let $X = \mathbb{R}^n$, $\bar{t} = \mathbb{Z}^n$, $K(s,u) = \psi(s-u)$ with an even function ψ (i.e. $\psi(u) = \psi(-u)$) satisfying (3.10) and (3.11). If $0 < \mathcal{L} < 1/4$ and $\bar{x}$ is Δ-dense for some $0 < \Delta \leq \tau$, then*

$$\lambda_{\bar{x}} \geq \frac{(\cos \mathcal{L}\pi - \sin \mathcal{L}\pi)^n c_0}{2^{1+n/2}} \mathcal{L}^{n/2} \Delta^{-n/2}.$$

Here τ is a constant independent of Δ and Proposition 1 is a consequence of Corollary 6 below where an explicit expression for τ (depending on $\mathcal{L}$) will be given.

Recall that the Shannon case corresponds to the choice $\psi = \phi$ with $n = 1, c_0 = 1, C_0 = (1+\pi)^6, \alpha = 6$, and $\|\cdot\|_K = \|\cdot\|_{L^2(\mathbb{R})}$. Then $\sum_{t \in \bar{t}} K(t,x)^2 \equiv 1$ and $\sigma^2 = \sum_{x \in \bar{x}} \sigma_x^2$. Combining Theorem 1 with Proposition 1 for $\mathcal{L} = 1/5$ (and the constant τ given in Corollary 6) yields the following.

Corollary 2. *Let $X = \mathbb{R}$, $\bar{t} = \mathbb{Z}$, $K(s,u) = \phi(s-u)$ where ϕ is the sinc function given in the Shannon Theorem. If $\bar{x}$ is Δ-dense for some $0 < \Delta \le 1/500$ and ρ satisfies special assumption, then for any $\varepsilon > 0$, the function $f_{\mathbf{z}}$ given in Theorem 1 satisfies*

$$Prob\Big\{ \|f_{\mathbf{z}} - f^*\|^2_{L^2(\mathbb{R})} \le 20^4\Delta^2\sigma^2 + \varepsilon \Big\} \ge 1 - \exp\Big\{ -\frac{\varepsilon}{800\Delta\mathcal{B}^2} \log\big(1 + \frac{\varepsilon}{20^4\Delta^2\sigma^2}\big) \Big\}.$$

If the data becomes dense such that $\Delta \to 0$ but $\Delta\mathcal{B}^2$ is kept bounded (e.g. $\bar{x}$ is quasi-uniform), then $\Delta^2\sigma^2 \to 0$ and Corollary 2 yields the convergence of $f_{\mathbf{z}}$ to f^* with confidence.

Notice that $\bar{x} \ne \bar{t}$ in general: $f^* \in \mathcal{H}_{K,\bar{t}}$, while $\bar{x}$ stands for the sampling points which can be much denser than $\bar{t}$.

In the above discussion, where $f^* \in \mathcal{H}_{K,\bar{t}}$, one may take either of two points of view. Start with ρ and let $f^* = f_\rho$ be the regression function as done in learning theory [27, 29, 14, 8, 18], or take a primary f^* as in sampling theory [2, 15] and hypothesize ρ as above.

Our learning process in Definition 2 is an example of a regularization scheme. Regularization schemes are often used for solving problems with ill-posed coefficient matrices or operators such as numerical solutions of integral and differential equations, stochastic ill-posed problems with operator equations, and empirical risk minimization problem for traditional learning. See *e.g.* [25, 16, 13].

Some preliminary estimates on $\lambda_{\bar{x}}$ will be provided in Sections 8 and 9. But we hope to give more satisfactory results in a subsequent work.

The authors would like to thank Akram Aldroubi for his conversations on the question of relating learning theory to sampling.

§4. The Algorithm

We give the proof of Theorem 2.

For $f : X \to \mathbb{R}$, it is natural to introduce an "error function"

$$\mathcal{E}(f) = \int_Z \big(f(x) - y\big)^2 d\rho. \tag{4.1}$$

For the empirical counterpart of $\mathcal{E}$, let $\mathbf{z} = (x, y_x)_{x\in\bar{x}}$ be a sample, so that x is defined by $\bar{x}$ and y_x is drawn at random from $f^*(x) + \rho_x$ as in (3.3). Then the **empirical error** is

$$\mathcal{E}_{\mathbf{z}}(f) = \sum_{x\in\bar{x}} w_x \big(f(x) - y_x\big)^2. \tag{4.2}$$

With the empirical error $\mathcal{E}_{\mathbf{z}}(f)$, our learning scheme (3.8) can be written as

$$f_{\mathbf{z},\gamma} := \arg \min_{f \in \mathcal{H}_{K,\bar{t}}} \Big\{ \mathcal{E}_{\mathbf{z}}(f) + \gamma \|f\|_K^2 \Big\}. \tag{4.3}$$

We show how to solve the minimization problem (4.3) or (3.8) by a linear algorithm.

Proof of Theorem 2. Consider the quadratic form

$$Q(c) := \mathcal{E}_{\mathbf{z}}(\sum_{t \in \bar{t}} c_t K_t) + \gamma \| \sum_{t \in \bar{t}} c_t K_t \|_K^2, \qquad c \in \ell^2(\bar{t}).$$

A simple computation yields

$$Q(c) =< \big(K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}\big) c, c >_{\ell^2(\bar{t})} - 2 < D_w K_{\bar{x},\bar{t}} c, y >_{\ell^2(\bar{x})} + < D_w y, y >_{\ell^2(\bar{x})} .$$

Taking the functional derivative as in [19] tells us that if c is a minimizer of Q in $\ell^2(\bar{t})$ then it satisfies

$$\big(K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}\big) c = K_{\bar{t},\bar{x}} D_w y, \qquad c \in \ell^2(\bar{t}). \tag{4.4}$$

By our assumption, $K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}$ is invertible, the system (4.4) has a unique solution: $c = \big(K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}\big)^{-1} K_{\bar{t},\bar{x}} D_w y$. It yields the unique minimizer of Q which represents the unique minimizer $f_{\mathbf{z},\gamma}$ of the functional $\mathcal{E}_{\mathbf{z}}(f) + \gamma \|f\|_K^2$ in $\mathcal{H}_{K,\bar{t}}$. □

Remark. *Standing hypothesis 1 can be weakened for the purpose of Theorem 2: the first case is $\gamma > 0$; the second case is $\bar{t} = \bar{x}$ and $\gamma = 0$. In both cases, the scheme (4.3) has a solution $f_{\mathbf{z},\gamma}$ lying in*

$$\mathcal{H}^o_{K,\bar{t}} := \{\sum_{t \in \bar{t}} c_t K_t : c \in \ell^2(\bar{t})\} \subseteq \mathcal{H}_{K,\bar{t}}$$

if and only if the system (4.4) is solvable. When the solvability of (4.4) holds, the solution in $\mathcal{H}^o_{K,\bar{t}}$ is unique and given by $f_{\mathbf{z},\gamma} = \sum_{t \in \bar{t}} c_t K_t$, independent of the choice of the solution c to (4.4). In fact, if c and d are both solutions to (4.4), then $\sum_{t \in \bar{t}} c_t K_t = \sum_{t \in \bar{t}} d_t K_t$: $K_{\bar{t},\bar{t}}(c - d) = 0$ for either $\gamma > 0$ or $\bar{t} = \bar{x}$.

In the following three sections we shall estimate the error $\|f_{\mathbf{z},\gamma} - f^*\|$.

§5. Probability Inequalities

In the following theorem, $m \in \mathbb{N}$ or $m = \infty$. When $m = \infty$, the product probability measure on the product space $\mathbb{R}^m$ can be defined in any sense such as the one defined by means of the Tikhonov topology, see *e.g.* [21].

Theorem 3. *Let $\{\xi_j\}_{j=1}^m$ be independent random variables on $\mathbb{R}$ with variances $\{\sigma_j^2\}_j$, and $w_j \geq 0$ with $\|w\|_\infty < \infty$. If $\sigma_w^2 := \sum_{j=1}^m w_j \sigma_j^2 < \infty$, and for each j there holds $|\xi_j - E(\xi_j)| \leq M$ almost everywhere, then for every $\varepsilon > 0$ the probability in the product space $\mathbb{R}^m$ satisfies*

$$Prob\left\{\left|\sum_{j=1}^m w_j \left[\xi_j - E(\xi_j)\right]\right| > \varepsilon\right\} \leq 2\exp\left\{-\frac{\varepsilon}{2\|w\|_\infty M}\log\left(1+\frac{M\varepsilon}{\sigma_w^2}\right)\right\}.$$

Corollary 3. *If $m < \infty$ and $\xi_1, \xi_2, \ldots, \xi_m$ are i.i.d. random variables with expected value μ, variance σ^2 satisfying $|\xi - \mu| \leq M$, then*

$$Prob\left\{\left|\frac{1}{m}\sum_{j=1}^m \xi_j - \mu\right| > \varepsilon\right\} \leq 2\exp\left\{-\frac{m\varepsilon}{2M}\log\left(1+\frac{M\varepsilon}{\sigma^2}\right)\right\}. \tag{5.1}$$

Proof of Theorem 3. Without loss of generality, we assume $E(\xi_j) = 0$. Then the variance of ξ_j is $\sigma_j^2 = E\big(\xi_j^2\big)$.

First we assume $m < \infty$. It is sufficient for us to prove the one-side inequality:

$$I := \text{Prob}\left\{\sum_{j=1}^m w_j \xi_j > \varepsilon\right\} \leq \exp\left\{-\frac{\varepsilon}{2\|w\|_\infty M}\log\left(1+\frac{M\varepsilon}{\sigma_w^2}\right)\right\}. \tag{5.2}$$

Let c be an arbitrary positive constant which will be determined later. Then by the independence,

$$\begin{aligned} I &= \text{Prob}\left\{\exp\left\{\sum_{j=1}^m c w_j \xi_j\right\} > e^{c\varepsilon}\right\} \\ &\leq e^{-c\varepsilon} E\left(\exp\left\{\sum_{j=1}^m c w_j \xi_j\right\}\right) = e^{-c\varepsilon}\Pi_{j=1}^m E\left(\exp\left\{c w_j \xi_j\right\}\right). \end{aligned}$$

Since $|\xi_j| \leq M$ almost everywhere, we have

$$E\left(\exp\left\{c w_j \xi_j\right\}\right) = 1 + \sum_{\ell=2}^{+\infty} \frac{c^\ell w_j^\ell E\big(\xi_j^\ell\big)}{\ell!} \leq 1 + \sum_{\ell=2}^{+\infty} \frac{c^\ell w_j^\ell M^{\ell-2}\sigma_j^2}{\ell!}.$$

As $w_j \le \|w\|_\infty$ and $1+t \le e^t$, there holds

$$E\bigg(\exp\Big\{cw_j\xi_j\Big\}\bigg) \le \exp\bigg\{\sum_{\ell=2}^{+\infty} \frac{c^\ell \|w\|_\infty^{\ell-1} M^{\ell-2} w_j \sigma_j^2}{\ell!}\bigg\}$$
$$= \exp\bigg\{\frac{e^{c\|w\|_\infty M} - 1 - c\|w\|_\infty M}{\|w\|_\infty M^2} w_j \sigma_j^2\bigg\}.$$

It follows that

$$I \le \exp\bigg\{-c\varepsilon + \frac{e^{c\|w\|_\infty M} - 1 - c\|w\|_\infty M}{\|w\|_\infty M^2} \sum_{j=1}^m w_j \sigma_j^2\bigg\}.$$

Now choose the constant c to be the minimizer of the bound on the above right hand side:

$$c = \frac{1}{\|w\|_\infty M} \log\bigg(1 + \frac{M\varepsilon}{\sum_{i=1}^m w_i \sigma_i^2}\bigg).$$

That is, $e^{c\|w\|_\infty M} - 1 = \frac{M\varepsilon}{\sigma_w^2}$. With this choice,

$$I \le \exp\bigg\{-\frac{\varepsilon}{\|w\|_\infty M}\bigg\{\bigg(1 + \frac{\sigma_w^2}{M\varepsilon}\bigg)\log\bigg(1 + \frac{M\varepsilon}{\sigma_w^2}\bigg) - 1\bigg\}\bigg\}. \tag{5.3}$$

If we set a function $g(\lambda)$ as

$$g(\lambda) := (1+\lambda)\log(1+\lambda) - \lambda, \qquad \lambda \ge 0,$$

then

$$I \le \exp\bigg\{-\frac{\sigma_w^2}{\|\overline{w}\|_\infty M^2} g\bigg(\frac{M\varepsilon}{\sigma_w^2}\bigg)\bigg\}. \tag{5.4}$$

We claim that

$$g(\lambda) \ge \frac{\lambda}{2}\log(1+\lambda), \qquad \forall \lambda \ge 0.$$

To see this, define a C^2 function on $\mathbb{R}_+$ as

$$f(\lambda) := 2\log(1+\lambda) - 2\lambda + \lambda\log(1+\lambda), \qquad \lambda \ge 0.$$

We can see that $f(0) = 0, f'(0) = 0$, and $f''(\lambda) = \lambda(1+\lambda)^{-2} \ge 0$ for $\lambda \ge 0$. Hence $f(\lambda) \ge 0$ and

$$\log(1+\lambda) - \lambda \ge -\frac{1}{2}\lambda\log(1+\lambda), \qquad \forall \lambda \ge 0.$$

It follows that

$$g(\lambda) = \lambda \log(1+\lambda) + \log(1+\lambda) - \lambda \geq \frac{\lambda}{2}\log(1+\lambda), \qquad \forall \lambda > 0.$$

This verifies our claim.

