Monograph:

Dodd, T.J. and Harrison, R.F. (2002) Gradient descent approach to approximation in reproducing kernel Hilbert spaces. Research Report. ACSE Research Report no. 821 . Automatic Control and Systems Engineering, University of Sheffield

Gradient Descent Approach to Approximation in Reproducing Kernel Hilbert Spaces

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

Research Report No. 821
May 2002

Abstract

Consider the bounded linear operator, $L : \mathcal{F} \to \mathcal{Z}$, where $\mathcal{Z} \subseteq \mathbb{R}^N$ and $\mathcal{F}$ are Hilbert spaces defined on a common field $\mathcal{X}$. L is made up of a series of N bounded linear evaluation functionals, $L_i : \mathcal{F} \to \mathbb{R}$. By the Riesz representation theorem, there exist functions, $k(x_i, \cdot) \in \mathcal{F} : L_i f = \langle f, k(x_i, \cdot) \rangle_{\mathcal{F}}$. The functions, $k(x_i, \cdot)$, are known as reproducing kernels and $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS). This is a natural framework for approximating functions given a discrete set of observations. In this paper the computational aspects of characterising such approximations are described and a gradient method presented for iterative solution. Such iterative solutions are desirable when N is large and the matrix computations involved in the basic solution become infeasible. This is also exactly the case where the problem becomes ill-conditioned. An iterative approach to Tikhonov regularisation is therefore also introduced. Unlike iterative solutions for the more general Hilbert space setting, the proofs presented make use of the spectral representation of the kernel.

1 Introduction

In many cases of interest the approximation of a function is equivalent to an inverse linear operator problem in Hilbert spaces. This is the case, for example, when the function is observed indirectly, such as is represented by a Fredholm integral equation of the first kind. In general, the mathematical treatment is then concerned with the infinite-dimensional problem where it is assumed that the function is observed over its whole domain. Numerical methods are then applied to form some approximate solution to the problem. In practice observations are often only available at a finite number of points in the function domain. These observations may be directly of the function at particular points or arise indirectly, such as in tomography (O'Sullivan 1986).

The solution of linear operator equations in Hilbert spaces is well known in the general case (Groetsch 1977) and also more specifically in connection with function approximation (Deutsch 2001). The case where the observation space is finite dimensional is well known to simplify the analysis (Kreyszig 1978). A full discussion of the solution of linear inverse problems with discrete observations was first made in (Bertero, De Mol, and Pike 1985) and the case where both the object and observation space are discrete is considered in (Hansen 1998). In (Bertero, De Mol, and Pike 1985) examples were given where the function of interest belongs to a reproducing kernel Hilbert space (RKHS). It is our opinion that RKHS do not simply belong to the realm of examples, but provide a (or even, the) natural framework for approximating functions given a discrete set of observations.

The basic theory and properties of RKHS is attributable to Aronszajn (Aronszajn 1950). RKHS have found widespread application in the solution of differential and integral equations (Saitoh 1997) and provide a unified framework for second-order stochastic processes and time series (Kailath 1971; Parzen 1961a). Examples of the application of RKHS to data interpolation and smoothing include splines (Kimeldorf and Wahba 1971; Wahba 1990), wavelets (Daubechies 1992), Paley-Wiener spaces (Yao 1967) and radial basis function type networks (Girosi, Jones, and Poggio 1993).

A common theme in these applications of RKHS to data interpolation and smoothing is the need to invert matrices which scale with the number of data. As the number of data increase the problem becomes ill-conditioned. This problem can be overcome by computing approximate regularised solutions (Hilgers 1976; Wahba 1977; Wahba 1990). However, this does not circumvent the computational burden which can become infeasible for very large data sets. In such cases it is necessary to use iterative approaches to solve the approximation problem.

In this paper we present iterative solutions, including complete proofs, to the approximation of functions in RKHS from discrete observations. The approaches are based on iterating in the direction of the negative gradient of the (regularised) error functional. This is an example of the method of successive approximations for linear operator equations, also known as gradient descent or weakest steepest descent (Groetsch 1977; Nashed 1970; Petryshyn 1962; Petryshyn 1963; Wolkowicz and Zlobec 1978). However, we present solutions which are readily computable without the need for approximation. Our approach to proving convergence is based on that introduced in (Parzen 1961b; Weiner 1965). Note that the authors of (Parzen 1961b; Weiner 1965) were concerned with a representation problem of continuous time series. In contrast, we are interested in the approximation problem of a continuous function given discrete data. The proofs make use of spectral results in RKHS which are not applicable to the general Hilbert space setting. Unlike previous descriptions of iterative approaches in Hilbert spaces we also generalise the method to solving the regularisation problem.

In the next section the natural framework for approximation of functions in RKHS from finite observations is presented. The generalised and regularised solutions to the approximation are described in Section 3. Computationally these solutions reduce to matrix computations in the case of finite observations as demonstrated by the series of results introduced in Section 4. In Sections 5 and 6 the iterative solutions are presented, the latter section is also concerned with the regularisation problem.

2 Function Approximation with Discrete Observations

Assume that we have some unknown function, f, of interest but that we are able to observe its behaviour at some finite number of points in its domain. The function belongs to some Hilbert space, $\mathcal{F}$, defined on a parameter set, $\mathcal{X}$. This set can be considered as an input set in the sense that for each $x \in \mathcal{X}$, $f(x)$ 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 $\{x_i\}_{i=1}^N$. 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 $[z_1, \ldots, z_N]^T$ 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.

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 (1).

