

GUARANTEED MESHFREE METHODS VIA CONES

FRED J. HICKERNELL

1. BACKGROUND

1.1. Spline Algorithm. Given the data (\mathbf{x}_i, y_i) , $i = 1, \dots, n$, where $y_i = f(\mathbf{x}_i)$, how can we best estimate the function $f : \mathcal{X} \rightarrow \mathbb{R}$? The spline (meshfree, radial basis function, kriging) approach is to construct the approximation as

$$(1a) \quad A_n(f)(\mathbf{x}) = \sum_{i=1}^n c_i K(\mathbf{x}, \mathbf{x}_i) = \mathbf{c}^T \mathbf{k}(\mathbf{x}) = \mathbf{k}^T(\mathbf{x}) \mathbf{c},$$

where $\mathbf{c} = (c_i)_{i=1}^n$, $\mathbf{k}(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}_i))_{i=1}^n$, and $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a *symmetric, positive definite kernel function*, i.e.,

$$K(\mathbf{x}, \mathbf{t}) = K(\mathbf{t}, \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{t} \quad \text{and} \quad \sum_{i,j=1}^m b_i b_j K(\mathbf{t}_i, \mathbf{t}_j) > 0 \quad \forall \mathbf{b} \neq \mathbf{0}, \quad \forall \mathbf{t}_1, \dots, \mathbf{t}_m.$$

A good choice of \mathbf{c} is

$$\mathbf{c} = \mathbf{K}^{-1} \mathbf{y}, \quad \text{where } \mathbf{K} = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^n, \quad \mathbf{y} = (y_i)_{i=1}^n.$$

Note that \mathbf{K} is a *symmetric, positive definite* matrix.

An equivalent way to write this is

$$(1b) \quad A_n(f)(\mathbf{x}) = A_n(f)(\mathbf{x}; \mathbf{y}) = A_n(f)(\mathbf{x}; \mathbf{y}, \mathcal{P}) = \sum_{i=1}^n y_i w_i(\mathbf{x}) = \mathbf{y}^T \mathbf{w}(\mathbf{x}) = \mathbf{w}^T(\mathbf{x}) \mathbf{y},$$

where the *cardinal functions* w_i are given by

$$\mathbf{w}(\mathbf{x}) = \mathbf{w}(\mathbf{x}; \mathcal{P}) = (w_i(\mathbf{x}))_{i=1}^n = \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}),$$

and $\mathcal{P} = \{\mathbf{x}_i\}_{i=1}^n$ is called the *design*. Note that in this form, it is obvious that $A_n(f)$ is *linear* in the function data \mathbf{y} . Thus, the approximation may also be written as

$$(1c) \quad A_n(f)(\mathbf{x}) = \mathbf{y}^T \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) = \mathbf{k}^T(\mathbf{x}) \mathbf{K}^{-1} \mathbf{y}.$$

The spline approximation depends on the function data, the design, and the kernel function.

It is recognized that this is a simplified version of splines. For example, it is often the case that there are low order polynomials included besides the nonparametric kernel part. However, even the simplified framework here is sufficient for posing questions of interest.

1.2. Optimality from a Deterministic Perspective [Fas07, Wen05]. The $A_n(f)$ constructed above has a couple of optimality properties when f is assumed to belong to the Hilbert space, \mathcal{H} , with reproducing kernel, K , i.e.,

$$K(\cdot, \mathbf{t}) \in \mathcal{H} \quad \text{and} \quad f(\mathbf{t}) = \langle K(\cdot, \mathbf{t}), f \rangle_{\mathcal{H}} \quad \forall \mathbf{t} \in \mathcal{X}, \quad f \in \mathcal{H}.$$

The spline, $A_n(f)$, has the property that

$$(2) \quad A_n(f) = \operatorname{argmin}_{g \in \mathcal{H}} \{ \|g\|_{\mathcal{H}} : g(\mathbf{x}_i) = y_i, \quad i = 1, \dots, n \}.$$

Furthermore, if $L : \mathcal{H} \rightarrow \mathbb{R}$ is any bounded linear functional, then

$$(3) \quad L(A_n(f)) = \underset{z}{\operatorname{argmin}} \sup_{g \in \mathcal{H}} \{ |L(g) - z| : g(\mathbf{x}_i) = y_i, \ i = 1, \dots, n, \text{ and } \|g\|_{\mathcal{H}} \leq V \},$$

where the constant $V \geq \|A_n(f)\|_{\mathcal{H}}$. This is an interesting property for, say, $L(f) = f(\mathbf{x})$.