The desired one-side inequality (5.2) follows from this claim and the bound for I in terms of g.

When $m = \infty$, the independence and the convergence of the series $\sum_{j=1}^{\infty} w_j \sigma_j^2$ tells us that$\{S_k := \sum_{j=1}^{k} w_j \xi_j\}_{k=1}^{\infty}$ is a Cauchy sequence in L^2:

$$\|S_k - S_\ell\|_{L^2} = \big(E(S_k - S_\ell)^2\big)^{1/2} = \big(\sum_{j=k}^{\ell} w_j^2 \sigma_j^2\big)^{1/2} \leq \big(\|w\|_\infty \sum_{j=k}^{\ell} w_j \sigma_j^2\big)^{1/2} \to 0$$

as $k, \ell \to \infty$. Then by the Cauchy Test in L^2 (see e.g. [21, p. 258]), the sequence $\{S_k\}$ converges in L^2 to a random variable. Since the convergence in L^2 implies the almost sure convergence, we write the limit random variable as $\sum_{j=1}^{\infty} w_j \xi_j$ and can understand the convergence of the series as in L^2 or almost surely. Thus, for every $\varepsilon > 0$, we have almost surely

$$\Big\{\Big|\sum_{j=1}^{\infty} w_j \xi_j\Big| > \varepsilon\Big\} \subseteq \cup_{\ell=1}^{\infty} \cap_{r=\ell}^{\infty} \Big\{\Big|\sum_{j=1}^{r} w_j \xi_j\Big| > \varepsilon\Big\}.$$

Then the inequality (5.2) for $\ell < \infty$ implies

$$\begin{aligned} &\text{Prob}\Big\{\Big|\sum_{j=1}^{\infty} w_j \xi_j\Big| > \varepsilon\Big\} \leq \liminf_{\ell\to\infty} \text{Prob}\Big\{\Big|\sum_{j=1}^{\ell} w_j \xi_j\Big| > \varepsilon\Big\} \\ &\leq \liminf_{\ell\to\infty} 2\exp\Big\{-\frac{\varepsilon}{2M \max_{i=1,\dots,\ell} w_i} \log\Big(1 + \frac{M\varepsilon}{\sum_{i=1}^{\ell} w_i \sigma_i^2}\Big)\Big\} \\ &= 2\exp\Big\{-\frac{\varepsilon}{2\|w\|_\infty M} \log\Big(1 + \frac{M\varepsilon}{\sum_{i=1}^{\infty} w_i \sigma_i^2}\Big)\Big\}. \end{aligned}$$

This proves our inequality. □

Remark. *(a) From (5.4), Bennett's inequality [5, 20] follows.*

(b) Corollary 3 always implies the Bernstein inequality up to a constant of $2/3$ *which states for i.i.d. random variables* $\xi_1, \dots, \xi_m$ *with mean* μ *and variance* σ^2 *that*

$$Prob\Big\{\Big|\frac{1}{m}\sum_{j=1}^{m} \xi_j - \mu\Big| > \varepsilon\Big\} \leq 2\exp\Big\{-\frac{m\varepsilon^2}{2(\sigma^2 + \frac{1}{3}M\varepsilon)}\Big\}.$$

To see this, notice that

$$\log(1+\lambda) \geq \frac{\lambda}{1+\frac{1}{2}\lambda}, \qquad \forall \lambda \geq 0. \tag{5.5}$$

Then (5.1) implies

$$Prob\left\{\left|\frac{1}{m}\sum_{j=1}^{m}\xi_j - \mu\right| > \varepsilon\right\} \leq 2\exp\left\{-\frac{m\varepsilon^2}{2(\sigma^2+\frac{1}{2}M\varepsilon)}\right\}.$$

This is the Bernstein inequality except for a loss of two-thirds. The Bernstein inequality can also be derived from (5.4) using the lower bound: $g(\lambda) \geq 3\lambda^2/(6+2\lambda)$.

(c) When the variance is small, the estimate in Corollary 3 (with $\xi_1, \ldots, \xi_m$ *identical) is much better than the Bernstein inequality. In particular, when the variance vanishes, i.e.,* $\sigma_j^2 = 0$ *for each* j*, then Corollary 3 yields* $\frac{1}{m}\sum_{j=1}^{m}\big[\xi_j - E(\xi_j)\big] = 0$ *in probability* 1 *while the Bernstein inequality only gives the estimate* $\frac{1}{m}\big|\sum_{j=1}^{m}\big[\xi_j - E(\xi_j)\big]\big| < \varepsilon$ *with confidence* $1 - 2e^{-m\varepsilon/M}$.

Because of its importance for function reconstruction, Theorem 3 has been developed in greater generality than needed for our immediate use in Theorem 4 below.

Bennett [5] has an early version of our Theorem 3. One may see Devroye, Györfi and Lugosi [12, p. 124] for an account which sketches a proof of (5.3) but with these differences: they have no weighting, there is an extra factor 2, and they use an average of the non-identical random variables. Also, Colin McDiarmid "concentration" Theorem 2.7 [17] is along the same line. The last two references were given to us by David McAllester.

§6. Sample Error

Define

$$\mathcal{E}_{\overline{x}}(f) := \sum_{x\in\overline{x}} w_x (f(x) - f^*(x))^2.$$

This is the empirical error (4.2) with $y_x = f^*(x)$. Then the corresponding minimizer for (4.3) becomes

$$f_{\overline{x},\gamma} := \arg\min_{f\in\mathcal{H}_{K,\overline{t}}}\left\{\mathcal{E}_{\overline{x}}(f) + \gamma\|f\|_K^2\right\}. \tag{6.1}$$

We see from Theorem 2 that $f_{\overline{x},\gamma}$ exists and is unique when $K_{\overline{t},\overline{x}}D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}$ is invertible.

Even when the variance vanishes, $f_{\overline{x},\gamma}$ is not f^* in general. But the error $\|f_{\overline{x},\gamma} - f^*\|^2$ is not caused by noise. It is a deterministic quantity. We shall bound this error in Section 7.

With the weighting, our assumption takes the following general form.

Special Assumption. *For each $x \in X$, ρ_x is a probability measure with zero mean supported on $[-M_x, M_x]$ with $\mathcal{B}_w := \left(\sum_{x\in\overline{x}} w_x M_x^2\right)^{1/2} < \infty$.*

The weighted richness is defined as

$$\lambda_{\overline{x},w} := \inf_{v\in\ell^2(\overline{t})} \|D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} v\|_{\ell^2(\overline{x})} / \|v\|_{\ell^2(\overline{t})}. \tag{6.2}$$

When $\lambda_{\overline{x},w} < \infty$, we have $\|D_w^{\frac{1}{2}} S_{\overline{x}} f\|_{\ell^2(\overline{x})} = \|D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} c\|_{\ell^2(\overline{x})} \geq \lambda_{\overline{x},w} \|c\|_{\ell^2(\overline{t})}$ for $f = \sum_{t\in\overline{t}} c_t K_t$. Hence the sampling operator $S_{\overline{x}}$ satisfies

$$\|D_w^{\frac{1}{2}} S_{\overline{x}} f\|_{\ell^2(\overline{x})} \geq \frac{\lambda_{\overline{x},w}\|f\|_K}{\sqrt{\|K_{\overline{t},\overline{t}}\|}}, \qquad \forall f \in \mathcal{H}_{K,\overline{t}}. \tag{6.3}$$

Corresponding to (3.6), the weighted variance of the system is defined as

$$\sigma_w^2 := \sum_{x\in\overline{x}} w_x \sigma_x^2 \sum_{t\in\overline{t}} K(t,x)^2 w_x \tag{6.4}$$

which is bounded by $\|K_{\overline{t},\overline{x}} D_w^{\frac{1}{2}}\|^2 \sum_{x\in\overline{x}} w_x \sigma_x^2 \leq \|K_{\overline{t},\overline{x}} D_w^{\frac{1}{2}}\|^2 \mathcal{B}_w^2$.

The **sample error** in the form of $\|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\|^2$ involves samples $y = (y_x)_{x\in\overline{x}}$, the weighting w, the point sets $\overline{x}$, $\overline{t}$, and γ. We can apply Theorem 3 to estimate the sample error. To do this, we use the expressions for $f_{\mathbf{z},\gamma}$ (and $f_{\overline{x},\gamma}$) given in Theorem 2. But we shall replace L by the linear operator $L_w : \ell^2(\overline{x}) \to \ell^2(\overline{t})$ defined by

$$L_w := \left(K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\right)^{-1} K_{\overline{t},\overline{x}} D_w^{\frac{1}{2}}. \tag{6.5}$$

It improves our error estimate and is natural: for the rich data case with $\gamma = 0$, L_w is exactly the Moore-Penrose inverse of the operator $D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}}$.