By assuming that $\mathcal{F}$ is a Hilbert space, and further, that 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)

$$L_i f = \langle f, \psi_i \rangle_{\mathcal{F}}, \quad i = 1, \ldots, N \tag{3}$$

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 (3) is satisfied.

A natural setting for such approximation problems is the case where $\mathcal{F}$ is a reproducing kernel Hilbert space (RKHS). The functions, $\{\psi_i\}_{i=1}^N$, then correspond to reproducing kernels. 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 positive 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. But, 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. Hence, $k(x_i, x_j)$ 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. We therefore have the following result.

Theorem 2.1 (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.*

The following definition of a RKHS is equivalent to that given above based on bounded linear functionals.

Definition 2.2 (Parzen, 1960) *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

$$f(\cdot) = \sum_{i=1}^{N} c_i k(x_i, \cdot) \tag{4}$$

for $N \in \mathbb{Z}^+$ and $c_i \in \mathbb{R}$. A well defined inner product for this collection is (Wahba 1990)

$$\left\langle \sum_i a_i k(x_i,\cdot), \sum_j b_j k(x_j,\cdot) \right\rangle_{\mathcal{F}} = \sum_{i,j} a_i b_j \langle k(x_i,\cdot), k(x_j,\cdot)\rangle_{\mathcal{F}} = \sum_{i,j} a_i b_j k(x_i, x_j).$$

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.

Suppose that $k(x_1, x_2)$ is continuous and

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k^2(x_1, x_2) dx_1 dx_2 < \infty \tag{5}$$

then there exists an orthonormal sequence of continuous eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$ in $L_2(\mathcal{X})$ with associated eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ such that (Wahba 1990)

$$\int_{\mathcal{X}} k(x_1, x_2)\phi_i(x_2) dx_2 = \lambda_i \phi_i(x_1), \quad i = 1, 2, \ldots \tag{6}$$

$$k(x_1, x_2) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x_1)\phi_i(x_2), \tag{7}$$

$$\int_{\mathcal{X}} \int_{\mathcal{X}} k^2(x_1, x_2) dx_1 dx_2 = \sum_{i=1}^{\infty} \lambda_i^2 < \infty. \tag{8}$$

Note that in the finite dimensional case (6)-(8) become

$$\begin{aligned} K\phi_i &= \lambda_i \phi_i, & (9)\\ K &= \Gamma\Lambda\Gamma^T, & (10)\\ \text{trace } R^2 &= \sum_{i=1}^{N} \lambda_i^2 & (11) \end{aligned}$$

where K is the $N \times N$ matrix with ijth entry $k(x_i, x_j)$, ϕ_i is the ith eigenvector, Λ is the diagonal matrix with iith entry λ_i, and Γ is the orthogonal matrix with ith column ϕ_i.

The following lemma then holds.

Lemma 2.3 (Wahba, 1990) *Suppose (5) holds. If we let*

$$f_i = \int_{\mathcal{X}} f(x)\phi_i(x)dx, \tag{12}$$

then $f \in \mathcal{F}$ if and only if

$$\sum_{i=1}^{\infty} \frac{f_i^2}{\lambda_i} < \infty \tag{13}$$

and

$$\|f\|_{\mathcal{F}}^2 = \sum_{i=1}^{\infty} \frac{f_i^2}{\lambda_i}. \tag{14}$$

We can also expand f in a Fourier series as

$$f(x) = \sum_i f_i \phi_i(x). \tag{15}$$

3 Generalised Inverses and Tikhonov Regularisation

Returning to the approximation problem, this was shown to be equivalent to the linear operator equation

$$Lf = z \tag{16}$$

where $f \in \mathcal{F}, z \in \mathcal{Z} \subseteq \mathbb{R}^N, L : \mathcal{F} \to \mathcal{Z}$ and $\mathcal{F}$ is a RKHS with reproducing kernel $k(\cdot,\cdot)$. We assume throughout that the range, $R(L)$, of L is closed. If the operator L has an inverse then (16) always has the unique solution $z = L^{-1}f$. More generally, we may have more than one solution ($N(L) \neq \{0\}$, where $N(L)$ is the null space of L) or no solution at all ($z \notin R(L)$). We must therefore seek, in some sense, a "best possible" solution, u, to the approximation problem.

Theorem 3.1 (Groetsch, 1977) *Suppose $L : \mathcal{F} \to \mathcal{Z}$ has closed range and $z \in \mathcal{Z}$, then the following conditions on $u \in \mathcal{F}$ are equivalent:*

(i) $Lu = Pz$;

(ii) $\|Lu - z\|_{\mathcal{Z}} \leq \|Lf - z\|_{\mathcal{Z}}$ *for any* $f \in \mathcal{F}$; *and*

(iii) $L^*Lu = L^*z$.

Here P denotes the projection of z onto $R(L)$ and L^* is the adjoint operator to L defined by $\langle Lf, z\rangle_{\mathcal{Z}} = \langle f, L^*z\rangle_{\mathcal{F}}$.

We call a vector $u \in \mathcal{F}$ which satisfies the equivalent conditions (i)-(iii) of Theorem 3.1 a least squares solution of $Lf = z$.

Assuming $R(L)$ is closed, a least squares solution of (16) always exists for each $z \in \mathcal{Z}$. But, we may have $N(L) \neq \{0\}$ (consider, for example, the case of a function which is zero at each observation). Then there are infinitely many least squares solutions of (16) since if u is a least squares solution so is $u+v$ for any $v \in N(L)$.