The proof of (2) follows by writing any $g \in \mathcal{H}$ as $A_n(f) + g_{\perp}$, where $g_{\perp} = g - A_n(f)$. Note that by definition g_{\perp} vanishes at the design, $\mathcal{P} = \{\mathbf{x}_i\}_{i=1}^n$, i.e.,

$$g_{\perp}(\mathbf{x}_i) = g(\mathbf{x}_i) - A_n(f)(\mathbf{x}_i) = y_i - y_i = 0, \quad i = 1, \dots, n.$$

Then the square norm of g satisfies

$$\|g\|_{\mathcal{H}}^2 = \langle g, g \rangle_{\mathcal{H}} = \langle A_n(f) + g_{\perp}, A_n(f) + g_{\perp} \rangle_{\mathcal{H}} = \|A_n(f)\|_{\mathcal{H}}^2 + \|g_{\perp}\|_{\mathcal{H}}^2 + 2 \langle A_n(f), g_{\perp} \rangle_{\mathcal{H}}.$$

This last inner product vanishes because

$$\langle A_n(f), g_{\perp} \rangle_{\mathcal{H}} = \langle \mathbf{c}^T \mathbf{k}, g_{\perp} \rangle_{\mathcal{H}} = \sum_{i=1}^n c_i \langle K(\cdot, \mathbf{x}_i), g_{\perp} \rangle_{\mathcal{H}} = \sum_{i=1}^n c_i g_{\perp}(\mathbf{x}_i) = 0.$$

Since

$$\|g\|_{\mathcal{H}}^2 = \|A_n(f)\|_{\mathcal{H}}^2 + \|g_{\perp}\|_{\mathcal{H}}^2,$$

it follows that $\|g\|_{\mathcal{H}}^2$ attains its minimum when g_{\perp} vanishes, i.e., $g = A_n(f)$. As an important aside, note that

$$\begin{aligned} \|A_n(f)\|_{\mathcal{H}}^2 &= \langle A_n(f), A_n(f) \rangle_{\mathcal{H}} = \langle \mathbf{c}^T \mathbf{k}, \mathbf{k}^T \mathbf{c} \rangle_{\mathcal{H}} = \sum_{i,j=1}^n \langle c_i K(\cdot, \mathbf{x}_i), K(\cdot, \mathbf{x}_j) c_j \rangle_{\mathcal{H}} \\ &= \sum_{i,j=1}^n c_i K(\mathbf{x}_i, \mathbf{x}_j) c_j = \mathbf{c}^T \mathbf{K} \mathbf{c} = \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y}. \end{aligned}$$

The proof of (3) follows by considering pairs of functions any $g_{\pm} \in \mathcal{H}$ given by $g_{\pm} = A_n(f) \pm g_{\perp}$, and noting that $\|g_{\pm}\|_{\mathcal{H}}^2 = \|A_n(f)\|_{\mathcal{H}}^2 + \|g_{\perp}\|_{\mathcal{H}}^2$. For any fixed L , z , $\mathbf{x}_1, \dots, \mathbf{x}_n$, and y_1, \dots, y_n , it follows that $A_n(f)$ is then fixed, and one may write

