Monograph:
Dodd, T.J. and Harrison, R.F. (2002) Some lemmas on reproducing kernel Hilbert spaces. Research Report. ACSE Research Report no. 819 . Automatic Control and System Engineering, University of Sheffield

Some Lemmas on Reproducing Kernel Hilbert Spaces

Tony J. Dodd and Robert F. Harrison
Department of Automatic Control and Systems Engineering
The University of Sheffield, Sheffield S1 3JD, UK
e-mail: {t.j.dodd, r.f.harrison}@shef.ac.uk

Research Report No. 819
March 2002

Abstract

Reproducing kernel Hilbert spaces (RKHS) provide a framework for approximation from finite data using the idea of bounded linear functionals. The approximation problem in this case can be viewed as the inverse problem of finding the optimum operator from the Euclidean space of observations to some subspace of the RKHS. In constructing the appropriate inverse operator use is made of both adjoint operators in RKHS and various norms. In this report a number of lemmas are given with respect to such adjoint operators and norms.

1 Introduction

Reproducing kernel Hilbert spaces (RKHS) provide a general framework for approximation of functions utilising only finite observations of the function. The important point is that in a RKHS we can express point evaluations of the function as inner products. These point evaluations are determined by a bounded linear operator for which the inner product is guaranteed to exist by the Riesz representation theorem. The finite observations determine a finite approximating subspace of the RKHS. A unique minimum norm approximation to the original function, within this approximating subspace, is guaranteed to exist. This approximation can be found using least squares or orthogonal projections under the action of the generalised inverse. It is also possible to construct regularised solutions.

A key feature of the solutions is the need exactly to determine the adjoint operator to the observation operator and compositions of these. These are required for practical implementations of the approximation solutions. These implementations may be batch or iterative. Further, in certain iterative solutions it is necessary to calculate various norms over functions in RKHS, possibly under the action of the observation operator and its adjoint. Fortunately, within a RKHS framework with finite observations it is possible to write down, in a practical way, what the adjoint operator, the composition operator, and various norms correspond to. In this report various lemmas will be described which provide the necessary results.

In the next section the general approximation problem with finite observations will be described. RKHS are introduced in Section 3 and the lemmas concerning adjoint operators are stated and proven in Section 4. Finally, various norm lemmas are proven in Section 5.

2 Approximation with Finite Observations

We assume that we have some unknown function, f, of interest but that we are able to observe its behaviour. The function belongs to some Hilbert space, $\mathcal{F}$, defined on some parameter set, $\mathcal{X}$. This set can be considered as an input set in the sense that for $x \in \mathcal{X}$, $f(x)$ represents the evaluation of f at x.

A finite set of observations $\{z_i\}_{i=1}^N$ of the function is made corresponding to inputs $\{x_i\}_{i=1}^N$. It is assumed that the space of all possible observations $\mathcal{Z} \subset \mathbb{R}^N$. Neglecting the effects of errors, the observations arise as follows

$$z_i = L_i f \tag{1}$$

where $\{L_i\}_{i=1}^N$ is a set of linear evaluation functionals, defined on $\mathcal{F}$, which associate real numbers to the function f. We can represent the complete set of observations, $[z_1, \dots, z_N]^T$, in vector form as follows

$$z = Lf = \sum_{i=1}^{N} (L_i f) e_i \tag{2}$$

where $e_i \in \mathbb{R}^N$ is the ith standard basis vector.

In general L_i permits indirect observation (e.g. via derivatives of f), but we are concerned with the case

$$z_i = f(x_i) \tag{3}$$

leading to the exact interpolation problem.

The approximation problem can then be formulated as follows (Bertero, De Mol, and Pike 1985): given a class, $\mathcal{F}$, of functions, and a set $\{z_i\}_{i=1}^N$ of values of linear functionals $\{L_i\}_{i=1}^N$ defined on $\mathcal{F}$, find in $\mathcal{F}$ a function, f, which satisfies Eq. 1.