Under the assumption that $K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}$ is invertible, our error bound is given by means of the quantity

$$\kappa := \|K_{\overline{t},\overline{t}}\| \, \|\left(K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\right)^{-1}\|^2. \tag{6.6}$$

Theorem 4 (Sample error). *Suppose* $K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}$ *is invertible. Under the assumption (3.3), let* $f_{\mathbf{z},\gamma} = \sum_{t\in\bar{t}} c_t K_t$ *be the solution of (4.3) given in Theorem 2 by* $c = Ly$. *Set* L_w *and* κ *as in (6.5) and (6.6) respectively. Then for every* $\varepsilon > 0$,

$$Prob\left\{ \|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \le \kappa\sigma_w^2 + \varepsilon \right\} \ge 1 - \exp\left\{ -\frac{\varepsilon}{2\|K_{\bar{t},\bar{t}} L_w\| \, \|L_w\| \mathcal{B}_w^2} \log\big(1 + \frac{\varepsilon}{\kappa\sigma_w^2}\big) \right\}.$$

Proof. Applying Theorem 2 to the sample $f^*|_{\bar{x}}$, we see that $f_{\bar{x},\gamma} = \sum_{t\in\bar{t}} \big(L\big(f^*|_{\bar{x}}\big)\big)_t K_t$. Hence

$$f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma} = \sum_{t\in\bar{t}} \big(L\big(y - f^*|_{\bar{x}}\big)\big)_t K_t = \sum_{t\in\bar{t}} \big(L_w D_w^{\frac{1}{2}} \big(y - f^*|_{\bar{x}}\big)\big)_t K_t$$

and

$$\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 = \big\langle K_{\bar{t},\bar{t}} L_w D_w^{\frac{1}{2}} \big(y - f^*|_{\bar{x}}\big), L_w D_w^{\frac{1}{2}} \big(y - f^*|_{\bar{x}}\big) \big\rangle_{\ell^2(\bar{t})}. \tag{6.7}$$

The expression (6.7) yields the bound

$$\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \le \|K_{\bar{t},\bar{t}} L_w\| \, \|L_w\| \, \|D_w^{\frac{1}{2}} \big(y - f^*|_{\bar{x}}\big)\|_{\ell^2(\bar{x})}^2 \le \|K_{\bar{t},\bar{t}} L_w\| \, \|L_w\| \mathcal{B}_w^2.$$

From (6.7) we also find that

$$\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \le \kappa \|K_{\bar{t},\bar{x}} D_w \big(y - f^*|_{\bar{x}}\big)\|_{\ell^2(\bar{t})}^2.$$

But

$$\|K_{\bar{t},\bar{x}} D_w \big(y - f^*|_{\bar{x}}\big)\|_{\ell^2(\bar{t})}^2 = \sum_{t\in\bar{t}} \left\{ \sum_{x\in\bar{x}} (y_x - f^*(x)) \big\langle K_{\bar{t},\bar{x}} D_w e_x, e_t \big\rangle_{\ell^2(\bar{t})} \right\}^2.$$

Since the random variables $\{y_x - f^*(x)\}_{x\in\bar{x}}$ are independent and have zero means, we see that $E\big((y_x - f^*(x))(y_{x'} - f^*(x'))\big) = \delta_{x,x'}\sigma_x^2$. It follows that

$$E\left(\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \right) \le \kappa \sum_{t\in\bar{t}} \sum_{x\in\bar{x}} w_x^2 \sigma_x^2 K(t,x)^2 = \kappa\sigma_w^2.$$

The one-side inequality of Corollary 3 with $m = 1$, $w = 1$ asserts that for a single random variable ξ satisfying $|\xi| \le M$, there holds for every $\varepsilon > 0$,

$$\text{Prob}\big\{\xi - E(\xi) > \varepsilon\big\} \le \exp\big\{ -\frac{\varepsilon}{2M} \log\big(1 + \frac{M\varepsilon}{\sigma^2(\xi)}\big)\big\}.$$

The random variable $\xi := \|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\|_K^2$ satisfies $0 \le \xi \le M := \|K_{\overline{t},\overline{t}} L_w\| \, \|L_w\| \mathcal{B}_w^2$ almost everywhere, $E(\xi) \le \kappa\sigma_w^2$ and $\sigma^2(\xi) \le ME(\xi) \le M\kappa\sigma_w^2$. Applying the above inequality, we see that with confidence at least $1 - \exp\big\{-\frac{\varepsilon}{2\|K_{\overline{t},\overline{t}}L_w\| \, \|L_w\|\mathcal{B}_w^2} \log\big(1+\frac{\varepsilon}{\kappa\sigma_w^2}\big)\big\}$, there holds $\xi = \|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\|_K^2 \le E(\xi) + \varepsilon \le \kappa\sigma_w^2 + \varepsilon$. □

Remark. *Another sample error estimate can be given by the Markov inequality which states for a nonnegative random variable ξ and $t > 0$ that Prob$\{\xi > t\} \le E(\xi)/t$. Applying this to the random variable $\xi = \|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\|_K^2$ and $t = E(\xi) + \varepsilon$, we have*

$$Prob\Big\{\|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\|_K^2 \le \kappa\sigma_w^2 + \varepsilon\Big\} \ge 1 - \frac{\kappa\sigma_w^2}{\varepsilon + \kappa\sigma_w^2}.$$

This bound is better when $\kappa\sigma_w^2$ is much smaller than ε.

Proposition 2. *The operator L_w defined by (6.5) satisfies*

$$\|L_w\| \le \min\Big\{\frac{1}{\lambda_{\overline{x},w}}, \quad \frac{\|K_{\overline{t},\overline{t}}^{-1}\| \, \|w\|_\infty^{1/2} \|K_{\overline{x},\overline{t}}\|}{\gamma}\Big\}.$$

Also,

$$\|\big(K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\big)^{-1}\| \le \min\Big\{\frac{1}{\lambda_{\overline{x},w}^2}, \quad \frac{\|K_{\overline{t},\overline{t}}^{-1}\|}{\gamma}\Big\}.$$

Proof. Let $v \in \ell^2(\overline{x})$ and $u = L_w v$. Then

$$\big(K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\big) u = K_{\overline{t},\overline{x}} D_w^{\frac{1}{2}} v.$$

Bounding the inner product

$$\big\langle \big(K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}} + \gamma K_{\overline{t},\overline{t}}\big) u, u\big\rangle_{\ell^2(\overline{t})} = < K_{\overline{t},\overline{x}} D_w^{\frac{1}{2}} v, u >_{\ell^2(\overline{t})} = < v, D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u >_{\ell^2(\overline{x})}$$

from below by inner products with the positive definite operators $K_{\overline{t},\overline{x}} D_w K_{\overline{x},\overline{t}}$ and $\gamma K_{\overline{t},\overline{t}}$ separately, we see that $\|D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u\|_{\ell^2(\overline{x})} \|v\|_{\ell^2(\overline{x})}$ is bounded from below by $\frac{\gamma}{\|K_{\overline{t},\overline{t}}^{-1}\|} \|u\|_{\ell^2(\overline{x})}^2$ and by $< D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u, D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u >_{\ell^2(\overline{x})} = \|D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u\|_{\ell^2(\overline{x})}^2 \ge \lambda_{\overline{x},w} \|u\|_{\ell^2(\overline{t})} \|D_w^{\frac{1}{2}} K_{\overline{x},\overline{t}} u\|_{\ell^2(\overline{x})}$. It follows that

$$\|u\|_{\ell^2(\overline{t})} \le \min\Big\{\frac{\|K_{\overline{t},\overline{t}}^{-1}\|}{\gamma} \|w\|_\infty^{1/2} \|K_{\overline{x},\overline{t}}\|, \quad \frac{1}{\lambda_{\overline{x},w}}\Big\} \|v\|_{\ell^2(\overline{x})}.$$

Thus the required estimate for $\|L_w\|$ follows. The proof for the second statement is the same. □

Remark. *When $\bar{t} = \bar{x}$, we do not require the standing hypothesis 1 for Theorem 2 and Theorem 4. Take*

$$L = L_{\bar{t},\bar{t}} = \big(K_{\bar{t},\bar{t}} + \gamma D_w^{-1}\big)^{-1}. \tag{6.8}$$

(the parameter γ can be zero when $K_{\bar{t},\bar{t}}$ is invertible) Moreover, we have

$$\|L_{\bar{t},\bar{t}}\| \leq \big(1/\|K_{\bar{t},\bar{t}}^{-1}\| + \gamma/\|w\|_\infty\big)^{-1}.$$

Proposition 2 combining with Theorem 4 presents estimates for the sample errors $\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|^2$ (for both rich and poor data cases). Even in the rich data case, the introduction of the parameter γ improves the well-posedness of the system in Theorem 1.

Proposition 3. *Assume (3.3) and that $K_{\bar{t},\bar{x}} D_w K_{\bar{x},\bar{t}} + \gamma K_{\bar{t},\bar{t}}$ is invertible. Then for every $0 < \delta < 1$, with confidence $1-\delta$ we have the sample error estimate*

$$\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \leq \mathcal{E}_{samp} := \kappa\sigma_w^2\, \alpha^{-1}\left(\frac{2\|K_{\bar{t},\bar{t}} L_w\|\ \|L_w\| \mathcal{B}_w^2}{\kappa\sigma_w^2} \log\frac{1}{\delta}\right), \tag{6.9}$$

where L_w and κ are given by (6.5) and (6.6) respectively, and α is the increasing function defined for $u > 1$ as $\alpha(u) = (u-1)\log u$. In particular, $\mathcal{E}_{samp} \to 0$ when γ tends to infinity or $\sigma_w^2 \to 0$.

Proof. Choose $u = \alpha^{-1}\big(\frac{2\|K_{\bar{t},\bar{t}} L_w\|\ \|L_w\|\mathcal{B}_w^2}{\kappa\sigma_w^2} \log\frac{1}{\delta}\big) > 1$. Then

$$\frac{\kappa\sigma_w^2}{2\|K_{\bar{t},\bar{t}} L_w\|\ \|L_w\|\mathcal{B}_w^2}(u-1)\log u = \log\frac{1}{\delta}.$$

Set $\varepsilon = \kappa\sigma_w^2(u-1)$. We have $\varepsilon > 0$ since $u > 1$. Also, there holds

$$-\frac{\varepsilon}{2\|K_{\bar{t},\bar{t}} L_w\|\ \|L_w\|\mathcal{B}_w^2}\log\big(1 + \frac{\varepsilon}{\kappa\sigma_w^2}\big) = -\frac{\kappa\sigma_w^2}{2\|K_{\bar{t},\bar{t}} L_w\|\ \|L_w\|\mathcal{B}_w^2}(u-1)\log u = \log\delta.$$

It follows from Theorem 4 that $\|f_{\mathbf{z},\gamma} - f_{\bar{x},\gamma}\|_K^2 \leq \kappa\sigma_w^2 + \varepsilon = \kappa\sigma_w^2 u$ with confidence $1-\delta$. But $\kappa\sigma_w^2 u = \mathcal{E}_{\mathrm{samp}}$. Then the stated sample error estimate follows.

When γ tends to infinity, we see that $\gamma^2\kappa \to \|K_{\bar{t},\bar{t}}\| \, \|K_{\bar{t},\bar{t}}^{-1}\|^2$ and $\gamma^2\|K_{\bar{t},\bar{t}}L_w\| \, \|L_w\| \to \|K_{\bar{t},\bar{x}}D_w^{\frac{1}{2}}\| \, \|K_{\bar{t},\bar{t}}^{-1}K_{\bar{t},\bar{x}}D_w^{\frac{1}{2}}\|$ while $\kappa\sigma_w^2 \to 0$, hence $\mathcal{E}_{\text{samp}} \to 0$.

When $\sigma_w^2 \to 0$, we have $\kappa\sigma_w^2 \to 0$. The definition of the function α tells us that $u \to \infty$ and $\kappa\sigma_w^2 = \frac{2\|K_{\bar{t},\bar{t}}L_w\| \, \|L_w\|\mathcal{B}_w^2 \log\frac{1}{\delta}}{(u-1)\log u}$. It follows that

$$\mathcal{E}_{\text{samp}} = \kappa\sigma_w^2 u = \frac{u}{u-1}\|K_{\bar{t},\bar{t}}L_w\| \, \|L_w\|\mathcal{B}_w^2 \frac{2\log\frac{1}{\delta}}{\log u}$$

which converges to zero. □

§7. Regularization Error and Integration Error

We finish the proof of Theorem 1 and give some estimates for the error $\|f_{\bar{x},\gamma} - f^*\|^2$. The first estimate depends (linearly) on the regularization parameter γ and we call it *regularization error*. Recall that $f^* \in \mathcal{H}_{K,\bar{t}}$.