The set of least squares solutions can be written, by Theorem 3.1, as

$$\{u \in \mathcal{F} : L^*Lu = L^*z\}. \tag{17}$$

By the continuity and linearity of L and L^* this set is a closed convex set which contains a unique vector of minimal norm (Groetsch 1977). We then define the generalised inverse of L as the mapping $L^\dagger : \mathcal{Z} \to \mathcal{F}$ such that $L^\dagger z = u$, where u is the least squares solution of minimal norm of the equation $Lf = z$. Note that if L is invertible we have $L^\dagger = L^{-1}$. The associated least squares solution, u, of minimal norm is the generalised solution, denoted $f^\dagger$.

The following theorem and associated corollary are important for characterising generalised solutions in our case.

Theorem 3.2 (Groetsch, 1977) *Suppose $L : \mathcal{F} \to \mathcal{Z}$ has closed range, then $L^\dagger = (L^*L)^\dagger L^* = L^*(LL^*)^\dagger$.*

Since, in our case, $\mathcal{Z}$ is finite dimensional we have the following corollary.

Corollary 3.3 *Suppose $L : \mathcal{F} \to \mathcal{Z}$ has closed range and $\mathcal{Z}$ is finite dimensional, then LL^* has a matrix representation and the computation of $L^\dagger$ reduces to the computation of the generalised inverse of a matrix.*

In a strict mathematical sense, since $\mathcal{Z}$ is finite dimensional, we have the following.

Theorem 3.4 (Bertero, 1985) *The generalised solution $L^\dagger$ depends continuously on the observations in the sense that if δz is the error in the observations and $\delta f^\dagger$ the induced error in $f^\dagger$, then $\|\delta f^\dagger\|_{\mathcal{F}} \to 0$ when $\|\delta z\|_{\mathcal{Z}} \to 0$.*

The generalised solution is therefore well-posed (Kirsch 1996) (a unique solution exists which depends continuously on the observations). However, a lack of numerical stability is possible. This can be seen from the following relation (Bertero, De Mol, and Pike 1985)

$$\frac{\|\delta f^\dagger\|_{\mathcal{F}}}{\|f^\dagger\|_{\mathcal{F}}} \leq C(L)\frac{\|\delta z\|_{\mathcal{Z}}}{\|z\|_{\mathcal{Z}}} \tag{18}$$

where

$$C(L) = \|L^\dagger\|\|L\|$$

is the condition number of L.

If $C(L)$ is large the problem of computing $f^\dagger$ is ill-conditioned. The generalised solution is therefore affected by numerical instability. Even small errors in the observations can produce a completely different and unphysical generalised solution. In such cases we must therefore seek some alternative stable solution.

Such solutions are generated by regularisation algorithms, of which many exist (Kirsch 1996). We consider the particular case of Tikhonov regularisation which is closely related to our existing generalised solution (Groetsch 1984).

We have already seen that $L^\dagger z$ is the vector, $u \in \mathcal{F}$, which minimises the functional $\|Lf - z\|$ and also has smallest norm amongst all such minimising vectors. The method of Tikhonov regularisation is approximately to minimise both the functional $\|Lf - z\|_{\mathcal{Z}}$ and the norm $\|f\|_{\mathcal{F}}$ by minimising the functional

$$\Phi[f] = \|Lf - z\|_{\mathcal{Z}}^2 + \rho\|f\|_{\mathcal{F}}^2 \tag{19}$$

where $\rho > 0$ is known as the regularisation parameter. Now, $\Phi[f]$ is Fréchet differentiable and we can therefore calculate the gradient $\nabla_f \Phi[f]$. The minimum of $\Phi[f]$ occurs at the unique stationary point u of Φ which satisfies $\nabla_f \Phi[u] = 0$ (Groetsch 1977). The gradient is given by

$$\nabla_f \Phi[f] = 2(L^* Lf - L^* z) + 2\rho f. \tag{20}$$

Hence the unique minimiser, f_{reg}, of (19) satisfies

$$f_{reg} = (\rho I + L^* L)^{-1} L^* z \tag{21}$$

or, equivalently,

$$f_{reg} = L^*(\rho I + LL^*)^{-1} z. \tag{22}$$

where I is the appropriate identity operator. Finally, the following theorem confirms a result which we would expect intuitively.

Theorem 3.5 (Groetsch, 1977) *If $L : \mathcal{F} \to \mathcal{Z}$ has closed range, then*

$$L^\dagger = \lim_{\rho \to 0^+} (\rho I + L^* L)^{-1} L^*$$

uniformly.

4 Characterisation of Operators in RKHS

Consider the operator $L : \mathcal{F} \to \mathcal{Z}$ where $\mathcal{Z}$ is the N dimensional Euclidean space with inner product $\langle g, h \rangle_{\mathcal{Z}} = \sum_{i=1}^N g_i h_i$, for $g, h \in \mathcal{Z}$. Then, for $z \in \mathcal{Z}, f \in \mathcal{F}$ the adjoint operator L^* is defined by

$$\langle Lf, z \rangle_{\mathcal{Z}} = \langle f, L^* z \rangle_{\mathcal{F}} \tag{23}$$

and transforms the observation vector, z, into an element of $\mathcal{F}$ or, more precisely, the finite dimensional subspace $\mathcal{F}_N$. In a RKHS the operator, L, acting on f has the form $Lf = \sum_{i=1}^N e_i \langle f, k(x_i, \cdot) \rangle_{\mathcal{F}}$, where $e_i \in \mathbb{R}^N$ is the ith standard basis vector. The following results apply to the operator, L, and its adjoint, L^*.

