Towards a Theory of Multivariate Interpolation using Spaces of Distributions

Henry Wayne

Submitted to the University of Leicester
for the degree of Doctor of Philosophy.

1996

Abstract

The research contained in this thesis concerns the study of multivariate interpolation problems. Given a discrete set of possibly complex-valued data, indexed by a set of interpolation nodes in Euclidean space, it is desirable to generate a function which agrees with the data at the nodes. Within this general framework, this work pursues and generalizes one approach to the problem. Based on a variational theory, we construct a parameterised family of Hilbert spaces of tempered distributions, detail the necessary evolution of the interpolation problem, and provide a general error analysis. Some of the more popular applications from the theory of radial basis functions are shown to arise naturally, but the theory admits many more examples, which are not necessarily radial. The general error analysis is applied to each of the applications, and taken further where possible. Connections with the theory of conditionally positive definite functions are highlighted, but are not central to the theme.

Preface

I will probably never be able to repay my debt of gratitude to the people who acted as counsellors, colleagues, peers, critics and above all, friends, but I should start somewhere, and this is as good a place as any.

First, and foremost, my gratitude goes to Professor Will Light, who has visibly grown wearier over the past three years through my incessant questioning, lack of understanding, and inability to listen. In spite of this, his guidance has been without question, and I count myself fortunate to have stumbled into his class on approximation theory during my undergraduate years. If there is but one thing I have learnt from Will, it is how little I understand.

Of course, special thanks must go to the EPSRC for their support of this work over the three years — it is difficult to thank a group of anonymous people, but without their help, I would not be here.

Thanks must also go to my colleagues and friends in the Department, here at Leicester — without their diversity of interests and enthusiasm, I would probably be financially wealthier, but socially poorer. A fair trade in my opinion. In parallel, my gratitude also extends to the people who have provided the everyday opportunities to succeed and learn, opportunities which I have readily grasped, and will never forget.

Finally, I must thank those who have given their support generously and freely, such as my old school teacher, Mr. Innes; a whole host of friends too numerous to mention;

fellow postgraduate students; Alastair and Vicky, whose hospitality is without peer, but most especially my parents to whom this work is dedicated.

Cheers,

H.

Contents

Notation and Conventions

$L^p(\Omega)$	Space of Lesbegue measurable functions whose p^{th} power is integrable over Ω.
$L^1_{\text{loc}}(\Omega)$	Space of Lesbegue measurable functions which are integrable over every compact subset of Ω.
$W^{k,p}(\Omega)$	Sobolev space of Lesbegue measurable functions whose derivatives upto and including order k lie in $L^p(\Omega)$.
$\|\cdot\|$	Generic Hilbert space norm.
$\|\cdot\|_p$, $\|\cdot\|_{p,\Omega}$	Lesbegue norms on $L^p(\mathbb{R}^n)$ and $L^p(\Omega)$ respectively.
$\|\cdot\|_{k,p,\Omega}$	Sobolev norm on $W^{k,p}(\Omega)$.
$\vert\cdot\vert_{p,\Omega}$, $\vert\cdot\vert_{k,p,\Omega}$	Semi-norms on $L^p(\Omega)$ and $W^{k,p}(\Omega)$ respectively.

Chapter 1

Introduction

The distinction between interpolation as a mathematical theory and a mathematical tool is very fine. As a tool, interpolation enjoys a wide variety of applications in the physical sciences — each one utilising a different aspect of the theory. The underlying theme though, is the same in every case. Given a discrete set of data from $\mathbb{C}$, indexed by a set of interpolation nodes in n-dimensional Euclidean space, the problem is to construct a function $u : \mathbb{R}^n \longrightarrow \mathbb{C}$ which agrees with the data at the nodes. Furthermore, it is often desirable to place certain requirements on the general behaviour of u, whether it be continuity, differentiability, or simply a bound on an 'energy' associated with the physical context of the problem.

The applications of such a scheme are many. Conceptually, the simplest is the reconstruction of a sampled signal, ensuring the result is, in some sense, more accessible to further analytical techniques. Data compression is another important application, but one where the focus is slightly different. In this case, the aim is to sample the data in such a way that the storage of the interpolant justifies the computational expense, whilst the level of reconstruction remains adequate for the application in hand. The ensuing conflicts between computation and reconstruction have provided many incentives for the development of a cohesive theory. Another application, whose motivation is balanced between those of the above, is that of modelling, where the problem requires a visual output of some surface, but the computational time is now the key issue.

In all of these applications, the interpolation nodes may be chosen by a method which is natural to the task at hand, or they may be the only ones at which the sampled data is considered accurate enough for use. In either case, these interpolation problems are sometimes referred to as having 'scattered data' — data, which in some sense, is generated from a set of nodes which may have no exploitable structure. In turn, this creates problems for prospective theories since it is known that, given any set of nodes in $\mathbb{R}^n$, $n > 1$, there is no finite dimensional linear space from which the interpolation problem is guaranteed to have a unique solution. To overcome this, the theory makes a trade-off with the application by supplying interpolants for a large class of configurations of the interpolation nodes. In multivariate theories, these configurations are sufficiently general to make the interpolation problem uniquely solvable in almost all cases that arise where the size of the data set is finite.