Proposition 4. *Assume the standing hypotheses. If $f^* \in \mathcal{H}_{K,\bar{t}}$ and $\lambda_{\bar{x},w} > 0$, then*

$$\|f_{\bar{x},\gamma} - f^*\|_K^2 \le \frac{\gamma\|K_{\bar{t},\bar{t}}\|\|f^*\|_K^2}{\lambda_{\bar{x},w}^2}.$$

Proof. According to the definition of $f_{\bar{x},\gamma}$, since $f^* \in \mathcal{H}_{K,\bar{t}}$ we have

$$\mathcal{E}_{\bar{x}}(f_{\bar{x},\gamma}) + \gamma\|f_{\bar{x},\gamma}\|_K^2 \le \mathcal{E}_{\bar{x}}(f^*) + \gamma\|f^*\|_K^2.$$

It follows from the fact $\mathcal{E}_{\bar{x}}(f^*) = 0$ that

$$\|f_{\bar{x},\gamma}\|_K^2 \le \|f^*\|_K^2 \tag{7.1}$$

and

$$\mathcal{E}_{\bar{x}}(f_{\bar{x},\gamma}) \le \gamma\|f^*\|_K^2. \tag{7.2}$$

But $\mathcal{E}_{\bar{x}}(f_{\bar{x},\gamma}) = \sum_{x\in\bar{x}} w_x\big(f_{\bar{x},\gamma}(x) - f^*(x)\big)^2 = \|D_w^{\frac{1}{2}}S_{\bar{x}}\big(f_{\bar{x},\gamma} - f^*\big)\|_{\ell^2(\bar{x})}^2$. Together with (6.3) and (7.2) this implies

$$\gamma\|f^*\|_K^2 \ge \frac{\lambda_{\bar{x},w}^2 \, \|f_{\bar{x},\gamma} - f^*\|_K^2}{\|K_{\bar{t},\bar{t}}\|}.$$

Then the desired estimate follows. □

Proof of Theorem 1. Since $\gamma = 0$ and $w \equiv 1$ in Theorem 1, the expression for $f_{\mathbf{z}}$ follows from Theorem 2 and we see from Proposition 4 that $f^* = f_{\overline{x},0}$. Moreover, the operator $L_w = L$ in Theorem 4 becomes $\big(K_{\overline{t},\overline{x}} K_{\overline{x},\overline{t}}\big)^{-1} K_{\overline{t},\overline{x}}$, the one given in Theorem 1. Also, $\sigma_w^2 = \sigma^2$.

Since $\lambda_{\overline{x},w} = \lambda_{\overline{x}} > 0$, Proposition 2 yields $\|L_w\| \leq 1/\lambda_{\overline{x}}$ and $\|\big(K_{\overline{t},\overline{x}} K_{\overline{x},\overline{t}}\big)^{-1}\| \leq 1/\lambda_{\overline{x}}^2$. Putting all these into Theorem 4, we know that for every $\varepsilon > 0$,

$$\mathrm{Prob}\big\{\|f_{\mathbf{z}} - f^*\|_K^2 \leq \kappa\sigma^2 + \varepsilon\big\} \geq 1 - \exp\big\{-\frac{\varepsilon\lambda_{\overline{x}}^2}{2\|K_{\overline{t},\overline{t}}\|\mathcal{B}^2} \log\big(1 + \frac{\varepsilon}{\kappa\sigma^2}\big)\big\}.$$

Here $\kappa \leq \frac{\|K_{\overline{t},\overline{t}}\|}{\lambda_{\overline{x}}^4}$. This proves Theorem 1. □

For the general situation including the poor data case, our estimate will be given under a Lipschitz continuity assumption involving the Voronoi of X. We call it *integration error* because the estimate comes from bounding the integral over X by sample values at $\overline{x}$.

Let $\overline{X} = (X_x)_{x \in \overline{x}}$ be the Voronoi of X associated with $\overline{x}$, and $w_x = \rho_X(X_x)$.

Define the Lipschitz norm on a subset $X' \subseteq X$ as

$$\|f\|_{\mathrm{Lip}(X')} := \|f\|_{L^\infty(X')} + \sup_{s,u \in X} \frac{|f(s) - f(u)|}{\|s - u\|_{\ell^\infty(\mathbb{R}^n)}}. \tag{7.3}$$

We shall assume that the inclusion map of $\mathcal{H}_{K,\overline{t}}$ into the Lipschitz space satisfies

$$C_{\overline{x}} := \sup_{f \in \mathcal{H}_{K,\overline{t}}} \frac{\sum_{x \in \overline{x}} w_x \|f\|_{\mathrm{Lip}(X_x)}^2}{\|f\|_K^2} < \infty. \tag{7.4}$$

This assumption is true if X is compact and the inclusion map of $\mathcal{H}_{K,\overline{t}}$ into the space of Lipschitz functions on X is bounded (This is the case when K is a C^2 Mercer kernel, see [33]). In fact, if $\|f\|_{\mathrm{Lip}(X)} \leq C_0\|f\|_K$ for each $f \in \mathcal{H}_{K,\overline{t}}$, then $C_{\overline{x}} \leq C_0^2 \rho_X(X)$.

When K is a convolution kernel satisfying a mild decay condition, (7.4) also holds. See Proposition 5 below and Example 5 in Section 8.

Theorem 5. *Assume the standing hypotheses. Let* $\overline{X} = (X_x)_{x\in\overline{x}}$ *be the Voronoi of* X *associated with* $\overline{x}$*, and* $w_x = \rho_X(X_x)$*. If* $\overline{x}$ *is* Δ*-dense,* $C_{\overline{x}} < \infty$*, and* $f^* \in \mathcal{H}_{K,\bar{t}}$*, then*

$$\|f_{\overline{x},\gamma} - f^*\|^2 \le \|f^*\|_K^2 \big(\gamma + 8C_{\overline{x}}\Delta\big).$$

Proof. Let $f \in \mathcal{H}_{K,\bar{t}}$. Then

$$\mathcal{E}_{\overline{x}}(f) = \sum_{x\in\overline{x}} w_x (f(x) - f^*(x))^2 = \sum_{x\in\overline{x}} (f(x) - f^*(x))^2 \int_{X_x} d\rho_X.$$

It follows that

$$\|f - f^*\|^2 \le \mathcal{E}_{\overline{x}}(f) + I_f,$$

where $I_f := \big|\sum_{x\in\overline{x}} \int_{X_x} (f(x) - f^*(x))^2 - (f(u) - f^*(u))^2 d\rho_X(u)\big|$.

For each $x \in \overline{x}$ and $u \in X_x$,

$$\big|(f(x) - f^*(x))^2 - (f(u) - f^*(u))^2\big| \le 2\|f - f^*\|^2_{\mathrm{Lip}(X_x)} \|x - u\|_{\ell^\infty(\mathbb{R}^n)}.$$

Since $\overline{x}$ is Δ-dense, we must have $\|x - u\|_{\ell^\infty(\mathbb{R}^n)} \le \Delta$, otherwise $u \in X_{x'}$ for some $x' \ne x$. Moreover, $\rho_X(X_x) = w_x$. Hence

$$I_f \le 2\big\{\sum_{x\in\overline{x}} w_x \|f - f^*\|^2_{\mathrm{Lip}(X_x)}\big\}\Delta \le 2C_{\overline{x}}\|f - f^*\|_K^2 \Delta.$$

Take f to be $f_{\overline{x},\gamma}$. Then

$$\|f_{\overline{x},\gamma} - f^*\|^2 \le \mathcal{E}_{\overline{x}}(f_{\overline{x},\gamma}) + 2C_{\overline{x}}\|f_{\overline{x},\gamma} - f^*\|_K^2 \Delta.$$

This in connection with (7.1) and (7.2) implies

$$\|f_{\overline{x},\gamma} - f^*\|^2 \le \gamma\|f^*\|_K^2 + 8C_{\overline{x}}\|f^*\|_K^2 \Delta.$$

This proves Theorem 5. □

From the proof of Theorem 5, we see that for $f \in \mathcal{H}_{K,\bar{t}}$ and $x \in \overline{x}$,

$$\int_{X_x} |f(u)|^2 d\rho_X \le \rho_X(X_x)\|f\|^2_{L^\infty(X_x)} \le w_x\|f\|^2_{\mathrm{Lip}(X_x)}.$$

Then the following holds.

Corollary 4. *Under the assumption of Theorem 5, there holds*

$$\|f\|^2 \le C_{\overline{x}} \|f\|_K^2, \qquad \forall f \in \mathcal{H}_{K,\overline{t}}.$$

Theorem 5 and Theorem 4 (together with the bounds in Corollary 4 and Proposition 3) proves the following error estimate.

Corollary 5. *Under the standing hypotheses and the assumption (3.3), let $\overline{X} = (X_x)_{x\in\overline{x}}$ be the Voronoi associated with $\overline{x}$ and $w_x = \rho_X(X_x)$. If $\overline{x}$ is Δ-dense, $C_{\overline{x}} < \infty$, and $f^* \in \mathcal{H}_{K,\overline{t}}$, then for every $0<\delta<1$, with confidence $1-\delta$ there holds*

$$\left\|f_{\mathbf{z},\gamma} - f^*\right\|^2 \le 2C_{\overline{x}}\mathcal{E}_{\text{samp}} + 2\gamma\|f^*\|_K^2 + 16C_{\overline{x}}\|f^*\|_K^2\Delta$$

where $\mathcal{E}_{\text{samp}}$ is given by (6.9) in Proposition 3.

Let us verify the condition (7.4) under some decay condition for K.

Proposition 5. *Assume standing hypothesis 1. Let $\overline{X} = (X_x)_{x\in\overline{x}}$ be the Voronoi associated with $\overline{x}$, and $w_x = \rho_X(X_x)$. If each K_t is Lipschitz on X_x satisfying*

$$B_{\overline{t}} := \sup_{x\in\overline{x}} \sum_{t\in\overline{t}} \|K_t\|_{Lip(X_x)} < \infty, \quad B_{\overline{x}} := \sup_{t\in\overline{t}} w_x \sum_{x\in\overline{x}} \|K_t\|_{Lip(X_x)} < \infty,$$

then

$$C_{\overline{x}} \le 4B_{\overline{t}}B_{\overline{x}}\|K_{\overline{t},\overline{t}}^{-1}\|.$$

Proof. Let $f = \sum_{t\in\overline{t}} c_t K_t \in \mathcal{H}_{K,\overline{t}}$ and $x \in \overline{x}$. Then for $u_1, u_2 \in X_x$,

$$|f(u_1) - f(u_2)| = \big|\sum_{t\in\overline{t}} c_t\big(K_t(u_1) - K_t(u_2)\big)\big| \le \sum_{t\in\overline{t}} |c_t|\|K_t\|_{\text{Lip}(X_x)}\|u_1 - u_2\|_{\ell^\infty(\mathbb{R}^n)}.$$

Also, $\|f\|_{L^\infty(X_x)} \le \sum_{t\in\overline{t}} |c_t|\|K_t\|_{L^\infty(X_x)} \le \sum_{t\in\overline{t}} |c_t|\|K_t\|_{\text{Lip}(X_x)}$. These in connection with the Schwartz inequality tell us that

$$\|f\|_{\text{Lip}(X_x)} \le 2\big\{\sum_{t\in\overline{t}} |c_t|^2\|K_t\|_{\text{Lip}(X_x)}\big\}^{1/2}\big\{\sum_{t\in\overline{t}} \|K_t\|_{\text{Lip}(X_x)}\big\}^{1/2}$$

can be bounded by $2\sqrt{B_{\overline{t}}}\big\{\sum_{t\in\overline{t}} |c_t|^2\|K_t\|_{\text{Lip}(X_x)}\big\}^{1/2}$. Therefore we have

$$\sum_{x\in\overline{x}} w_x\|f\|_{\text{Lip}(X_x)}^2 \le 4B_{\overline{t}} \sum_{t\in\overline{t}} |c_t|^2\big\{\sum_{x\in\overline{x}} w_x\|K_t\|_{\text{Lip}(X_x)}\big\} \le 4B_{\overline{t}}B_{\overline{x}}\|c\|_{\ell^2(\overline{t})}^2.$$

But $\|c\|^2_{\ell^2(\bar{t})} \leq \|K^{-1}_{\bar{t},\bar{t}}\| \|f\|^2_K$. Then our conclusion follows. □

For the poor data situation, the integration error can be bad. In fact, if $K_{\overline{x},\bar{t}}c = 0$ for some $c \in \ell^2(\bar{t})$, set $f^* = \sum_{t\in\bar{t}} c_t K_t \in \mathcal{H}_{K,\bar{t}}$. Then $f^*(x) = 0$ for each $x \in \overline{x}$. Hence $f_{\overline{x},\gamma} = 0$ and $\|f_{\overline{x},\gamma} - f^*\| = \|f^*\|$ for any $\gamma > 0$.