Theorem 4.1 *Given the operator L and its adjoint L^* defined by (23) then, in a RKHS with $Lf = \sum_{i=1}^N e_i \langle f, k(x_i, \cdot)\rangle_{\mathcal{F}}$, the adjoint L^* is given by*

$$L^* z = \sum_{i=1}^N z_i k(x_i, \cdot). \tag{24}$$

Proof. Solving for the LHS of (23)

$$\langle Lf, z\rangle_{\mathcal{Z}} = \sum_{i=1}^N z_i \langle f, k(x_i, \cdot)\rangle_{\mathcal{F}} = \sum_{i=1}^N f(x_i) z_i. \tag{25}$$

By assumption we set $L^* z = \sum_{i=1}^N z_i k(x_i, \cdot)$ and solving for the RHS of (23)

$$\langle f, L^* z\rangle_{\mathcal{F}} = \left\langle f, \sum_{i=1}^N z_i k(x_i, \cdot) \right\rangle_{\mathcal{F}} = \sum_{i=1}^N z_i \langle f, k(x_i, \cdot)\rangle_{\mathcal{F}} \tag{26}$$

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

Corollary 4.2 *The representation of a function in a RKHS, (4), for some $c \in \mathbb{R}^N$ is equivalent to*

$$f(\cdot) = L^* c.$$

Theorem 4.3 *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:

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

using the definition of L. As $z_i \in \mathbb{R}$ we can write this as

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

$\square$

Corollary 4.4 *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 (Gram) matrix $[K]_{ij} = k(x_i, x_j)$.*

Theorem 4.5 *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 Theorem 4.1 and the definition of the operator, L, we have

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

□

Given the above results we are now in a position to provide computable solutions to approximation in RKHS.

Corollary 4.6 *For the linear operator equation $Lf = z$, where $f \in \mathcal{F}, z \in \mathcal{Z} \subseteq \mathbb{R}^N$ and $\mathcal{F}$ is a RKHS, the generalised solution, $f^\dagger = L^*(LL^*)^\dagger z$, is given by*

$$f^\dagger(x) = k^T K^{-1} z$$

where k is the vector $[k(x, x_1), \ldots, k(x, x_N)]^T$.

Corollary 4.7 *The associated regularisation solution, $f_{reg} = L^*(\rho I + LL^*)^{-1} z$, is given by*

$$f_{reg}(x) = k^T (\rho I + K)^{-1} z.$$

5 Iterative Solutions

Consider now an iterative solution for $f^\dagger$, then, defining a sequence of estimates as $\{f^n\}_{n=1}^{\infty}$, the method of successive approximations estimates f^{n+1} in terms of f^n as

$$f^{n+1} = f^n - \gamma_n \tilde{f}^n \tag{27}$$

where $f^0 \in \mathcal{F}$, $\gamma_n \in \mathbb{R}^+$ and $\tilde{f}^n$ is the residual

$$\tilde{f}^n = L^*Lf^n - L^*z. \tag{28}$$

We motivate this iteration in terms of the negative gradient of the error functional as follows.

Consider again the error functional

$$\|Lf^n - z\|^2_{\mathcal{Z}}. \tag{29}$$

Expanding we have

$$\begin{aligned}\|Lf^n - z\|^2_{\mathcal{Z}} &= \langle Lf^n, Lf^n\rangle_{\mathcal{Z}} - 2\langle Lf^n, z\rangle_{\mathcal{Z}} + \langle z, z\rangle_{\mathcal{Z}} \\ &= \langle f^n, L^*Lf^n\rangle_{\mathcal{F}} - 2\langle f^n, L^*z\rangle_{\mathcal{F}} + \langle z, z\rangle_{\mathcal{Z}}.\end{aligned}$$

The gradient with respect to f^n is then given by

$$\nabla_{f^n}\|Lf^n - z\|^2_{\mathcal{Z}} = 2L^*Lf^n - 2L^*z \tag{30}$$

where the term $\langle z, z\rangle_{\mathcal{Z}}$, being independent of f^n is treated as a constant. This is simply twice the residual, i.e.

$$\nabla_f\|Lf^n - z\|^2_{\mathcal{Z}} = 2\tilde{f}^n. \tag{31}$$

In practice the iterations must be made on finite dimensional objects. Returning to the basic solution in RKHS, (4), f^n can be expressed, using the adjoint operator, as a linear combination of the c_i

$$f^n = L^*c^n \tag{32}$$

where $c^n = [c^n_1, \ldots, c^n_N]^T$. Also

$$\tilde{f}^n = L^*\tilde{c}^n, \quad \tilde{c}^n = LL^*c^n - z. \tag{33}$$

The method of successive approximations then finds estimates of the coefficients as

$$c^0 \in \mathbb{R}^N, \quad c^{n+1} = c^n - \gamma_n\tilde{c}^n \tag{34}$$

where the γ_n are chosen as below. The function at each iteration is determined by $f^n = L^*c^n = \sum_{j=1}^N c^n_j k(x_j, \cdot)$.

To complete the iterative scheme we need to define a schedule for the parameters γ_n and together with this prove convergence in the sense that $\|\tilde{f}^n\|^2 \to 0$ when $n \to \infty$.

Theorem 5.1 *Let* $\{\gamma_n\}_{n=0}^\infty$ *satisfy:*

(1) $0 < \gamma_n < 2/\lambda_{max}$, *for all* n, *where* λ_{max} *is the largest eigenvalue of* $LL^* = K$*; and*

(2) $\sum_{n=0}^\infty \gamma_n = \infty$.