$$\begin{aligned} &\sup_{g \in \mathcal{H}} \{ |L(g) - z| : g(\mathbf{x}_i) = y_i, \ i = 1, \dots, n, \text{ and } \|g\|_{\mathcal{H}} \leq V \} \\ &= \sup_{g_{\perp} \in \mathcal{H}} \left\{ \max_{\pm} |L(g_{\pm}) - z| : g_{\perp}(\mathbf{x}_i) = 0, \ i = 1, \dots, n, \text{ and } \|g_{\perp}\|_{\mathcal{H}}^2 \leq V^2 - \|A_n(f)\|_{\mathcal{H}}^2 \right\} \\ &= \sup_{g_{\perp} \in \mathcal{H}} \left\{ \max_{\pm} |L(A_n(f)) \pm L(g_{\perp}) - z| : g_{\perp}(\mathbf{x}_i) = 0, \ i = 1, \dots, n, \text{ and } \|g_{\perp}\|_{\mathcal{H}}^2 \leq V^2 - \|A_n(f)\|_{\mathcal{H}}^2 \right\} \\ &= |L(A_n(f)) - z| + \sup_{g_{\perp} \in \mathcal{H}} \left\{ |L(g_{\perp})| : g_{\perp}(\mathbf{x}_i) = 0, \ i = 1, \dots, n, \text{ and } \|g_{\perp}\|_{\mathcal{H}}^2 \leq V^2 - \|A_n(f)\|_{\mathcal{H}}^2 \right\}. \end{aligned}$$

Only the first term depends on z , and this is minimized by the choice $z = L(A_n(f))$. This completes the proof of (3). Note that since $L(A_n(f))$ is the best estimate of $L(f)$, and $L(A_n(f))$ is linear in the function data \mathbf{y} , it is also the best linear estimator of $L(f)$.

Continuing the above argument one may derive an error bound by expressing $L(f)$ in terms of the reproducing kernel and applying the Cauchy-Schwartz inequality. Let $\eta \in \mathcal{H}$ be the representer of L , i.e.,

$$L(f) = \langle \eta, f \rangle_{\mathcal{H}} \quad \forall f \in \mathcal{H}.$$

It then follows that

$$\eta(\mathbf{x}) = \langle K(\cdot, \mathbf{x}), \eta \rangle_{\mathcal{H}} = L(K(\cdot, \mathbf{x})).$$

The error of the approximation to $L(f)$ may then be written as

$$\begin{aligned} & \sup_{f \in \mathcal{H}} \{|L(f) - L(A_n(f))| : f(\mathbf{x}_i) = y_i, \ i = 1, \dots, n, \text{ and } \|f\|_{\mathcal{H}} \leq V\} \\ &= \sup_{f \in \mathcal{H}} \left\{ \left| \left\langle \eta - \sum_{i=1}^n L(w_i)K(\cdot, \mathbf{x}_i), f \right\rangle_{\mathcal{H}} \right| : \|f\|_{\mathcal{H}} \leq V \right\} \\ &= \left\| \eta - \sum_{i=1}^n L(w_i)K(\cdot, \mathbf{x}_i) \right\|_{\mathcal{H}} V. \end{aligned}$$

The pieces of this norm are

$$\begin{aligned} \langle \eta, \eta \rangle_{\mathcal{H}} &= L(\eta) = L \circ L^* K(\cdot, \cdot) \\ \left\langle \eta, \sum_{i=1}^n L(w_i)K(\cdot, \mathbf{x}_i) \right\rangle_{\mathcal{H}} &= \sum_{i=1}^n L(w_i) \langle \eta, K(\cdot, \mathbf{x}_i) \rangle_{\mathcal{H}} = \sum_{i=1}^n L(w_i) \eta(\mathbf{x}_i) \\ &= \sum_{i=1}^n L(w_i) L(K(\cdot, \mathbf{x}_i)) = \boldsymbol{\eta}^T \mathbf{K}^{-1} \boldsymbol{\eta} \end{aligned}$$

where $\boldsymbol{\eta} = (L(K(\cdot, \mathbf{x}_i)))_{i=1}^n$. Moreover,

$$\begin{aligned} \left\langle \sum_{i=1}^n L(w_i)K(\cdot, \mathbf{x}_i), \sum_{j=1}^n L(w_j)K(\cdot, \mathbf{x}_j) \right\rangle_{\mathcal{H}} &= \sum_{i,j=1}^n \langle L(w_i)K(\cdot, \mathbf{x}_i), L(w_j)K(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}} \\ &= \sum_{i,j=1}^n L(w_i)K(\mathbf{x}_i, \mathbf{x}_j)L(w_j) = \boldsymbol{\eta}^T \mathbf{K}^{-1} \boldsymbol{\eta}. \end{aligned}$$