Summarizing, our main goal of the error estimate is to bound the difference $f_{\mathbf{z},\gamma} - f^*$ (either $\|f_{\mathbf{z},\gamma} - f^*\|_K$ or even $\|f_{\mathbf{z},\gamma} - f^*\|_{L^2_{\rho_X}}$). But

$$\|f_{\mathbf{z},\gamma} - f^*\| \leq \|f_{\mathbf{z},\gamma} - f_{\overline{x},\gamma}\| + \|f_{\overline{x},\gamma} - f^*\|$$

Each of the two summands on the right is estimated separately, the first via Theorem 4 and the second in two cases: $\lambda_{\overline{x},w} > 0$ by Proposition 4; in general by Theorem 5.

§8. Convolution Kernels

Some estimates for $\lambda_{\overline{x}}$ will be given for convolution kernels having $\|K^{-1}_{\bar{t},\bar{t}}\| < \infty$. We consider now the setting with $X = \mathbb{R}^n$, $w \equiv 1$ and $\bar{t} = \mathbb{Z}^n$ (The more general situation of $X \subset \mathbb{R}^n$ can be analyzed as in the discussion in Section 3).

The convolution kernels take the form:

$$K(s,u) = \psi(s-u), \quad s,u \in \mathbb{R}^n \quad \text{with } \psi \in L^2(\mathbb{R}^n) \text{ being continuous and even.} \tag{8.1}$$

For these kernels, $K(s,s) = \psi(0)$ for any s. Then K is Mercer if and only if ψ has nonnegative Fourier transform $\hat{\psi}(\xi) \geq 0$. See [6]. The Gaussian is an example of a convolution kernel. More examples can be seen in [3, 8, 14, 27, 32].

Proposition 6. *Let $X = \mathbb{R}^n$, $\bar{t} = \mathbb{Z}^n$, and K be as in (8.1). Then both $\|K_{\bar{t},\bar{t}}\|$ and $\|K^{-1}_{\bar{t},\bar{t}}\|$ are finite if and only if for some $0 < a \leq b < \infty$,*

$$a \leq \sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi) \leq b, \qquad \forall \xi. \tag{8.2}$$

Note that the function $\sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi)$ is 2π-periodic. From Proposition 6, one can easily find "kernels" which satisfy our standing hypotheses but are not Mercer kernels on X: take ψ whose Fourier transform is not nonnegative but satisfies (8.2) for positive constants a, b.

The proof of Proposition 6 follows from the expressions for $\|K_{\bar{t},\bar{t}}\|$ and $\|K^{-1}_{\bar{t},\bar{t}}\|$ in Lemma 1 which give the sharp bounds for a and b.

Lemma 1. *Let* $\bar{t} = \mathbb{Z}^n$ *and* $K(s,u) = \psi(s-u)$ *with some continuous even function* $\psi \in L^2(\mathbb{R}^n)$ *satisfying (8.2) for* $a, b > 0$. *Then standing hypothesis 1 holds. In particular,*

(a) $\|K_{\bar{t},\bar{t}}\| = \left\| \sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi) \right\|_{L^\infty} \leq b.$

(b) $\|K_{\bar{t},\bar{t}}^{-1}\| = \left\| \left(\sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi) \right)^{-1} \right\|_{L^\infty} \leq \frac{1}{a}.$

Proof. Note that

$$< K_{\bar{t},\bar{t}}c, c >_{\ell^2(\bar{t})} = \sum_{t,t'\in\bar{t}} \psi(t-t')c_t c_{t'} = (2\pi)^{-n} \int_{\mathbb{R}^n} \hat{\psi}(\xi) \left| \sum_{t\in\bar{t}} c_t e^{i\xi\cdot t} \right|^2 d\xi$$
$$= (2\pi)^{-n} \int_{[-\pi,\pi]^n} \left(\sum_{\ell\in\mathbb{Z}^n} \hat{\psi}(\xi + 2\ell\pi) \right) \left| \sum_{t\in\bar{t}} c_t e^{i\xi\cdot t} \right|^2 d\xi \geq 0.$$

Then $K_{\bar{t},\bar{t}}$ is positive. From the identity

$$\left\{ (2\pi)^{-n} \int_{[-\pi,\pi]^n} \left| \sum_{t\in\bar{t}} c_t e^{-i\xi\cdot t} \right|^2 d\xi \right\}^{1/2} = \|c\|_{\ell^2(\bar{t})}, \qquad \forall c \in \ell^2(\bar{t}),$$

we see that the upper bounds for the norms hold. The lower bounds can be seen by taking for each $\varepsilon > 0$, a sequence $c \in \ell^2(\bar{t})$ whose Fourier series is the characteristic function of the set $\{\xi \in [\pi,\pi]^n : |F(\xi)| \geq \|F\|_{L^\infty} - \varepsilon\}$. Here $F(\xi) = \sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi)$ or $\left(\sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi) \right)^{-1}$. □

Remark. *The same norm expressions hold when one scales the set* $\mathbb{Z}^n$ *by a constant* $H > 0$*: if* $\bar{t} = H\mathbb{Z}^n$ *and* $\Psi(\xi) := \sum_{j\in\mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi/H) \geq 0$, *then* $\|K_{\bar{t},\bar{t}}\| = H^{-n}\|\Psi\|_{L^\infty}$ *and* $\|K_{\bar{t},\bar{t}}^{-1}\| = H^n\|\Psi^{-1}\|_{L^\infty}$.

Turn to the Shannon example. Here K is a convolution kernel generated by the sinc function ϕ whose Fourier transform $\hat{\phi}$ is the characteristic function of the interval $[-\pi,\pi]$.

Example 4. *Let* $n = 1$ *and* $\psi(x) = \phi(x) = \sin(\pi x)/(\pi x)$ *be the sinc function and* K *given by (8.1). Then for* $\bar{t} = \mathbb{Z}$, $\{K_j\}_{j\in\mathbb{Z}}$ *is an orthonormal basis of* $\mathcal{H}_{K,\bar{t}}$, $\|K_{\bar{t},\bar{t}}\| = \|K_{\bar{t},\bar{t}}^{-1}\| = 1$, *and*

$$\mathcal{H}_{K,\bar{t}} = \left\{ \sum_{j\in\mathbb{Z}} c_j \frac{\sin\pi(x-j)}{\pi(x-j)} : \quad c \in \ell^2(\mathbb{Z}) \right\}.$$

Moreover, as subspaces of $L^2(\mathbb{R})$, *we have*

$$\mathcal{H}_K = \mathcal{H}_{K,\bar{t}} = V := \{f \in L^2(\mathbb{R}) : \hat{f}(\xi) = 0 \quad \forall \xi \notin [-\pi,\pi]\}.$$

Proof. Take the inner product on V to be the one inherited from $L^2(\mathbb{R})$, we see from the Plancheral formula and the fact $\hat{\psi}(\xi) = \chi_{[-\pi,\pi]}$ that

$$< K_t, K_s >_{L^2} = (2\pi)^{-1} < \widehat{K}_t, \widehat{K}_s >_{L^2} = (2\pi)^{-1} \int_{\mathbb{R}} |\hat{\psi}(\xi)|^2 e^{i\xi(t-s)} d\xi = \psi(t-s) = K(t,s).$$

Thus, $< K_t, K_s >_{L^2} = < K_t, K_s >_K$. Also, $\widehat{K}_t = e^{-it\xi}\hat{\psi}(\xi)$ is supported on $[-\pi,\pi]$, hence $K_t \in V$ for any t. Moreover, for each $f \in V$, we have $\hat{f}$ supported on $[-\pi,\pi]$ and given on this interval by $\sum_{j\in\mathbb{Z}} c_j e^{-ij\xi}$ for some $c \in \ell^2$. Hence $\hat{f} = \sum_{j\in\mathbb{Z}} c_j \widehat{K}_j$, and $f = \sum_{j\in\mathbb{Z}} c_j K_j$. Therefore, $\mathcal{H}_K = \mathcal{H}_{K,\bar{t}} = (V, \|\cdot\|_{L^2(\mathbb{R})})$. □

Denote $C_{n,\alpha} := 2^n\big(1 + n^{\alpha/2}/(\alpha - n)\big)$. For $\mathcal{L} \in (0, 1/4)$, we set in the following

$$C_- := \big(\cos\mathcal{L}\pi - \sin\mathcal{L}\pi\big)^n, \qquad C_+ := \big(2 - \cos\mathcal{L}\pi + \sin\mathcal{L}\pi\big)^n.$$

We expand the setting now where we do not have a kernel. In this new setting, just a continuous function $\psi \in L^2(\mathbb{R}^n)$ (not necessarily even) is involved. Then the operator $K_{\overline{x},\bar{t}}$ is replaced by $\mathcal{C}_{\overline{x},\bar{t}} : \ell^2(\bar{t}) \to \ell^2(\overline{x})$ defined as

$$\big(\mathcal{C}_{\overline{x},\bar{t}}a\big)_x = \sum_{t\in\bar{t}} \psi(x-t)a_t. \tag{8.3}$$

The constant $\lambda_{\overline{x}}$ is also defined similarly by

$$\lambda_{\overline{x}} := \inf_{v\in\ell^2(\bar{t})} \|\mathcal{C}_{\overline{x},\bar{t}}v\|_{\ell^2(\overline{x})}/\|v\|_{\ell^2(\bar{t})}. \tag{8.4}$$

Theorem 6. *Let* $0 < \mathcal{L} < 1/4$, $\bar{t} = \mathbb{Z}^n$, $h > 0$ *with* $1/h \in \mathbb{N}$, *and* $\overline{u} = \{u_j\}_{j\in\mathbb{Z}^n}$ *satisfy* $\|u_j - hj\|_{\ell^\infty(\mathbb{R}^n)} \le \mathcal{L}h$ *for every* $j \in \mathbb{Z}^n$. *Suppose* ψ *is an* L^2 *function on* $\mathbb{R}^n$ *satisfying*

$$\big|\hat{\psi}(\xi)\big| \le C_0(1+|\xi|)^{-\alpha} \qquad \forall \xi \in \mathbb{R}^n \tag{8.5}$$

for some $C_0 > 0, \alpha > n$. *Define* $\mathcal{C}_{\overline{u},\bar{t}}$ *by (8.3) and* $\lambda_{\overline{u}}$ *by (8.4) with* $\overline{x} = \overline{u}$. *Then*

$$\|\mathcal{C}_{\overline{u},\bar{t}}\| \le 2C_+C_0C_{n,\alpha}h^{-n/2}. \tag{8.6}$$

If moreover, for some $0 < c_0 \le C_0$, $h \le \big(\frac{C_-c_0^2}{5C_+C_0^2C_{n,\alpha}}\big)^{2/(2\alpha-n)}$ *and*

$$\sum_{j\in\mathbb{Z}^n} \big|\hat{\psi}(\xi + 2j\pi)\big|^2 \ge c_0^2 \qquad \forall \xi, \tag{8.7}$$

then the constant $\lambda_{\overline{u}}$ *can be bounded from below as*

$$\lambda_{\overline{u}} \ge \frac{C_-c_0}{2}h^{-n/2}.$$

Note that C_0 depends on α. For general $\overline{x}$, we get the following consequences.