Define the iteration $f^n = L^*c^n = \sum_{i=1}^N c^n_i k(x_i, \cdot)$ *together with* $f^0 \in \mathcal{F}$ *(i.e.* $c^0 \in \mathbb{R}^N$*) arbitrary,* $c^{n+1} = c^n - \gamma_n\tilde{c}^n$, $\tilde{c}^n = LL^*c^n - z$, *then*

$$\|\tilde{f}^n\|^2_{\mathcal{F}} = \|L^*\tilde{c}^n\|^2_{\mathcal{F}} \to 0.$$

as $n \to \infty$.

Proof. (a) *Monotonicity.*

$$\tilde{f}^{n+1} = L^*Lf^{n+1} - L^*z \tag{35}$$

but $f^{n+1} = L^*c^{n+1}$ and $c^{n+1} = c^n - \gamma_n \tilde{c}^n$, therefore $f^{n+1} = L^*(c^n - \gamma_n \tilde{c}^n)$ from which

$$\begin{aligned}\tilde{f}^{n+1} &= L^*LL^*(c^n - \gamma_n \tilde{c}^n) - L^*z \\ &= L^*Lf^n - L^*z - \gamma_n L^*L\tilde{f}^n = \tilde{f}^n - \gamma_n L^*L\tilde{f}^n.\end{aligned}$$

Define

$$\begin{aligned}\Delta\|\tilde{f}^n\|^2_{\mathcal{F}} &= \|\tilde{f}^n\|^2_{\mathcal{F}} - \|\tilde{f}^{n+1}\|^2_{\mathcal{F}} \\ &= \|\tilde{f}^n\|^2_{\mathcal{F}} - \|\tilde{f}^n - \gamma_n L^*L\tilde{f}^n\|^2_{\mathcal{F}}\end{aligned}$$

and thus

$$\begin{aligned}\Delta\|\tilde{f}^n\|^2_{\mathcal{F}} &= \|\tilde{f}^n\|^2_{\mathcal{F}} - \|\tilde{f}^n\|^2_{\mathcal{F}} - \gamma_n^2\langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} \\ &\quad +2\gamma_n\langle \tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} \\ &= 2\gamma_n\langle \tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} - \gamma_n^2\langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}}.\end{aligned}$$

Since L is bounded it is also compact (Kreyszig 1978) then L^*L is a non-negative self-adjoint map and (Edmunds and Evans 1987; Groetsch 1984),

$$L^*L\tilde{f}^n = \sum_j \tilde{f}^n_j \lambda_j \phi_j \tag{36}$$

where λ_j and ϕ_j, as defined in (6), correspond to the eigenvalues and associated eigenvectors of L^*L [1] and

$$\tilde{f}^n_j = \int_{\mathcal{X}} \tilde{f}^n(x)\phi_j(x)dx.$$

Using results (14) and (36) we also have

$$\begin{aligned}\langle L^*L\tilde{f}^n, \tilde{f}^n\rangle_{\mathcal{F}} &= \sum_{j=1}^N \frac{(\tilde{f}^n_j)^2\lambda_j}{\lambda_j} = \sum_{j=1}^N (\tilde{f}^n_j)^2, \\ \langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} &= \sum_{j=1}^N \frac{(\tilde{f}^n_j)^2\lambda_j^2}{\lambda_j} = \sum_{j=1}^N (\tilde{f}^n_j)^2\lambda_j.\end{aligned}$$

Therefore

$$\frac{\langle L^*L\tilde{f}^n, \tilde{f}^n\rangle_{\mathcal{F}}}{\langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}}} = \frac{\sum_{j=1}^N (\tilde{f}^n_j)^2}{\sum_{j=1}^N (\tilde{f}^n_j)^2\lambda_j} \tag{37}$$

[1] Note that L^*L has the same positive eigenvalues as LL^* with the same multiplicity (Bertero, De Mol, and Pike 1985; Edmunds and Evans 1987). The main result of the theorem applies to λ_{max} which is the same for both L^*L and LL^* (but is readily computable for $LL^* = K$).

from which we have

$$\frac{2\langle L^*L\tilde{f}^n, \tilde{f}^n\rangle_{\mathcal{F}}}{\langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}}} \geq \frac{2\sum_{j=1}^{N}(\tilde{f}_j^n)^2}{\lambda_{max}\sum_{j=1}^{N}(\tilde{f}_j^n)^2} = \frac{2}{\lambda_{max}}. \tag{38}$$

But by assumption $\gamma_n < 2/\lambda_{max}$, hence

$$\Delta\|\tilde{f}^n\|_{\mathcal{F}}^2 = 2\gamma_n\langle \tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} - \gamma_n^2\langle L^*L\tilde{f}^n, L^*L\tilde{f}^n\rangle_{\mathcal{F}} \geq 0. \tag{39}$$

(b) *Convergence.*

It was shown above that the residuals satisfy

$$\tilde{f}^{n+1} = L^*Lf^n - \gamma_n L^*L\tilde{f}^n - L^*z. \tag{40}$$

Hence

$$\begin{aligned}
\tilde{f}^{n+1} &= \tilde{f}^n - \gamma_n L^*L\tilde{f}^n = (I - \gamma_n L^*L)\tilde{f}^n \\
&= \prod_{k=0}^{n}(I - \gamma_k L^*L)\tilde{f}^0 = \prod_{k=0}^{n}(I - \gamma_k L^*L)g
\end{aligned}$$

where $g = \tilde{f}^0 \in \mathcal{F}$ which can be expanded in terms of the orthonormal basis of eigenfunctions, ϕ_i, as (c.f. (15))

$$g = \sum_i g_i\phi_i(x) \tag{41}$$

where the g_i are given by (12), and since L^*L is self-adjoint (with L bounded) (Groetsch 1984)

$$L^*Lg = \sum_i g_i\lambda_i\phi_i(x) \tag{42}$$

where λ_i refers to the eigenvalues of L^*L.