By assuming that $\mathcal{F}$ is a Hilbert space, and further, the $\{L_i\}_{i=1}^N$ are continuous (hence bounded), it follows from the Riesz representation theorem that we can express the observations as (Akhiezer and Glazman 1981)

$$L_i f = \langle f, \psi_i \rangle_{\mathcal{F}}, \quad i = 1, \dots, N \tag{4}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ denotes the inner product in $\mathcal{F}$. The $\{\psi_i\}_{i=1}^N$ are a set of functions each belonging to $\mathcal{F}$ and uniquely determined by the functionals $\{L_i\}_{i=1}^N$.

The approximation problem can now be stated as follows: given the Hilbert space of functions, $\mathcal{F}$, the set of functions, $\{\psi_i\}_{i=1}^N \subset \mathcal{F}$, and the observations, $\{z_i\}_{i=1}^N$, find a function, $f \in \mathcal{F}$, such that Eq. 4 is satisfied.

Without proof we now state the solution to the approximation problem in a least squares sense (Groetsch 1977). Given the observations, the approximation of minimum norm is given by

$$\hat{f} = (L^*L)^{-1}L^*z = L^*(LL^*)^{-1}z \tag{5}$$

where

$$\hat{f} = \arg\min_{f \in \mathcal{F}} ||Lf - z||_{\mathcal{Z}}.$$

Similarly we can find a regularised solution to

$$\hat{f}_{reg} = \arg\min_{f \in \mathcal{F}} \left\{ ||Lf - z||_{\mathcal{Z}}^2 + \rho ||f||_{\mathcal{F}}^2 \right\}.$$

as

$$\hat{f}_{reg} = (L^*L + \rho I)^{-1}L^*z = L^*(LL^* + \rho I)^{-1}z. \tag{6}$$

Note that, in each case, in order to calculate the approximation, it is necessary to use the adjoint operator L^* (defined in Section 4) and compositions of L and L^*.

3 Reproducing Kernel Hilbert Spaces

Formally a RKHS is a Hilbert space of functions on some parameter set, $\mathcal{X}$, with the property that for each $x \in \mathcal{X}$ the evaluation functional, L_i, which associates

f with $f(x_i)$, $L_i f \to f(x_i)$, is a bounded linear functional (Wahba 1990). The boundedness means that there exists a scalar M such that

$$|L_i f| = |f(x_i)| \leq M||f||_{\mathcal{F}} \text{ for all } f \text{ in the RKHS}$$

where $||\cdot||_{\mathcal{F}}$ is the norm in the Hilbert space. To satisfy the Riesz representation theorem the L_i must be bounded hence any Hilbert space satisfying the Riesz theorem will be a RKHS.

We use $k(x_i, \cdot)$ to refer to ψ_i (i.e. the evaluation of the function $k(x_i, \cdot) = \psi_i$ at x_j is $k(x_i, x_j)$). The inner product $\langle k(x_i, \cdot), k(x_j, \cdot)\rangle_{\mathcal{F}}$ must equal $k(x_i, x_j)$ by the Riesz representation theorem. This leads to the following important result: $k(x_i, x_j)$ is positive definite since, for any $x_1, \ldots, x_n \in \mathcal{X}$, $a_i, \ldots, a_n \in \mathbb{R}$,

$$\begin{aligned} \sum_{i,j} a_i a_j k(x_i . x_j) &= \sum_{i,j} a_i a_j \langle k(x_i, \cdot), k(x_j, \cdot)\rangle_{\mathcal{F}} \\ &= \left\| \sum a_i k(x_i, \cdot) \right\|_{\mathcal{F}}^2 \geq 0 \end{aligned}$$

where $||\cdot||_{\mathcal{F}}$ is the corresponding norm in the RKHS. The following is then a standard theorem on RKHS.

Theorem 3.1 *(Aronszajn 1950) To every RKHS there corresponds a unique positive-definite function (the reproducing kernel) and conversely given a positive-definite function k on $\mathcal{X} \times \mathcal{X}$ we can construct a unique RKHS of real-valued functions on $\mathcal{X}$ with k as its reproducing kernel.*

We then have a more common definition for RKHS.