Putting all the pieces together gives

$$(4) \quad \sup_{f \in \mathcal{H}} \{|L(f) - L(A_n(f))| : f(\mathbf{x}_i) = y_i, \ i = 1, \dots, n, \text{ and } \|f\|_{\mathcal{H}} \leq V\} = \Phi_L V,$$

$$\text{where } \Phi_L^2 = L \circ L^* K(\cdot, \cdot) - \boldsymbol{\eta}^T \mathbf{K}^{-1} \boldsymbol{\eta}.$$

This may be expressed as a tight error bound

$$(5) \quad |L(f) - L(A_n(f))|^2 \leq \Phi_L^2 \|f - A_n(f)\|_{\mathcal{H}}^2 = \Phi_L^2 \left[\|f\|_{\mathcal{H}}^2 - \|A_n(f)\|_{\mathcal{H}}^2 \right] \leq \Phi_L^2 \|f\|_{\mathcal{H}}^2.$$

The space \mathcal{H} may be decomposed into two orthogonal subspaces, $\mathcal{H}(K_{\mathcal{P}})$ and $\mathcal{H}(K_{\perp})$. Define two kernel functions:

$$K_{\mathcal{P}}(\mathbf{x}, \mathbf{t}) = \mathbf{k}^T(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{t}), \quad K_{\perp}(\mathbf{x}, \mathbf{t}) = K(\mathbf{x}, \mathbf{t}) - K_{\mathcal{P}}(\mathbf{x}, \mathbf{t}) = K(\mathbf{x}, \mathbf{t}) - \mathbf{k}^T(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{t}).$$

These two kernels are symmetric by definition. The first is positive semidefinite because

$$\begin{aligned} \sum_{i,j=1}^n b_i K_{\mathcal{P}}(\mathbf{t}_i, \mathbf{t}_j) b_j &= \sum_{i,j=1}^n b_i \mathbf{k}^T(\mathbf{t}_i) \mathbf{K}^{-1} \mathbf{k}(\mathbf{t}_j) b_j \\ &= \mathbf{b}^T \left((K(\mathbf{x}_i, \mathbf{t}_j))_{i,j=1}^{n,m} \right)^T \mathbf{K}^{-1} (K(\mathbf{x}_i, \mathbf{t}_j))_{i,j=1}^{n,m} \mathbf{b} \geq 0, \quad \forall \mathbf{b}, \ \forall \mathbf{t}_1, \dots, \mathbf{t}_m. \end{aligned}$$

For each fixed \mathbf{t} , $K_{\mathcal{P}}(\cdot, \mathbf{t})$ is a linear combination of the functions comprising the vector \mathbf{k} . Thus, $\mathcal{H}(K_{\mathcal{P}})$ consists of the span of the components of \mathbf{k} . Notice that for all $i, j = 1, \dots, n$, it follows that $K_{\mathcal{P}}(\mathbf{x}, \mathbf{x}_i) = \mathbf{k}^T(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_i) = \mathbf{k}^T(\mathbf{x}) \mathbf{e}_i = K(\mathbf{x}, \mathbf{x}_i)$, and

$$\begin{aligned} \langle K(\cdot, \mathbf{x}_i), K(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}(K_{\mathcal{P}})} &= \langle K_{\mathcal{P}}(\cdot, \mathbf{x}_i), K_{\mathcal{P}}(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}(K_{\mathcal{P}})} = K_{\mathcal{P}}(\mathbf{x}_i, \mathbf{x}_j) \\ &= K(\mathbf{x}_i, \mathbf{x}_j) = \langle K(\cdot, \mathbf{x}_i), K(\cdot, \mathbf{x}_j) \rangle_{\mathcal{H}}. \end{aligned}$$

Thus, $\mathcal{H}(K_{\mathcal{P}})$ is a subspace of \mathcal{H} .