We can then write

$$\prod_{k=0}^{n}(I - \gamma_k L^*L)g = \sum_i g_i \prod_{k=0}^{n}(1 - \gamma_k\lambda_i)\phi_i \tag{43}$$

and hence

$$\|\tilde{f}^{n+1}\|_{\mathcal{F}}^2 = \left\|\prod_{k=0}^{n}(I - \gamma_k L^*L)g\right\|_{\mathcal{F}}^2 = \sum_i \frac{g_i^2}{\lambda_i}\prod_{k=0}^{n}(1 - \gamma_k\lambda_i)^2 \tag{44}$$

where we have used (14) and (15).

Using the assumed inequality on γ_k

$$1 - \frac{2\lambda_i}{\lambda_{max}} < 1 - \gamma_k\lambda_i < 1 \Rightarrow (1 - \gamma_k\lambda_i)^2 < 1. \tag{45}$$

Then for any $l \in \mathbb{Z}^+$

$$\|\tilde{f}^{n+1}\|_{\mathcal{F}}^2 \leq \sum_{i=1}^{l} \frac{g_i^2}{\lambda_i} \prod_{k=0}^{n} (1 - \gamma_k \lambda_i)^2 + \sum_{i>l} \frac{g_i^2}{\lambda_i}. \tag{46}$$

For fixed l, let $n \to \infty$, and since $\sum_{i=0}^{\infty} \gamma_i = \infty$ (by assumption) and $(1 - \gamma_k \lambda_i)^2 < 1$, the first term tends to 0. Now let $l \to \infty$, $g \in \mathcal{F} \Rightarrow \sum_i g_i^2/\lambda_i < \infty$ (Lemma 2.3), and therefore the second term is the tail of a convergent series and therefore tends to 0. □

If we further define $c^0 = 0$ (and therefore $f^0 = 0$) then for any n, $\|f^n\|_{\mathcal{F}} \leq \|f^{n+1}\|_{\mathcal{F}}$ and therefore $\|f^n\|_{\mathcal{F}} \leq \|f^\dagger\|_{\mathcal{F}}$ (Bertero, De Mol, and Pike 1988). It therefore follows that the method of successive approximations defines a regularisation scheme where the inverse of the number of iterations plays the role of the regularisation parameter.

6 Iterative Regularised Solutions

Consider now, instead, the regularised error functional

$$\|Lf^n - z\|_{\mathcal{Z}}^2 + \rho \|f^n\|_{\mathcal{F}}^2. \tag{47}$$

Expanding

$$\begin{aligned} \|Lf^n - z\|_{\mathcal{Z}}^2 + \rho\|f^n\|_{\mathcal{F}}^2 &= \langle Lf^n, Lf^n \rangle_{\mathcal{Z}} - 2\langle Lf^n, z\rangle_{\mathcal{Z}} + \langle z, z\rangle_{\mathcal{Z}} + \rho\langle f^n, f^n\rangle_{\mathcal{F}} \\ &= \langle f^n, L^*Lf^n\rangle_{\mathcal{F}} - 2\langle f^n, L^*z\rangle_{\mathcal{F}} + \langle z, z\rangle_{\mathcal{Z}} + \rho\langle f^n, f^n\rangle_{\mathcal{F}} \end{aligned}$$

The gradient with respect to f^n is given by

$$\nabla_{f^n}\{\|Lf^n - z\|_{\mathcal{Z}}^2 + \rho\|f^n\|_{\mathcal{F}}^2\} = 2L^*Lf^n - 2L^*z + 2\rho f^n \tag{48}$$

and therefore the direction of the gradient is $L^*Lf^n - L^*z + \rho f^n$ which is once again the (now regularised) residual.

We now apply the method of successive approximations to estimate f^{n+1} in terms of f^n as

$$f^{n+1} = f^n - \gamma_n \tilde{f}^n \tag{49}$$

where $f^0 \in \mathcal{F}, \gamma_n \in \mathbb{R}^+$ and now $\tilde{f}^n$ is given by

$$\tilde{f}^n = (L^*L + \rho I)f^n - L^*z. \tag{50}$$

Again, we express the f^n in terms of the adjoint operator as

$$f^n = L^*c^n. \tag{51}$$

Before proceeding with the basic theorem and proof we require the following lemma.

Lemma 6.1 *(Spectral Mapping Theorem) Consider a linear operator, $T : \mathcal{F} \to \mathcal{F}$, and p, a polynomial, then, for the spectrum, σ,*

$$\sigma(p(T)) = p(\sigma(T)) = \{p(\lambda) : \lambda \in \sigma(T)\}.$$

This simply states that the eigenvalues of a polynomial function of the linear operator, T, are equal to the same polynomial applied to the eigenvalues of the original operator, T.

The gradient iteration for the regularised case is then summarised in the following theorem.