Corollary 6. *Let $\overline{t} = \mathbb{Z}^n$, and ψ be an L^2 function on $\mathbb{R}^n$ satisfying (8.5) and (8.7) for some $\alpha > n, 0 < c_0 \le C_0$. If $\overline{x}$ is Δ-dense for some $0 < \Delta \le \frac{\mathcal{L}}{2}\big(\frac{C_- c_0^2}{5C_+ C_0^2 C_{n,\alpha}}\big)^{2/(2\alpha - n)}$ and $0 < \mathcal{L} < 1/4$, then*

$$\lambda_{\overline{x}} \ge \frac{C_- c_0}{2^{1+n/2}} \mathcal{L}^{\frac{n}{2}} \Delta^{-\frac{n}{2}}.$$

Proof. Since $\Delta/\mathcal{L} \le 1$, we can choose some h satisfying $\Delta \le \mathcal{L}h \le 2\Delta$ and $1/h \in \mathbb{N}$. Then $\overline{x}$ is $\mathcal{L}h$-dense. For each $j \in \mathbb{Z}^n$, there is some $u_j \in \overline{x}$ such that $\|hj - u_j\|_{\ell^\infty(\mathbb{R}^n)} \le \mathcal{L}h$. It means that $\overline{u} := \{u_j\}_{j \in \mathbb{Z}^n}$ satisfies the requirement in Theorem 6. As $h \le 2\Delta/\mathcal{L} \le \big(\frac{C_- c_0^2}{5C_+ C_0^2 C_{n,\alpha}}\big)^{2/(2\alpha - n)}$, we conclude by Theorem 6 that for each $c \in \ell^2(\overline{t})$,

$$\|\mathcal{C}_{\overline{x},\overline{t}} c\|_{\ell^2(\overline{x})} \ge \|\mathcal{C}_{\overline{u},\overline{t}} c\|_{\ell^2(\overline{u})} \ge \frac{C_- c_0}{2} h^{-\frac{n}{2}} \|c\|_{\ell^2(\overline{t})}.$$

Hence $\lambda_{\overline{x}} \ge \frac{C_- c_0}{2} h^{-\frac{n}{2}} \ge \frac{C_- c_0}{2^{1+n/2}} \mathcal{L}^{\frac{n}{2}} \Delta^{-\frac{n}{2}}$. □

Now we can see that Proposition 1 follows from Corollary 6: (3.11) in connection with (3.10) tells us that $\sum_{j \in \mathbb{Z}^n} |\hat{\psi}(\xi + 2j\pi)|^2 \ge |\hat{\psi}(\xi)|^2 \ge c_0^2$ for $\xi \in [-\pi, \pi]^n$, hence (8.7) holds.

Standing hypothesis 2 requires the norm $\|K_{\overline{x},\overline{t}}\|$. In the current general setting, we can estimate the norm $\mathcal{C}_{\overline{x},\overline{t}}$ which involves the separation of $\overline{x}$, defined as

$$\mathrm{Sep}_{\overline{x}} := \inf_{x \ne y \in \overline{x}} \|x - y\|_{\ell^\infty(\mathbb{R}^n)}.$$

Corollary 7. *Let $\overline{t} = \mathbb{Z}^n$ and ψ be a function on $\mathbb{R}^n$ satisfying (8.5) for some $C_0 > 0, \alpha > n$. For any discrete set $\overline{x} \subset X$ and $0 < \mathcal{L} < 1/4$, we have*

$$\|\mathcal{C}_{\overline{x},\overline{t}}\| \le 2C_+ C_0 C_{n,\alpha} \left(\max\{\frac{4}{\mathrm{Sep}_{\overline{x}}}, \frac{2}{\mathcal{L}}\}\right)^{n/2}.$$

Proof. Let h be a positive constant with $1/h \in \mathbb{N}$ which will be determined later. Take a set of multi-integers $\Sigma := \big([-\frac{1}{4\mathcal{L}} - \frac{1}{2}, \frac{1}{4\mathcal{L}} + \frac{1}{2}] \cap \mathbb{Z}\big)^n$. We separate the set $\overline{x}$ into $\{\overline{x}^{(\alpha)}\}_{\alpha \in \Sigma}$ where $\overline{x}^{(\alpha)} = \overline{x} \cap (h\mathbb{Z}^n + h\Omega_\alpha)$. Here for $\alpha \in \Sigma$, $\Omega_\alpha = \big((-\mathcal{L}, \mathcal{L}]^n + 2\mathcal{L}\alpha\big) \cap (-\frac{1}{2}, \frac{1}{2}]^n$. Then

$$\Big\|\sum_{t \in \overline{t}} c_t \psi(x - t)\Big\|^2_{\ell^2(\overline{x})} = \sum_{\alpha \in \Sigma} \Big\|\sum_{t \in \overline{t}} c_t \psi(x - t)\Big\|^2_{\ell^2(\overline{x}^{(\alpha)})}.$$

The definition of $\mathrm{Sep}_{\overline{x}}$ tells us that for each $\alpha \in \Sigma$ and $j \in \mathbb{Z}^n$, the set $\overline{x}^{(\alpha)} \cap (hj + h\Omega_\alpha)$ contains at most $S := ([(2\mathcal{L}h)/\mathrm{Sep}_{\overline{x}}] + 1)^n$ points. Thus we can divide the set $\overline{x}^{(\alpha)}$ into S subsets $\{\overline{x}_k^{(\alpha)}\}_{k=1}^S$ such that $\overline{x}_k^{(\alpha)} \cap (hj + h\Omega_\alpha)$ contains at most one point for each $j \in \mathbb{Z}^n$.

Fix α and k. Then there are $J \subseteq \mathbb{Z}^n$ and $\{\theta_j\} \subset [-\mathcal{L}, \mathcal{L}]^n$ such that

$$\overline{x}_k^{(\alpha)} = 2\mathcal{L}\alpha h + \{hj + h\theta_j\}_{j \in J}.$$

Let $\overline{u}^{(\alpha)} = \{hj + h\theta_j\}_{j \in \mathbb{Z}^n}$ where $\theta_j = 0$ for $j \notin J$. Consider the linear operator $\mathcal{C}_{\overline{u}^{(\alpha)}, \overline{t}}$ defined by (8.3) with $\overline{x}$ replaced by $\overline{u}^{(\alpha)}$ and ψ by $\psi(2\mathcal{L}\alpha h + \cdot)$. As $|\psi(2\mathcal{L}\alpha h + \cdot)\hat{}(\xi)| = |\hat{\psi}(\xi)|$, we apply Theorem 6 and conclude that

$$\Big\|\sum_{t \in \overline{t}} c_t \psi(x - t)\Big\|_{\ell^2(\overline{x}_k^{(\alpha)})} \leq \|\mathcal{C}_{\overline{u}^{(\alpha)}, \overline{t}} c\|_{\ell^2(\overline{u}^{(\alpha)})} \leq 2C_+ C_0 C_{n,\alpha} h^{-n/2} \|c\|_{\ell^2(\overline{t})}.$$

This is true for each α, k. Therefore,

$$\|\mathcal{C}_{\overline{x}, \overline{t}} c\|^2_{\ell^2(\overline{x})} = \Big\|\sum_{t \in \overline{t}} c_t \psi(x - t)\Big\|^2_{\ell^2(\overline{x})} = \sum_{\alpha \in \Sigma} \sum_{k=1}^{S} \Big\|\sum_{t \in \overline{t}} c_t \psi(x - t)\Big\|^2_{\ell^2(\overline{x}_k^{(\alpha)})}$$

can be bounded by $(2 + 1/(2\mathcal{L}))^n S \big(2C_+ C_0 C_{n,\alpha} h^{-n/2} \|c\|_{\ell^2(\overline{t})}\big)^2$. Hence

$$\|\mathcal{C}_{\overline{x}, \overline{t}}\| \leq 2C_+ C_0 C_{n,\alpha} \Big(\frac{\mathcal{L} + 1}{\mathrm{Sep}_{\overline{x}}} + \frac{2 + 1/(2\mathcal{L})}{h}\Big)^{n/2} \leq 2C_+ C_0 C_{n,\alpha} \Big(\frac{2}{\mathrm{Sep}_{\overline{x}}} + \frac{1}{\mathcal{L} h}\Big)^{n/2}.$$

When $\mathrm{Sep}_{\overline{x}} \geq 2\mathcal{L}$, we choose $h = 1$ and obtain $\|\mathcal{C}_{\overline{x}, \overline{t}}\| \leq 2C_+ C_0 C_{n,\alpha} \big(2/\mathcal{L}\big)^{n/2}$.

When $\mathrm{Sep}_{\overline{x}} < 2\mathcal{L}$, we choose some h satisfying $1/h \in \mathbb{N}$ and $\mathrm{Sep}_{\overline{x}}/(2\mathcal{L}) \leq h < \mathrm{Sep}_{\overline{x}}/\mathcal{L}$ and obtain $\|\mathcal{C}_{\overline{x}, \overline{t}}\| \leq 2C_+ C_0 C_{n,\alpha} \big(4/\mathrm{Sep}_{\overline{x}}\big)^{n/2}$. This proves Corollary 7. $\square$

Remark. *Note that $\lambda_{\overline{x}} \leq \|\mathcal{C}_{\overline{x}, \overline{t}}\|$. Then we see from the lower bound for $\lambda_{\overline{x}}$ given in Corollary 6 and the upper bound for $\|\mathcal{C}_{\overline{x}, \overline{t}}\|$ stated in Corollary 7 that our estimates are sharp up to a constant depending on the ratio $\Delta/Sep_{\overline{x}}$.*

Remark. *The lower bound in Corollary 6 and the upper bound in Corollary 7 can be established for general convolution kernels without the decay (8.5).*

Remark. *One may consider more general $\overline{t}$. For example, choose $\overline{t}$ to be a subset of $\mathbb{R}^n$ such that $\{e^{-i\xi \cdot t}\}_{t \in \overline{t}}$ is a Riesz system in $L^2([-\pi/H, \pi/H]^n)$ for some $H > 0$. Then similar upper and lower bounds hold with constants depending on H. Here for a Hilbert space $\mathcal{H}$, we say that a sequence of elements $\{\phi_t : t \in \overline{t}\} \subset \mathcal{H}$ is a Riesz system in $\mathcal{H}$ if there are two positive constants $C_1, C_2 > 0$ such that*

$$C_1 \|c\|_{\ell^2(\overline{t})} \leq \Big\|\sum_{t \in \overline{t}} c_t \phi_t\Big\|_{\mathcal{H}} \leq C_2 \|c\|_{\ell^2(\overline{t})}, \qquad \forall c \in \ell^2(\overline{t}).$$

The Riesz system is called a Riesz basis of $\mathcal{H}$ if moreover, span$\{\phi_t\}_{t\in\bar{t}}$ is dense in $\mathcal{H}$.

To prove Theorem 6, we need Kadec's $\frac{1}{4}$-Theorem. See [30], and [24] for the multivariate version:

Let $\mathcal{L} < 1/4$. If $\|x_j - j\|_{\ell^\infty(\mathbb{R}^n)} \le \mathcal{L}$ for each $j \in \mathbb{Z}^n$, then

$$\begin{aligned}(2\pi)^n C_-^2 \|f\|^2_{L^2([-\pi,\pi]^n)} &\le \sum_{j\in\mathbb{Z}^n} \left| < f, e^{-i\xi\cdot x_j} >_{L^2([-\pi,\pi]^n)} \right|^2 \\ &\le (2\pi)^n C_+^2 \|f\|^2_{L^2([-\pi,\pi]^n)}, \qquad \forall f \in L^2([-\pi,\pi]^n).\end{aligned} \tag{8.8}$$

This is the frame property of the Riesz basis $\{e^{-i\xi\cdot x_j}\}_{j\in\mathbb{Z}^n}$ of $L^2([-\pi,\pi]^n)$.