Let \mathcal{H}_{\perp} denote the orthogonal complement of $\mathcal{H}(K_{\mathcal{P}})$. It is now shown that K_{\perp} as defined above is the reproducing kernel for this Hilbert space. To demonstrate this, it is first shown that $K_{\perp}(\cdot, \mathbf{x}) \in \mathcal{H}_{\perp}$ by showing that $K_{\perp}(\cdot, \mathbf{x})$ is perpendicular to all functions residing in $\mathcal{H}(K_{\mathcal{P}})$. Such functions take the form $\mathbf{b}^T \mathbf{k}$ for some \mathbf{b} , and

$$\langle \mathbf{b}^T \mathbf{k}(\cdot), K_{\perp}(\cdot, \mathbf{x}) \rangle_{\mathcal{H}} = \langle \mathbf{b}^T \mathbf{k}(\cdot), K(\cdot, \mathbf{x}) - \mathbf{k}^T(\cdot) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) \rangle_{\mathcal{H}} = \mathbf{b}^T \mathbf{k}(\mathbf{x}) - \mathbf{b}^T \mathbf{K} \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}) = 0.$$

Now consider any $f \in \mathcal{H}_{\perp}$. This means that $\langle K(\cdot, \mathbf{x}_i), f \rangle_{\mathcal{H}} = 0$ for $i = 1, \dots, n$, and so,

$$f(\mathbf{x}) = \langle K(\cdot, \mathbf{x}), f \rangle_{\mathcal{H}} = \langle K(\cdot, \mathbf{x}) - \mathbf{k}(\cdot) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}), f \rangle_{\mathcal{H}} = \langle K_{\perp}, f \rangle_{\mathcal{H}_{\perp}}.$$

Thus, K_{\perp} obeys the reproducing property for \mathcal{H}_{\perp} , and thus must be its reproducing kernel.

2. ERROR ESTIMATION VIA CONES

Let $L_{\mathbf{x}}$ be the evaluation functional, i.e., $L_{\mathbf{x}}(f) = f(\mathbf{x})$. For this case we have an error bound

$$\begin{aligned} (6) \quad & \sup_{0 \neq f \in \mathcal{H}} \frac{|f(\mathbf{x}) - A_n(f)(\mathbf{x})|^2}{\|f\|_{\mathcal{H}}^2} \rho(\mathbf{x}) \, d\mathbf{x} \\ & \leq \int_{\mathcal{X}} \Phi_{L_{\mathbf{x}}}^2 \rho(\mathbf{x}) \, d\mathbf{x} \\ & = \int_{\mathcal{X}} [K(\mathbf{x}, \mathbf{x}) - \mathbf{k}^T(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x})] \rho(\mathbf{x}) \, d\mathbf{x} \\ & = \int_{\mathcal{X}} K(\mathbf{x}, \mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x} - \text{trace}(\mathbf{K}^{-1} \tilde{\mathbf{K}}) \\ & =: h^2(n) \end{aligned}$$

where $\tilde{\mathbf{K}} = \int_{\mathcal{X}} \mathbf{k}(\mathbf{x}) \mathbf{k}^T(\mathbf{x}) \rho(\mathbf{x}) \, d\mathbf{x}$. Note that the left hand side is no smaller than

$$\sup_{0 \neq f \in \mathcal{H}} \frac{\|f - A_n(f)\|_{\mathcal{L}_2}}{\|f\|_{\mathcal{H}}}.$$

Now suppose that $T : \mathcal{H} \rightarrow \mathcal{L}_2 = \mathcal{L}_2(\mathcal{X})$ is some bounded linear operator, i.e., $\|T(f)\|_{\mathcal{L}_2} \leq C \|f\|_{\mathcal{H}}$ for all $f \in \mathcal{H}$. Here the \mathcal{L}_2 -norm is defined in terms of a weight, ρ :

$$\|f\|_{\mathcal{L}_2}^2 = \int_{\mathcal{X}} |f(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x}.$$