Definition 3.1 *(Parzen 1961) A Hilbert space $\mathcal{F}$ is said to be a reproducing kernel Hilbert space, with reproducing kernel k, if the members of $\mathcal{F}$ are functions on some set, $\mathcal{X}$, and if there is a kernel, k, on $\mathcal{X} \times \mathcal{X}$ having the following two properties; for every $x \in \mathcal{X}$ (where $k(\cdot, x_2)$ is the function defined on $\mathcal{X}$, with value at x_1 in $\mathcal{X}$ equal to $k(x_1, x_2)$):*

1. $k(\cdot, x_2) \in \mathcal{F}$; *and*
2. $\langle f, k(\cdot, x_2)\rangle_{\mathcal{F}} = f(x_2)$

for every f in $\mathcal{F}$.

We can then associate with $k(\cdot, \cdot)$ a unique collection of functions of the form

$$f(\cdot) = \sum_{i=1}^{N} c_i k(x_i, \cdot) \tag{7}$$

for $N \in \mathbb{Z}^+$ and $c_i \in \mathbb{R}$. Strictly this defines a finite dimensional subspace, $\mathcal{F}_N$, of $\mathcal{F}$. A well defined inner product for this collection is (Wahba 1990)

$$\left\langle \sum_{i=1}^{N} a_i k(x_i, \cdot), \sum_{j=1}^{N} b_j k(x_j, \cdot) \right\rangle_{\mathcal{F}} = \sum_{i,j=1}^{N} a_i b_j \langle k(x_i, \cdot), k(x_j, \cdot) \rangle_{\mathcal{F}} = \sum_{i,j=1}^{N} a_i b_j k(x_i, x_j). \tag{8}$$

For this collection, norm convergence implies pointwise convergence and we can therefore adjoin all limits of Cauchy sequences of functions which are well defined as pointwise limits (Wahba 1990). The resulting Hilbert space is then a RKHS.

4 Adjoint Operators in RKHS

In this section we prove various lemmas relating to the observation operator, L, and its adjoint, L^*, defined as follows.

Definition 4.1 *For some bounded linear operator $L : \mathcal{F} \to \mathcal{Z}$, where $\mathcal{F}$ and $\mathcal{Z}$ are Hilbert spaces, the adjoint operator L^* of L is the operator*

$$L^* : \mathcal{Z} \to \mathcal{F} \tag{9}$$

such that, for all $f \in \mathcal{F}$ and $z \in \mathcal{Z}$,

$$\langle Lf, z \rangle_{\mathcal{Z}} = \langle f, L^* z \rangle_{\mathcal{F}}. \tag{10}$$

We have already seen that the adjoint operator plays a role in constructing approximations of functions in Section 2. We therefore make the following additional assumptions regarding the operator L and the spaces $\mathcal{F}$ and $\mathcal{Z}$.

Assumption 4.1 *$\mathcal{Z} \subset \mathbb{R}^N$ with inner product $\langle g, h \rangle_{\mathcal{Z}} = \sum_{i=1}^{N} g_i h_i$, for any $g, h \in \mathcal{Z}$.*

Assumption 4.2 *$\mathcal{F}$ is a RKHS with reproducing kernel $k(\cdot, \cdot)$ and inner product given by Eq. 8.*

Assumption 4.3 *The operator L acting on f has the form*

$$Lf = \sum_{i=1}^{N} \langle f, k(x_i, \cdot) \rangle_{\mathcal{F}} \cdot e_i$$

where $e_i \in \mathbb{R}^N$ is the ith standard basis vector.

The following results then apply to the operator L and its adjoint L^*.