Proof of Theorem 6. Notice that

$$\sum_{j\in\mathbb{Z}^n} (1+|j|)^{-\alpha} \le C_{n,\alpha}.$$

Let $x, t \in \mathbb{R}^n$. Apply the inverse Fourier transform, we obtain for $c \in \ell^2(\bar{t})$,

$$\sum_{t\in\bar{t}} c_t \psi(x-t) = (2\pi)^{-n} \int_{\mathbb{R}^n} \left(\hat{\psi}(\xi) \sum_{t\in\bar{t}} c_t e^{-i\xi\cdot t} \right) e^{i\xi\cdot x} d\xi.$$

Denote $\tilde{c}(\xi) := \sum_{t\in\bar{t}} c_t e^{-i\xi\cdot t}$, $g(\xi) := \hat{\psi}(\xi)\tilde{c}(\xi)$. Then the above expression is

$$(2\pi)^{-n} \int_{\mathbb{R}^n} g(\xi) e^{i\xi\cdot x} d\xi = \sum_{\ell\in\mathbb{Z}^n} (2\pi)^{-n} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]^n} g(\xi + \frac{2\ell\pi}{h}) e^{i\xi\cdot x} e^{i\frac{2\ell\pi}{h}\cdot x} d\xi.$$

If we denote for $\ell \in \mathbb{Z}^n$,

$$I_\ell(g) := \left\{ \sum_{j\in\mathbb{Z}^n} \left| (2\pi)^{-n} \int_{[-\frac{\pi}{h},\frac{\pi}{h}]^n} g(\xi + 2\ell\pi/h) e^{i\xi\cdot u_j} d\xi \right|^2 \right\}^{1/2},$$

then $\|\mathcal{C}_{\overline{u},\bar{t}} c\|_{\ell^2(\overline{u})} = \|\sum_{t\in\bar{t}} c_t \psi(u-t)\|_{\ell^2(\overline{u})}$ can be bounded from above and below as

$$I_0(g) - \sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} I_\ell(g) \le \|\mathcal{C}_{\overline{u},\bar{t}} c\|_{\ell^2(\overline{u})} \le I_0(g) + \sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} I_\ell(g).$$

Let us first derive upper bounds from (8.5) by means of Kadec's $\frac{1}{4}$-Theorem (8.8). The condition on $\overline{u}$ tells us that $\{u_j/h\}_{j\in\mathbb{Z}^n}$ satisfies the condition for (8.8). Applying the upper bound of (8.8) to the functions $g(\xi/h + 2\ell\pi/h)$, we know that

$$I_\ell(g) \le (\sqrt{2\pi} h)^{-n} C_+ \|g(\frac{\xi}{h} + \frac{2\ell\pi}{h})\|_{L^2([-\pi,\pi]^n)}, \qquad \forall \ell \in \mathbb{Z}^n.$$

As $\tilde{c}(\xi)$ is 2π-periodic, $\tilde{c}(\xi + 2\ell\pi/h) = \tilde{c}(\xi)$ because of $1/h \in \mathbb{N}$. Then we see that $h^{-n/2}\|g(\xi/h + 2\ell\pi/h)\|_{L^2([-\pi,\pi]^n)}$ is

$$\begin{aligned}\Big\|g(\xi + \frac{2\ell\pi}{h})\Big\|_{L^2([-\frac{\pi}{h},\frac{\pi}{h}]^n)} &= \Big\|\hat{\psi}(\xi + \frac{2\ell\pi}{h})\tilde{c}(\xi)\Big\|_{L^2([-\frac{\pi}{h},\frac{\pi}{h}]^n)} \\ &\le \Big\{\sum_{s\in[-1/(2h),1/(2h)]^n} \big\|\hat{\psi}(\xi + 2s\pi + \frac{2\ell\pi}{h})\tilde{c}(\xi)\big\|^2_{L^2([-\pi,\pi]^n)}\Big\}^{1/2}.\end{aligned}$$

If we set the quantity A_ℓ^ψ as

$$A_\ell^\psi := \Big\{\sum_{s\in[-1/(2h),1/(2h)]^n} \big\|\hat{\psi}(\xi + 2s\pi + \frac{2\ell\pi}{h})\big\|^2_{L^\infty([-\pi,\pi]^n)}\Big\}^{1/2},$$

we find that

$$I_\ell(g) \le \big(\sqrt{2\pi h}\big)^{-n} C_+ A_\ell^\psi \|\tilde{c}(\xi)\|_{L^2([-\pi,\pi]^n)} \le h^{-n/2} C_+ A_\ell^\psi \|c\|_{\ell^2(\bar{t})}.$$

By the decay condition (8.5), we have $A_0^\psi \le C_0\sqrt{C_{n,\alpha}}$, and

$$\sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} A_\ell^\psi \le \sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} \big(1/h + 1\big)^{n/2} C_0\big(1 + \frac{|\ell|\pi}{2h}\big)^{-\alpha} \le h^{\alpha-n/2} C_0 C_{n,\alpha}$$

which yields

$$\sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} I_\ell(g) \le h^{\alpha-n} C_+ C_0 C_{n,\alpha}\|c\|_{\ell^2(\bar{t})}.$$

Thus, we have

$$\|\mathcal{C}_{\overline{u},\overline{t}}c\|_{\ell^2(\overline{u})} \le I_0(g) + \sum_{\ell\in\mathbb{Z}^n\setminus\{0\}} I_\ell(g) \le 2C_+ C_0 C_{n,\alpha} h^{-n/2}\|c\|_{\ell^2(\bar{t})}.$$

This proves (8.6).

Next we provide a lower bound for $I_0(g)$. Apply the lower bound of (8.8) to the functions $g(\xi/h)$, we find

$$I_0(g) \ge \big(\sqrt{2\pi h}\big)^{-n} C_- \|g\|_{L^2([-\pi/h,\pi/h]^n)}.$$

Observe that

$$\|g\|_{L^2([-\pi/h,\pi/h]^n)} \ge \int_{[-\pi,\pi]^n} \sum_{s\in[-\frac{1}{2h}+\frac{1}{2},\frac{1}{2h}-\frac{1}{2}]^n} \big|\hat{\psi}(\xi + 2s\pi)\big|^2 |\tilde{c}(\xi)|^2 d\xi.$$

But for $\xi \in [-\pi, \pi]^n$,

$$\sum_{s \notin [-\frac{1}{2h}+\frac{1}{2}, \frac{1}{2h}-\frac{1}{2}]^n} \left|\hat{\psi}(\xi + 2s\pi)\right|^2 \le \sum_{|s| \ge 1/(2h)} C_0^2 (1 + |\xi + 2s\pi|)^{-2\alpha} \le C_0^2 C_{n,\alpha} h^{\alpha}.$$

This in connection with (8.7) implies

$$\|g\|^2_{L^2([-\pi/h,\pi/h]^n)} \ge \left(c_0^2 - C_0^2 C_{n,\alpha} h^{\alpha}\right) \|\tilde{c}\|^2_{L^2([-\pi,\pi]^n)}.$$

It follows that

$$I_0(g) \ge h^{-n/2} C_- \sqrt{c_0^2 - C_0^2 C_{n,\alpha} h^{\alpha}} \|c\|_{\ell^2(\bar{t})}.$$

When $h^{\alpha - n/2} \le C_- c_0^2/(5C_+ C_0^2 C_{n,\alpha})$, we have $c_0^2 - C_0^2 C_{n,\alpha} h^{\alpha} \ge c_0^2/2$, and

$$\sum_{\ell \in \mathbb{Z}^n \setminus \{0\}} I_\ell(g) \le (1 - 1/\sqrt{2}) I_0(g), \qquad I_0(g) \ge \frac{c_0}{\sqrt{2}} C_- h^{-n/2} \|c\|_{\ell^2(\bar{t})}.$$

Therefore,

$$\|\mathcal{C}_{\overline{u},\bar{t}} c\|_{\ell^2(\overline{u})} \ge \frac{1}{\sqrt{2}} I_0(g) \ge \frac{C_- c_0}{2} h^{-n/2} \|c\|_{\ell^2(\bar{t})}.$$

Hence $\lambda_{\overline{u}} \ge \frac{C_- c_0}{2} h^{-n/2}$ and the proof of Theorem 6 is complete. □

We study for the convolution kernel the last quantity $C_{\overline{x}}$ required by (7.4). We shall apply Proposition 5 involving the decay of the kernel.

Example 5. *Let $X = \mathbb{R}^n$, $\bar{t} = \mathbb{Z}^n$, $\overline{X} = (X_x)_{x \in \overline{x}}$ be the Voronoi associated with $\overline{x}$ and $w_x = \rho_X(X_x)$. If ρ_X is the Lebesgue measure, $\overline{x}$ is Δ-dense, and ψ is a continuous even function on $\mathbb{R}^n$ satisfying $\sum_{j \in \mathbb{Z}^n} \hat{\psi}(\xi + 2j\pi) \ge c_0 > 0$ for every ξ and*

$$|\psi(x)| + |\nabla \psi(x)| \le C_0 (1 + |x|)^{-\alpha}$$

for some $C_0 > 0, \alpha > n$, then for the kernel $K(s,u) = \psi(s-u)$ we have

$$C_{\overline{x}} \le 8(1+n)(4n)^{\alpha} \left(2^{\alpha} C_{n,\alpha} + 3^n\right)\left(C_{n,\alpha} + 1\right) \frac{C_0^2}{c_0} (\Delta + 1)^{2\alpha}.$$

Proof. Let $t \in \bar{t}$ and $x \in \overline{x}$. Then the decay condition gives

$$\|K_t\|_{\mathrm{Lip}(X_x)} \le C_0(1+\sqrt{n})\left(1 + \inf_{u \in X_x - t} |u|\right)^{-\alpha} \le C_0(1+\sqrt{n})\left(1 + \max\{0, |x-t| - \sqrt{n}\Delta\}\right)^{-\alpha}.$$

It follows immediately that

$$B_{\overline{t}} \leq C_0(1+\sqrt{n}) \sup_{y\in[0,1]^n} \sum_{t\in\mathbb{Z}^n} \big(1+\max\{0, |y-t|-\sqrt{n}\Delta\}\big)^{-\alpha}$$

is bounded by $C_0(1+\sqrt{n})\big(\sqrt{n}(\Delta+1)\big)^{\alpha}\big(2^{\alpha}C_{n,\alpha}+3^n\big)$.

Concerning $B_{\overline{x}}$ we fix $t \in \overline{t}$, and see from $w_x = \rho_X(X_x)$ that

$$\sum_{x\in\overline{x}} w_x \|K_t\|_{\mathrm{Lip}(X_x)} \leq \sum_{x\in\overline{x}} C_0(1+\sqrt{n}) \int_{X_x} \big(1+\max\{0, |x-t|-\sqrt{n}\Delta\}\big)^{-\alpha} d\rho_X$$

can be bounded by $C_0(1+\sqrt{n})\int_X \big(1+\max\{0,|y-t|-2\sqrt{n}\Delta\}\big)^{-\alpha} d\rho_X$. As ρ_X is the Lebesgue measure, the integral is bounded by

$$\int_{\mathbb{R}^n} \big(1+\max\{0, |y|-2\sqrt{n}\Delta\}\big)^{-\alpha} dy \leq (2+4\sqrt{n}\Delta)^{\alpha} C_{n,\alpha} + (4\sqrt{n}\Delta)^n.$$

Therefore,

$$B_{\overline{x}} \leq C_0(1+\sqrt{n})(2+4\sqrt{n}\Delta)^{\alpha}\big(C_{n,\alpha}+1\big).$$

Then the estimate for $C_{\overline{x}}$ follows from Proposition 5 and Lemma 1. □

More general decay conditions such as the Wiener amalgam spaces [15, 2] can be used for the condition (8.5) on ψ or the decay of ρ_x.