Theorem 6.2 *Let $\{\gamma_n\}$ satisfy:*

(1) $0 < \gamma_n < \frac{2}{\lambda_{max}+\rho}$, for all n, where λ_{max} is the largest eigenvalue of $LL^* = K$; and

(2) $\sum_{n=0}^{\infty} \gamma_n = \infty$.

Define the iteration, $f^n = L^* c^n = \sum_{i=1}^{N} c_i^n k(x_i, \cdot)$, together with $f^0 \in \mathcal{F}$ (i.e. $c^0 \in \mathbb{R}^N$) arbitrary, $c^{n+1} = c^n - \gamma_n \tilde{c}^n$, $\tilde{c}^n = (LL^* + \rho I)c^n - z$, then

$$\|\tilde{f}^n\|_{\mathcal{F}}^2 = \|L^* \tilde{c}^n\|_{\mathcal{F}}^2 \to 0$$

as $n \to \infty$.

Proof. (a) *Monotonicity.*

$$\tilde{f}^n = (L^*L + \rho I) f^n - L^* z \tag{52}$$

but $f^{n+1} = L^* c^{n+1}$ and $c^{n+1} = c^n - \gamma_n \tilde{c}^n$, therefore $f^{n+1} = L^*(c^n - \gamma_n \tilde{c}^n)$ from which

$$\begin{aligned} \tilde{f}^{n+1} &= (L^*L + \rho I)L^*(c^n - \gamma_n \tilde{c}^n) - L^* z \\ &= (L^*L + \rho I) f^n - L^* z - \gamma_n (L^*L + \rho I)\tilde{f}^n \\ &= \tilde{f}^n - \gamma_n (L^*L + \rho I)\tilde{f}^n \\ &= (I - \gamma_n (L^*L + \rho I))\tilde{f}^n. \end{aligned}$$

Define

$$\begin{aligned} \Delta \|\tilde{f}^n\|_{\mathcal{F}}^2 &= \|\tilde{f}^n\|_{\mathcal{F}}^2 - \|\tilde{f}^{n+1}\|_{\mathcal{F}}^2 \\ &= \|\tilde{f}^n\|_{\mathcal{F}}^2 - \|(I - \gamma_n (L^*L + \rho I))\tilde{f}^n\|_{\mathcal{F}}^2 \end{aligned}$$

which can be expanded, thus

$$\begin{aligned} \Delta \|\tilde{f}^n\|_{\mathcal{F}}^2 &= \|\tilde{f}^n\|_{\mathcal{F}}^2 - \|\tilde{f}^n\|_{\mathcal{F}}^2 - \gamma_n^2 \langle (L^*L + \rho I)\tilde{f}^n, (L^*L + \rho I)\tilde{f}^n \rangle_{\mathcal{F}} \\ &\quad + 2\gamma_n \langle \tilde{f}^n, (L^*L + \rho I)\tilde{f}^n \rangle_{\mathcal{F}} \\ &= 2\gamma_n \langle \tilde{f}^n, (L^*L + \rho I)\tilde{f}^n \rangle_{\mathcal{F}} - \gamma_n^2 \langle (L^*L + \rho I)\tilde{f}^n, (L^*L + \rho I)\tilde{f}^n \rangle_{\mathcal{F}}. \end{aligned}$$

Analogous to the proof in the previous section

$$\frac{\langle \tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}}}{\langle (L^*L + \rho I)\tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}}} = \frac{\sum_{j=1}^N \lambda_j(\lambda_j + \rho)(\tilde{f}_j^n)^2}{\sum_{j=1}^N \lambda_j(\lambda_j + \rho)^2(\tilde{f}_j^n)^2} \tag{53}$$

where the λ_j are again the eigenvalues of L^*L and we have applied the Spectral Mapping Theorem. Therefore

$$\frac{2\langle \tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}}}{\langle (L^*L + \rho I)\tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}}} \geq \frac{2\sum_{j=1}^N \lambda_j(\lambda_j + \rho)(\tilde{f}_j^n)^2}{(\lambda_{max} + \rho)\sum_{j=1}^N \lambda_j(\lambda_j + \rho)(\tilde{f}_j^n)^2} \tag{54}$$

where the latter term is simply equal to $\frac{2}{\lambda_{max}+\rho}$. But, by assumption $\gamma_n < \frac{2}{\lambda_{max}+\rho}$ from which

$$\Delta\|\tilde{f}^n\|_{\mathcal{F}}^2 = 2\gamma_n\langle \tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}} - \gamma_n^2\langle (L^*L + \rho I)\tilde{f}^n, (L^*L + \rho I)\tilde{f}^n\rangle_{\mathcal{F}} \geq 0.$$

(b) *Convergence.*

Now, in general, the residuals satisfy

$$\tilde{f}^{n+1} = (I - \gamma_n(L^*L + \rho I))\tilde{f}^n$$

and therefore, by iteration,

$$\begin{aligned}\tilde{f}^{n+1} &= \prod_{k=0}^n (I - \gamma_k(L^*L + \rho I))\tilde{f}^0 \\ &= \prod_{k=0}^n (I - \gamma_k(L^*L + \rho I))g\end{aligned}$$

where $g = \tilde{f}^{n+1} \in \mathcal{F}$. Again we can express g and L^*Lg as

$$g = \sum_i g_i\phi_i(x) \tag{55}$$

and

$$L^*Lg = \sum_i g_i\lambda_i\phi_i(x). \tag{56}$$

Using the Spectral Mapping Theorem

$$(L^*L + \rho I)g = \sum_i g_i(\lambda_i + \rho)\phi_i(x) \tag{57}$$

and therefore

$$\prod_{k=0}^n (I - \gamma_k(L^*L + \rho I))g = \sum_i g_i \prod_{k=0}^n (1 - \gamma_k(\lambda_i + \rho))\phi_i. \tag{58}$$

Hence

$$\|\tilde{f}^{n+1}\|_{\mathcal{F}}^2 = \left\| \prod_{k=0}^{n} (I - \gamma_k(L^*L + \rho I))g \right\|_{\mathcal{F}}^2 = \sum_i \frac{g_i^2}{\lambda_i} \prod_{k=0}^{n} (1 - \gamma_k(\lambda_i + \rho))^2 \quad (59)$$

where we have, again, used (14) and (15).