The linear functional $T_{\mathbf{x}} : f \mapsto T(f)(\mathbf{x})$ is assumed to be bounded, and its square norm is $\|T_{\mathbf{x}}\| = \sqrt{T_{\mathbf{x}} T_{\mathbf{x}}^* K(\cdot, \cdot)}$. Note that

$$\begin{aligned} \|T\|^2 &= \sup_{\|f\|_{\mathcal{H}} \leq 1} \|T(f)\|_{\mathcal{L}_2}^2 = \sup_{\|f\|_{\mathcal{H}} \leq 1} \int_{\mathcal{X}} |T(f)(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x} \\ &\leq \int_{\mathcal{X}} \sup_{\|f\|_{\mathcal{H}} \leq 1} |T(f)(\mathbf{x})|^2 \rho(\mathbf{x}) \, d\mathbf{x} = \int_{\mathcal{X}} \|T_{\mathbf{x}}\|^2 \rho(\mathbf{x}) \, d\mathbf{x} = \| \|T_{\cdot}\| \|_{\mathcal{L}_2}^2. \end{aligned}$$

It is assumed that the right hand side is finite. Let $\xi(\cdot, \mathbf{x})$ denote the representer for $T_{\mathbf{x}}$. This means that $\xi(\mathbf{t}, \mathbf{x}) = T_{\mathbf{x}}(K(\cdot, \mathbf{t})) = T_{\mathbf{x}}(K(\mathbf{t}, \cdot))$.

Let $\|\cdot\|_{\tilde{\mathcal{H}}}$ be a weaker norm defined by $\|f\|_{\tilde{\mathcal{H}}} = \|T(f)\|_{\mathcal{L}_2}$. For a fixed design \mathcal{P} , define the algorithm for estimating the weaker norm by $\tilde{H}_n(f) = \|T(A_n(f))\|_{\mathcal{L}_2}$. Defining

$$\boldsymbol{\xi}(\mathbf{x}) = (\xi(\mathbf{x}, \mathbf{x}_i))_{i=1}^n = (T_{\mathbf{x}} K(\cdot, \mathbf{x}_i))_{i=1}^n,$$

it follows that $T(A(f))(x) = \mathbf{c}^T \boldsymbol{\xi}(x)$, and so

$$\tilde{H}_n^2(f) = \|\mathbf{c}^T \boldsymbol{\xi}(\cdot)\|_{\mathcal{L}_2}^2 = \mathbf{y}^T \mathbf{K}^{-1} \tilde{\mathbf{H}} \mathbf{K}^{-1} \mathbf{y}, \quad \text{where } \tilde{\mathbf{H}} := \int_{\mathcal{X}} \boldsymbol{\xi}(x) \boldsymbol{\xi}^T(x) \rho(x) dx.$$

By the work in the previous section it follows that

$$\begin{aligned} \left| \|f\|_{\tilde{\mathcal{H}}} - \tilde{H}_n(f) \right|^2 &= \left| \|T(f)\|_{\mathcal{L}_2} - \|T(A_n(f))\|_{\mathcal{L}_2} \right|^2 \\ &\leq \|T(f) - T(A_n(f))\|_{\mathcal{L}_2}^2 = \int_{\mathcal{X}} [T(f)(x) - T(A_n(f))(x)]^2 \rho(x) dx \\ &\leq \|f\|_{\mathcal{H}}^2 \int_{\mathcal{X}} \Phi_{T_x}^2 \rho(x) dx = \|\Phi_T\|_{\mathcal{L}_2}^2 \|f\|_{\mathcal{H}}^2. \end{aligned}$$

Let $\zeta(\mathbf{t}, \mathbf{x}) = T_{\mathbf{t}} T_{\mathbf{x}} K(\cdot, \cdot)$. Note that the $\boldsymbol{\eta}$ above becomes $\boldsymbol{\xi}(x)$ for $L = T_x$. Then one may write

$$\Phi_{T_x}^2 = \zeta(x, x) - \boldsymbol{\xi}^T(x) \mathbf{K}^{-1} \boldsymbol{\xi}(x).$$