Lemma 4.1 *The adjoint operator L^* is given by*

$$L^* z = \sum_{i=1}^{N} z_i k(x_i, \cdot). \tag{11}$$

Proof Solving for the LHS of Eq. 10

$$\langle Lf, z \rangle_{\mathcal{Z}} = \sum_{i=1}^{N} \langle f, k(x_i, \cdot) \rangle_{\mathcal{F}} \cdot z_i = \sum_{i=1}^{N} f(x_i) z_i. \tag{12}$$

By assumption we set $L^* z = \sum_{i=1}^{N} z_i k(x_i, \cdot)$ and solving for the RHS of Eq. 10

$$\langle f, L^* z \rangle_{\mathcal{F}} = \left\langle f, \sum_{i=1}^{N} z_i k(x_i, \cdot) \right\rangle_{\mathcal{F}} = \sum_{i=1}^{N} z_i \langle f, k(x_i, \cdot) \rangle_{\mathcal{F}} \tag{13}$$

the latter owing to the linearity property of the inner product. But this is simply equal to $\sum_{i=1}^{N} z_i f(x_i)$ by the reproducing property in a RKHS. □

Lemma 4.2 *For the operator LL^* we have*

$$LL^* z = \sum_{j=1}^{N} \sum_{i=1}^{N} k(x_i, x_j) e_j z_i.$$

Proof The operator LL^* acting on z can be expressed, using the previous results, as follows:

$$\begin{aligned} LL^* z &= L\left(\sum_{i=1}^{N} z_i k(x_i, \cdot) \right) \\ &= \sum_{j=1}^{N} \left\langle \sum_{i=1}^{N} z_i k(x_i, \cdot), k(x_j, \cdot) \right\rangle_{\mathcal{F}} \cdot e_j \end{aligned}$$

using the definition of L. As $z_i \notin \mathcal{F}$ we can write this as

$$\begin{aligned} LL^* z &= \sum_{j=1}^{N} \sum_{i=1}^{N} z_i \langle k(x_i, \cdot), k(x_j, \cdot) \rangle_{\mathcal{F}} \cdot e_j \\ &= \sum_{j=1}^{N} \sum_{i=1}^{N} z_i k(x_i, x_j) e_j. \end{aligned}$$

□

Since $LL^* : \mathcal{Z} \to \mathcal{Z}$ has domain and range equal to a finite dimensional space we can express LL^* as the matrix $LL^* = \sum_{j=1}^{N} \sum_{i=1}^{N} k(x_i, x_j) e_j e_i^T$. This is equivalent to $LL^* = K$ where K is defined as the matrix $[K]_{ij} = k(x_i, x_j)$.

Lemma 4.3 *The operator L^*L is given by*

$$L^*Lf = \sum_{i=1}^{N} f(x_i)k(x_i, \cdot).$$

Proof Using the result in Lemma 4.1 and the definition of the operator, L, we have

$$\begin{aligned} L^*Lf &= L^*\left(\sum_{i=1}^{N} \langle f, k(x_i, \cdot)\rangle_{\mathcal{F}} \cdot e_i\right) \\ &= L^*\left(\sum_{i=1}^{N} f(x_i)e_i\right) \\ &= \sum_{i=1}^{N} f(x_i)k(x_i, \cdot). \end{aligned}$$

□

5 Norms in RKHS

Using the assumptions and results of the previous sections we now provide results on various norms in connection with functions in RKHS. These norms are of use in quantifying the errors in approximations and also for calculating learning rates in iterative approximation schemes.

Lemma 5.1 *For any function $f \in \mathcal{F}$ expressed in the form $f = \sum_{i=1}^{N} a_i k(x_i, \cdot)$ its norm is given by*

$$||f||_{\mathcal{F}}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j k(x_i, x_j).$$

Proof Using the definition of an inner product in a RKHS (Eq. 8) we have

$$\begin{aligned} ||f||_{\mathcal{F}}^2 &= \left\langle \sum_{i=1}^{N} a_i k(x_i, \cdot), \sum_{j=1}^{N} a_j k(x_j, \cdot) \right\rangle_{\mathcal{F}} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \langle k(x_i, \cdot), k(x_j, \cdot)\rangle_{\mathcal{F}} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j k(x_i, x_j) \end{aligned}$$

owing to the reproducing property of the kernel. □

Note that this norm can also be expressed as

$$||f||^2_{\mathcal{F}} = a^T K a \tag{14}$$

where K is the usual kernel (Gram) matrix defined by $[K]_{ij} = k(x_i, x_j)$ and a is the vector $[a_1, \dots, a_N]^T$.