§9. Estimating the Operator Norms for Compact Domains

When X is compact, the richness $\lambda_{\overline{x}}$ can be easily bounded from below. Moreover, it will be shown that $\lambda_{\overline{x}} \to \infty$ when $\overline{x}$ becomes dense. Denote

$$N_\sigma(\overline{x}) := \sup\{d \in \mathbb{N}: \text{ for each } x \in X, \text{ there are } (x_i)_{i=1}^d \subset \overline{x} \text{ satisfying } |x_i - x| \leq \sigma\}.$$

Proposition 7. *Let $\overline{t} = (t_i)_{i=1}^s$ be finite. Then for sufficiently small $\sigma > 0$ there holds*

$$\big|K(u,t') - K(t,t')\big| \leq \frac{1}{2s\|K^{-1}_{\overline{t},\overline{t}}\|}, \qquad \forall t \in \overline{t}, u \in X \quad \textit{with} \quad |u-t| \leq \sigma \tag{9.1}$$

for each $t' \in \overline{t}$. In this case,

$$\lambda_{\overline{x}} \geq \frac{\sqrt{N_\sigma(\overline{x})}}{2\|K^{-1}_{\overline{t},\overline{t}}\|}.$$

In particular, $\lambda_{\overline{x}} \to \infty$ *when* $N_\sigma(\overline{x}) \to \infty$.

Proof. The continuity of K tells us that for sufficiently small $\sigma > 0$, (9.1) holds for each $t' \in \overline{t}$.

Let $0 < \sigma < \frac{1}{2}\mathrm{Sep}_{\overline{t}}$. By the definition of $N_\sigma(\overline{x}) =: N$, for each $t \in \overline{t}$ there are $(u_t^{(j)})_{j=1}^N \subset \overline{x}$ such that $|u_t^{(j)} - t| \le \sigma$. As $\sigma < \frac{1}{2}\mathrm{Sep}_{\overline{t}}$, we know that $(u_t^{(j)})_{j=1}^N \cap (u_{t'}^{(j)})_{j=1}^N = \emptyset$ when $t \neq t'$.

Fix $j \in \{1, \ldots, N\}$. The set $\overline{u}^{(j)} = (u_t^{(j)})_{t \in \overline{t}}$ satisfies $|u_t^{(j)} - t| \le \sigma$. By (9.1), we see that

$$\left| \left(K_{\overline{u}^{(j)}, \overline{t}} c\right)_{u_t^{(j)}} - \left(K_{\overline{t}, \overline{t}} c\right)_t \right| = \left| \sum_{t' \in \overline{t}} c_{t'} \left(K(u_t^{(j)}, t') - K(t, t') \right) \right| \le \|c\|_{\ell^2(\overline{t})} \frac{1}{2\|K_{\overline{t},\overline{t}}^{-1}\| \sqrt{s}}.$$

Therefore,

$$\|K_{\overline{u}^{(j)}, \overline{t}} c - K_{\overline{t}, \overline{t}} c\|_{\ell^2(\overline{t})} \le \frac{\|c\|_{\ell^2(\overline{t})}}{2\|K_{\overline{t},\overline{t}}^{-1}\|}, \qquad \forall c \in \ell^2(\overline{t}),$$

and

$$\|K_{\overline{u}^{(j)}, \overline{t}} c\|_{\ell^2(\overline{u})} \ge \frac{1}{\|K_{\overline{t},\overline{t}}^{-1}\|} \|c\|_{\ell^2(\overline{t})} - \frac{\|c\|_{\ell^2(\overline{t})}}{2\|K_{\overline{t},\overline{t}}^{-1}\|}.$$

It follows that

$$\|K_{\overline{x}, \overline{t}} c\|_{\ell^2(\overline{x})}^2 \ge \sum_{j=1}^N \|K_{\overline{u}^{(j)}, \overline{t}} c\|_{\ell^2(\overline{u}^{(j)})}^2 \ge N_\sigma(\overline{x}) \left(\frac{\|c\|_{\ell^2(\overline{t})}}{2\|K_{\overline{t},\overline{t}}^{-1}\|} \right)^2.$$

Then our conclusion follows. □

§10. Extension to a Setting without a Kernel

Our study can be extended to a setting without a kernel K.

Let $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ be a Hilbert space of continuous function on X, finite or infinite dimensional. Let $\{\phi_t : t \in \overline{t}\}$ be an orthonormal basis. Then $\mathcal{H}_{K,\overline{t}}$ is replaced by $\mathcal{H}$ and $K_{\overline{t},\overline{t}}$ by the identity operator on $\mathcal{H}$, hence standing hypothesis 1 holds. Now the linear operator $\mathcal{C}_{\overline{x},\overline{t}} : \ell^2(\overline{t}) \to \ell^2(\overline{x})$ is given by the matrix $\left(\phi_t(x)\right)_{x \in \overline{x}, t \in \overline{t}}$, and only standing hypothesis 2 is required, where $\mathcal{C}_{\overline{x},\overline{t}}$ replaces $K_{\overline{x},\overline{t}}$. The main results are still true. For example, take $\sigma^2 := \sum_{x \in \overline{x}} \sigma_x^2 \sum_{t \in \overline{t}} (\phi_t(x))^2$. Corresponding to Theorem 1, we have

Theorem 7. *Assume* $f^* \in \mathcal{H}$ *with* $\mathcal{H}, X, \rho, \{\phi_t\}_{t\in\bar{t}}$ *as above,* y *as in (3.3). If* $\bar{x}$ *provides rich data, then the optimization problem* $\arg\min_{f\in\mathcal{H}} \sum_{x\in\bar{x}} \big(f(x) - y_x\big)^2$ *can be solved:*

$$f_{\mathbf{z}} = \sum_{t\in\bar{t}} a_t \phi_t, \quad a = Ly \quad \text{and} \quad L = \big(\mathcal{C}^T_{\bar{x},\bar{t}} \mathcal{C}_{\bar{x},\bar{t}}\big)^{-1} \mathcal{C}^T_{\bar{x},\bar{t}}.$$

Moreover, for every $\varepsilon > 0$, *there holds*

$$Prob\Big\{ \|f_{\mathbf{z}} - f^*\|^2_{\mathcal{H}} \le \frac{\sigma^2}{\lambda^4_{\bar{x}}} + \varepsilon \Big\} \ge 1 - \exp\Big\{ -\frac{\varepsilon \lambda^2_{\bar{x}}}{2\mathcal{B}^2} \log\big(1 + \frac{\varepsilon\lambda^4_{\bar{x}}}{\sigma^2}\big) \Big\}.$$

Examples of finite dimensional spaces $\mathcal{H}$ include polynomial spaces for the purpose of interpolation. Examples of infinite dimensional spaces include the Fourier series (the most classical!); function spaces on a 2-dimensional rectangle (with eigenfunctions of Laplacian being the orthonormal basis); and wavelet spaces (with an orthonormal basis of wavelets or shifts of refinable functions).

Next suppose that $\{\phi_t : t \in \bar{t}\}$ is only a Riesz basis of $\mathcal{H}$. Then the mapping $\mathcal{K} : \ell^2(\bar{t}) \to \mathcal{H}$ given by $\mathcal{K}c = \sum_{t\in\bar{t}} c_t \phi_t$ is an isomorphism. This isomorphism plays the role of $K_{\bar{t},\bar{t}}$. The setting is now similar to the one with standing hypothesis 1 satisfied. One example is generated by a (stable, but not necessarily orthogonal) scaling function φ of a multiresolution analysis in wavelet analysis. Take $k \in \mathbb{Z}, \bar{t} = \mathbb{Z}^n$, and $\phi_t = \varphi(2^k \cdot -t)$, the scaled shifts of φ. Then estimates for $\lambda_{\bar{x}}$ can be given as in Section 8, which would lead to sample error estimates like Theorem 1. The regularization error and integration error estimates can be obtained from the approximation properties of multiresolution analysis [10, 23].

Remark. *In this paper we study the error* $\|f_{\bar{x},\gamma} - f^*\|^2$ *(regularization error or integration error estimates) under the assumption* $f^* \in \mathcal{H}_{K,\bar{t}}$. *It would be interesting to have some estimates for the error without this assumption. One situation is when* $\mathcal{H}_{K,\bar{t}}$ *is a closed subspace of a RKHS* $\mathcal{H}_K$ *generated by a Mercer kernel* K *and* $f^* \in \mathcal{H}_K$. *One may study the error even for* f^* *to be outside* $\mathcal{H}_K$, *as done for the approximation error in [22, 26].*

References

[1] A. Aldroubi, Non-uniform weighted average sampling and reconstruction in shift-invariant and wavelet spaces, Appl. Comput. Harmon. Anal. **13** (2002), 151–161.

[2] A. Aldroubi and K. Gröchenig, Non-uniform sampling and reconstruction in shift-invariant spaces, SIAM Review **43** (2001), 585–620.

[3] N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc. **68** (1950), 337–404.

[4] R. F. Bass and K. Gröchenig, Random sampling of multivariate trigonometric polynomials, preprint, 2003.

[5] G. Bennett, Probability inequalities for the sum of independent random variables, J. Amer. Statis. Assoc. **57** (1962), 33–45.

[6] S. Bochner, Hilbert distances and positive definite functions, Annals Math. **42** (1941), 647–656.

[7] T. Chan, J. Shen, and L. Vese, Variational PDE models in image processing, Notices Amer. Math. Soc. **50** (2003), 14–26.

[8] F. Cucker and S. Smale, On the mathematical foundations of learning, Bull. Amer. Math. Soc. **39** (2001), 1–49.

[9] F. Cucker and S. Smale, Best choices for regularization parameters in learning theory, Foundat. Comput. Math. **2** (2002), 413–428.

[10] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[11] J. P. Dedieu, Newton's method and some complexity aspects of the zero-finding problem, in *Foundations of Computational Mathematics*, R. A. DeVore, A. Iserles, and E. Süli (eds.), Cambridge University Press, 2001, pp. 45–67.

[12] L. Devroye, L. Györfi, and G. Lugosi, A Probabilistic Theory of Pattern Recognition, Spinger-Verlag, New York, 1996.

[13] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Mathematics and Its Applications, **375**, Kluwer, Dordrecht, 1996.

[14] T. Evgeniou, M. Pontil, and T. Poggio, Regularization networks and support vector machines, Adv. Comput. Math. **13** (2000), 1–50.

[15] H. G. Feichtinger, Banach convolution algebras of Wiener type, in *Functions, Series, and Operators*, Vol. I, II, North-Holland, Amsterdam, 1983, pp. 509–524.

[16] V. Ivanov, The Theory of Approximate Methods and Their Application to Numerical Solution of Singular Integral Equations, Nordhoff International, 1976.

[17] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, Springer-Verlag, Berlin, 1998, pp. 195–248.

[18] P. Niyogi, The Informational Complexity of Learning, Kluwer, 1998.

[19] T. Poggio and S. Smale, The mathematics of learning: dealing with data, Notices Amer. Math. Soc. **50** (2003), 537–544.

[20] D. Pollard, Convergence of Stochastic Processes, Springer-Verlag, 1984.

[21] A. N. Shiryaev, Probability, Springer-Verlag, 1996.

[22] S. Smale and D. X. Zhou, Estimating the approximation error in learning theory, Anal. Appl. **1** (2003), 17–41.

[23] G. Strang and G. Fix, A Fourier analysis of the finite element variational method, in *Constructive Aspects of Functional Analysis*, G. Geymonat (ed.), C. I. M. E., 1971, pp. 796–830.

[24] W. C. Sun and X. W. Zhou, On the stability of multivariate trigonometric systems, J. Math. Anal. Appl. **235** (1999), 159–167.

[25] A. Tikhonov and V. Arsenin, Solutions of Ill-posed Problems, W. H. Winston, 1977.

[26] M. Unser, Sampling-50 years after Shannon, Proc. IEEE **88** (2000), 569–587.

[27] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, 1998.

[28] G. Voronoi, Recherches sur les parallelodres primitives, J. Reine Angew. Math. **134** (1908), 198–287.

[29] G. Wahba, Spline Models for Observational Data, SIAM, 1990.

[30] R. Young, An Introduction to Non-Harmonic Fourier Series, Academic Press, 1980.

[31] A. I. Zayed, Advances in Shannon's Sampling Theory, CRC Press, 1993.

[32] D. X. Zhou, The covering number in learning theory, J. Complexity **18** (2002), 739–767.

[33] D. X. Zhou, The capacity of reproducing kernel spaces in learning theory, IEEE Trans. Inform. Theory **49** (2003), 1743–1752.