This error bound may be written as

$$\begin{aligned} \tilde{h}^2(n) &:= \int_{\mathcal{X}} \Phi_{T_x}^2 \rho(x) dx \\ &= \int_{\mathcal{X}} [\zeta(x, x) - \boldsymbol{\xi}^T(x) \mathbf{K}^{-1} \boldsymbol{\xi}(x)] \rho(x) dx \\ &= \int_{\mathcal{X}} \zeta(x, x) \rho(x) dx - \text{trace}(\mathbf{K}^{-1} \tilde{\mathbf{H}}) \end{aligned}$$

Let $\mathcal{C}_\tau = \{f \in \mathcal{H} : \|f\|_{\mathcal{H}} \leq \tau \|f\|_{\tilde{\mathcal{H}}}\}$. Then for all $f \in \mathcal{C}_\tau$,

$$\left| \|f\|_{\tilde{\mathcal{H}}} - \tilde{H}_n(f) \right| \leq \tilde{h}(n) \|f\|_{\mathcal{H}} \leq \tau \tilde{h}(n) \|f\|_{\tilde{\mathcal{H}}},$$

which implies

$$\|f\|_{\tilde{\mathcal{H}}} \leq \frac{\tilde{H}_n(f)}{1 - \tau \tilde{h}(n)}, \quad \|f\|_{\mathcal{H}} \leq \frac{\tau \tilde{H}_n(f)}{1 - \tau \tilde{h}(n)},$$

given that $1 - \tau \tilde{h}(n) > 0$. This implies that

$$\left[\int_{\mathcal{X}} \sup_{\substack{g \in \mathcal{H} \\ \|g\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} \\ g(\mathbf{x}_i) = f(\mathbf{x}_i)}} |g(x) - A_n(f)(x)|^2 \rho(x) dx \right]^{1/2} \leq \frac{\tau h(n) \tilde{H}_n(f)}{1 - \tau \tilde{h}(n)}$$

Algorithm 1. Given $\tau > 0$, $\varepsilon > 0$, and $f \in \mathcal{C}_\tau$, choose n_1 such that $\tilde{h}(n_1) < 1/\tau$

Stage 1: Check if $\tau h(n_i) \tilde{H}_{n_i}(f) \leq \varepsilon[1 - \tau \tilde{h}(n_i)]$. If so, return the answer $A_{n_i}(f)$.

Stage 2: If not, increment i by one and return to Stage 1.

3. EXAMPLE

3.1. Example 1. Consider the univariate case and let $\mathcal{H} = \{f | f(0) = 0, f \text{ is absolutely continuous and } f' \in \mathcal{L}_2((0, 1))\}$. Define the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ as

$$\langle f, g \rangle_{\mathcal{H}} = \frac{1}{\gamma} \int_0^1 f'(x) g'(x) dx.$$

It can be easily verified that \mathcal{H} is a reproducing kernel Hilbert space with the reproducing kernel

$$K(x, y) = \gamma \min(x, y).$$

Suppose we take T to be the embedding operator, i.e., $T : \mathcal{H} \rightarrow \mathcal{L}_2((0, 1))$ is defined by $T(f) = f$. Then $\tilde{H} = \tilde{K}$, $\tilde{h} = h$, and hence the error bound in the previous section becomes

$$(7) \quad \left[\int_0^1 |f(x) - A_n(f)(x)|^2 dx \right]^{1/2} \leq \frac{\tau h(n) \tilde{H}(f)}{1 - \tau h(n)}.$$

As a side note, the i, j -th entry of \tilde{K} is

$$\begin{aligned} \int_0^1 K(x, x_i) K(x, x_j) dx &= \gamma^2 \int_0^1 \min(x, x_i) \min(x, x_j) dx \\ &= \gamma^2 \left(\frac{\min^3(x_i, x_j)}{3} + \frac{\min(x_i, x_j)(\max^2(x_i, x_j) - \min^2(x_i, x_j))}{2} + x_i x_j (1 - \max(x_i, x_j)) \right). \end{aligned}$$