Lemma 5.2 *Under the action of the operator L the norm is given by*

$$||Lf||^2_{\mathcal{Z}} = \sum_{i=1}^{N} f^2(x_i)$$

where $f^2(x_i) = (f(x_i))^2$.

Proof Using the definition of the operator L this follows immediately from

$$\begin{aligned} ||Lf||^2_{\mathcal{Z}} &= \left\| \sum_{i=1}^{N} (L_i f) e_i \right\|^2_{\mathcal{Z}} \\ &= \left\| \sum_{i=1}^{N} f(x_i) e_i \right\|^2_{\mathcal{Z}} \\ &= \sum_{i=1}^{N} f^2(x_i). \end{aligned}$$

□

Defining $f_N = \sum_{i=1}^{N} f(x_i)e_i = [f(x_1), \dots, f(x_N)]^T$ this norm can also be expressed as

$$||Lf||^2_{\mathcal{Z}} = f_N^T f_N. \tag{15}$$

Lemma 5.3 *For any $z \in \mathbb{R}^N$*

$$||LL^* z||^2_{\mathcal{Z}} = \sum_{i=1}^{N} \sum_{j=1}^{N} z_i z_j k^2(x_i, x_j)$$

where $k^2(x_i, x_j) = (k(x_i, x_j))^2$.

Proof Straightforward from Lemma 4.2

$$\begin{aligned} ||LL^* z||^2_{\mathcal{Z}} &= \left\| \sum_{j=1}^{N} \sum_{i=1}^{N} z_i k(x_i, x_j) e_j \right\|^2_{\mathcal{Z}} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} z_i z_j k(x_i, x_j) k(x_i, x_j). \end{aligned}$$

□

The norm can also be expressed in matrix-vector form as

$$||LL^* z||^2_{\mathcal{Z}} = ||Kz||^2_{\mathcal{Z}} = z^T K^T K z = z^T K^2 z. \tag{16}$$

Lemma 5.4 *For any $f \in \mathcal{F}$*

$$||L^*Lf||^2_{\mathcal{F}} = \sum_{i=1}^{N}\sum_{j=1}^{N} f(x_i)f(x_j)k(x_i, x_j)$$

Proof Using Lemma 4.3 and the definition of the inner product, Eq. 8,

$$\begin{aligned} ||L^*Lf||^2_{\mathcal{F}} &= \left\langle \sum_{i=1}^{N} f(x_i)k(x_i,\cdot), \sum_{j=1}^{N} f(x_j)k(x_j,\cdot) \right\rangle_{\mathcal{F}} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} f(x_i)f(x_j)\langle k(x_i,\cdot), k(x_j,\cdot)\rangle_{\mathcal{F}} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} f(x_i)f(x_j)k(x_i, x_j). \end{aligned}$$

□

Using f_N as defined above this norm can also be expressed as

$$||L^*Lf||^2_{\mathcal{F}} = f_N^T K f_N. \tag{17}$$

6 Conclusions

Reproducing kernel Hilbert spaces provide an important framework for the approximation of functions given finite observations of the function. The approximations in the least squares sense and in a certain regularised case are given in terms of an adjoint operator. Various lemmas were proven relating to the adjoint operator and associated norms. In the case where the RKHS is finite dimensional closed form practical equations exist for these which allow us to calculate exactly the associated approximations in RKHS from finite data.

Acknowledgements

The authors would like to thank the UK EPSRC for their financial support under Grant No. GR/R15726/01.

References

Akhiezer, N. and I. Glazman (1981). *Theory of Linear Operators in Hilbert Space*, Volume I. Pitman.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society 68*, 337–404.

Bertero, M., C. De Mol, and E. Pike (1985). Linear inverse problems with discrete data. I: General formulation and singular system analysis. *Inverse Problems 1*, 301–330.

Groetsch, C. (1977). *Generalized Inverses of Linear Operators.* Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker.

Parzen, E. (1961). An approach to time series analysis. *Annals of Mathematical Statistics 32*, 951–989.

Wahba, G. (1990). *Spline Models for Observational Data*, Volume 50 of *Series in Applied Mathematics.* Philadelphia: SIAM.