Using the assumed inequality on γ_k

$$1 - 2\left(\frac{\lambda_i + \rho}{\lambda_{max} + \rho}\right) < 1 - \gamma_k(\lambda_i + \rho) < 1 \Rightarrow (1 - \gamma_k(\lambda_i + \rho))^2 < 1. \quad (60)$$

Then, for any $l \in \mathbb{Z}^+$

$$\|\tilde{f}^{n+1}\|_{\mathcal{F}}^2 \leq \sum_i^{l} \frac{g_i^2}{\lambda_i} \prod_{k=0}^{n} (1 - \gamma_k(\lambda_i + \rho))^2 + \sum_{i>l} \frac{g_i^2}{\lambda_i}. \quad (61)$$

By similar arguments to the previous iteration proof, for fixed l, let $n \to \infty$, and since $\sum_{i=0}^{\infty} \gamma_i = \infty$ (by assumption) and $(1 - \gamma_k(\lambda_i + \rho))^2 < 1$ the first term tends to zero. Now let $l \to \infty$, $g \in \mathcal{F} \Rightarrow \sum_i g_i^2/\lambda_i < \infty$ (Lemma 2.3), and the second term is the tail of a convergent series and therefore tends to zero. □

7 Conclusions

A natural framework for approximating functions, given a discrete set of observations, has been presented. The function of interest belongs to a RKHS and the observations arise as bounded linear functionals mapping to the reals. The computational aspects of characterising such approximations were described and a gradient method presented for iterative solution. Such iterative solutions are desirable when the number of observations is large and the matrix computations involved in the direct solution become infeasible. This is also exactly the case where the problem becomes ill-conditioned. An iterative approach to Tikhonov regularisation was therefore also presented. Detailed proofs were given for the convergence of the iterative solutions. These iterative solutions are novel in the context of approximating functions in RKHS with discrete data. Further, the proofs of convergence are unique to RKHS in making use of a particular eigenexpansion of the kernel.

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.

Bertero, M., C. De Mol, and E. Pike (1988). Linear inverse problems with discrete data. II: Stability and regularisation. *Inverse Problems 4*, 573–594.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM.

Deutsch, F. (2001). *Best Approximation in Inner Product Spaces*, Volume 7 of *CMS Books in Mathematics*. Springer-Verlag.

Edmunds, D. and W. Evans (1987). *Spectral Theory and Differential Operators*. Oxford Mathematical Monographs. Clarendon Press.

Girosi, F., M. Jones, and T. Poggio (1993). Priors, stabilizers and basis functions: from regularization to radial, tensor and additive splines. Technical Report C.B.C.L. Paper No. 75, Artificial Intelligence Laboratory, MIT.

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

Groetsch, C. (1984). *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Volume 105 of *Research Notes in Mathematics Series*. Pitman.

Hansen, P. (1998). *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM Monographs on Mathematical Modeling and Computation. SIAM.

Hilgers, J. (1976). On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems. *SIAM Journal on Numerical Analysis 13*(2), 172–184.

Kailath, T. (1971). RKHS approach to detection and estimation problems - Part I: Deterministic signals in Gaussian noise. *IEEE Transactions on Information Theory IT-17*(5), 530–549.

Kimeldorf, G. and G. Wahba (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications 33*, 82–95.

Kirsch, A. (1996). *An Introduction to the Mathematical Theory of Inverse Problems*, Volume 120 of *Applied Mathematical Sciences*. Springer.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley & Sons.

Nashed, M. (1970). Steepest descent for singular linear operator equations. *SIAM Journal on Numerical Analysis 7*(3), 358–362.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistial Science 1*(4), 502–527.

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

Parzen, E. (1961b). Regression analysis of continuous parameter time series. In *Proceedings of the Fourth Berkeley Symposium on Probability and Mathematical Statistics*, Volume 1, Berkeley, pp. 469–489. University of California Press.

Petryshyn, W. (1962). Direct and iterative methods for the solution of linear operator equations in Hilbert space. *Transactions of the American Mathematical Society 105*, 136–175.

Petryshyn, W. (1963). On a general iterative method for the approximate solution of linear operator equations. *Mathematics of Computation 17*(81), 1–10.

Saitoh, S. (1997). *Integral Transforms, Reproducing Kernels and Their Applications*, Volume 369 of *Pitman Research Notes in Mathematics Series*. Longman.

Wahba, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis 14*(4), 651–667.

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

Weiner, H. (1965). The gradient iteration in time series analysis. *Journal of the Society for Industrial and Applied Mathematics 13*(4), 1096–1101.

Wolkowicz, H. and S. Zlobec (1978). Calculating the best approximate solution of an operator equation. *Mathematics of Computation 32*(144), 1183–1213.

Yao, K. (1967). Applications of reproducing kernel Hilbert spaces - bandlimited signal models. *Information and Control 11*, 429–444.