Let the test function $f \in \mathcal{H}$ be given by $f(x) = \sin(ax)$, $a \in \mathbb{R}$. We can compute that

$$\|f\|_{\mathcal{H}} = \left(\frac{1}{\gamma} \int_0^1 a^2 \cos^2(ax) dx \right)^{1/2} = a \sqrt{\frac{1}{\gamma} \left(\frac{1}{2} + \frac{\sin(2a)}{4a} \right)},$$

and that

$$\|f\|_{\tilde{\mathcal{H}}} = \|f\|_{\mathcal{L}_2} = \left(\int_0^1 \sin^2(ax) dx \right)^{1/2} = \sqrt{\frac{1}{2} - \frac{\sin(2a)}{4a}}.$$

We want $f \in \mathcal{C}_\tau$, or equivalently, $\|f\|_{\mathcal{H}} \leq \tau \|f\|_{\tilde{\mathcal{H}}}$, which implies that

$$a \sqrt{\frac{1}{\gamma} \left(\frac{1}{2} + \frac{\sin(2a)}{4a} \right)} \leq \tau \sqrt{\frac{1}{2} - \frac{\sin(2a)}{4a}}.$$

Solving this inequality for τ , we have that

$$\tau \geq a \sqrt{\frac{2a + \sin(2a)}{\gamma(2a - \sin(2a))}}.$$

For example, if we take $\gamma = 1$ and $a = \pi$, then $\tau = \pi$ such that $f \in \mathcal{C}_\tau$. Once τ is chosen, we can proceed to compute the error bound in (7).

3.2. Example 2. Consider the univariate case and define the Gaussian kernel $\Phi(x-t) = K(x, t) = \exp(-\gamma^2(x-t)^2)$. Let $\mathcal{H} = \{f \in \mathcal{L}_2(\mathbb{R}) \cap C(\mathbb{R}) : \}$ and define the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ as

$$\langle f, g \rangle_{\mathcal{H}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{\hat{f}(\omega) \overline{\hat{g}(\omega)}}{\hat{\Phi}(\omega)} d\omega.$$

Then by Theorem 13.4 from [Fas07] \mathcal{H} is a reproducing kernel Hilbert space with the kernel $K(x, y)$. We take again T to be the embedding operator and let the test function $f \in \mathcal{H}$ be given by

$$f(x) = \begin{cases} 0 & x \leq 1 \\ \frac{1}{2}x^2 - x + \frac{1}{2} & 1 < x \leq 4 \\ -\frac{1}{2}x^2 + 7x - \frac{31}{2} & 4 < x \leq 10 \\ \frac{1}{2}x^2 - 13x + \frac{169}{2} & 10 < x \leq 13 \\ 0 & x > 13 \end{cases}.$$

If we take the density function

$$\rho(x) = \begin{cases} 1/\pi & 0 \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases},$$

we can proceed to compute $\|f\|_{\mathcal{H}}$. However in this case since ρ has the support $[0, \pi]$, to construct a tighter error bound, instead of computing $\|f\|_{\mathcal{H}}$, we only need to find τ such that

$$\left(\frac{1}{\sqrt{2\pi}} \int_0^\pi \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega)} d\omega \right)^{1/2} \leq \tau \|f\|_{\tilde{\mathcal{H}}}.$$

If we take $\gamma = 1$, then we compute that

$$\begin{aligned} \left(\frac{1}{\sqrt{2\pi}} \int_0^\pi \frac{|\hat{f}(\omega)|^2}{\hat{\Phi}(\omega)} d\omega \right)^{1/2} &\simeq 1.3833 \\ \text{and } \|f\|_{\tilde{\mathcal{H}}} &= \int_0^\pi \frac{1}{\pi} \left(\frac{1}{2}x^2 - x + \frac{1}{2} \right)^2 \simeq .8561. \end{aligned}$$

We can choose $\tau = 2$, which satisfies the above inequality and proceed to compute the error bound in (7).

REFERENCES

- [Fas07] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*, Interdisciplinary Mathematical Sciences, vol. 6, World Scientific Publishing Co., Singapore, 2007.
- [Wen05] H. Wendland, *Scattered data approximation*, Cambridge Monographs on Applied and Computational Mathematics, no. 17, Cambridge University Press, Cambridge, 2005.