In determining the effectiveness of the interpolation method, there are many considerations. Almost immediately, we are confronted by the question, how do we measure the effectiveness, and what are the units of measurement? To answer the former, we must return to the nature of the problem. Is it appropriate to measure the relative pointwise error between a function f and its interpolant u, given by $|f(x)-u(x)|$? Does this quantity even have meaning? If we know that f and u lie in some Sobolev space $W^{k,p}(\mathbb{R}^n)$, should we measure $\|f-u\|_{k,p}$? In many applications, the method of measuring is given by, or is a consequence of the pointwise error, but this has an important effect on the theory — we must deal with functions for which point evaluation is a *meaningful* operation.

Turning to the subject of units of measurement, the immediate impact on a prospective theory is to enforce some uniform condition on the functions to which interpolation is desirable. Every function must be 'measurable' in this loose sense for the effectiveness to have meaning. In practice, these assumptions usually arise from the application, where they maybe interpreted as restrictions on a quantifiable 'energy'. In terms of the mathematical theory, they often relate to smoothness conditions on the functions, and their derivatives.

We will take a brief look at two of the existing methods and approaches to multivariate interpolation, in order to properly place the work of this thesis in the on-going research initiatives. Both methods concentrate on building the interpolant from a particular type of function, which we now discuss.

Radial basis function interpolants

Given a set of interpolation nodes $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$, the subspace from which interpolants are formed is frequently, in its simplest form, given by

$$\text{span}\,\{\Psi(\cdot - a_r) : r = 1, \ldots, m\}.$$

Here, Ψ is a complex-valued function on $\mathbb{R}^n$, and we refer to it as the basis function. Interpolants are then built up as linear combinations:

$$u(x) = \sum_{r=1}^{m} \alpha_r \Psi(x - a_r), \qquad x \in \mathbb{R}^n,$$

where the coefficients $\{\alpha_j\}$ may be complex-valued. In radial basis function interpolation, Ψ is given the simple form $\Psi(x) = \phi(|x|)$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^n$ and $\phi \in C[0, \infty)$. However, the unique solvability of the interpolation equations may not be guaranteed — it is often the case that additional polynomial terms and conditions on the coefficients α_r, $r = 1, \ldots, m$ are required to ensure this. In detail, let π_k denote the space of polynomials of degree at most k, with dimension ℓ and assume that $p_1, \ldots, p_\ell$

form a basis for this space. Then interpolants may be constructed with the following form,

$$u(x) = \sum_{r=1}^{m} \alpha_r \Psi(x - a_r) + \sum_{s=1}^{\ell} \beta_s p_s(x), \qquad x \in \mathbb{R}^n,$$

where the degree of polynomials k is now chosen to ensure solvability. This mysterious form holds yet more surprises when the 'natural' conditions for solvability come to light. Suppose the interpolation data is given by $d_1, \ldots, d_m$. Then the choice of coefficients $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_\ell$ is restricted so that they satisfy

$$u(a_r) = d_r,\ r = 1, \ldots, m \text{ and } \sum_{r=1}^{m} \alpha_r p_s(a_r) = 0,\ s = 1, \ldots, m. \tag{1.1}$$

Furthermore, the interpolation nodes are now required to form a π_k-*unisolvent set*, that is, choose $a_1, \ldots, a_m$ so that, if p can be found in π_k such that $p(a_r) = 0$, $r = 1, \ldots, m$, then p must be the zero function.

Both methods we will discuss give rise to this form of interpolant — the theories also explain the mysterious nature and origin of the interpolant, but the common core of the two schemes arise from differing perspectives.

Surface splines

The theory of spline approximation is relatively modern, yet increasingly well-understood. It concerns the efficient construction of piecewise polynomial approximations to functions which satisfy certain smoothness conditions. For example, the cubic splines in one dimension are comprised of cubic sections, which are joined in a particular way so that the overall interpolant is twice continuously differentiable everywhere, with each section depending on all of the interpolation data. If restrictions are imposed at either end of the interval containing the interpolation nodes, so that the first and second derivatives of the interpolant vanish, then the resulting cubic splines are called natural, and arise from a well-documented variational problem (c.f. Holladay [17]). One of the main uses of these splines occurs in computer generated images of curves, where the B-spline basis greatly increases computational efficiency (c.f. Schumaker [32]).

The multivariate analogues of the natural splines are the surface splines. Here, problems arising in the aeronautical industry prompted research into thin-plate splines [15] in the early 1970s. These interpolants were chosen so that they minimized the bending energy of a thin lamina as it was stretched over a skeletal frame — each point of contact being an interpolation node. This concept of minimization runs deep through the theory, but was made precise when progress continued with the work of Atteia [4], and especially Duchon, whose two papers [10, 11] are seminal in this area. Duchon introduced a family of semi-Hilbert spaces in which interpolation problems were well-defined. From a variational theory he then constructed the interpolants and showed that they satisfied the conditions mentioned earlier. In the particular case [11], Duchon considered the space of Schwartz distributions whose k^{th} derivatives lie in $L^2(\mathbb{R}^n)$. From the resulting analysis, Duchon showed that interpolation with the basis functions defined, for all x in $\mathbb{R}^n$, by

$$\Psi(x) = \begin{cases} |x|^{2k-n} \ln |x|, & 2k - n \text{ is an even integer,} \\ |x|^{2k-n}, & \text{otherwise,} \end{cases}$$

converged in the L^p norm at a certain rate. We should make precise the meaning of this convergence, in order to obtain a better feel for the process of error estimation. Given a function f, its interpolant u would be constructed from the data given at the interpolation points in the set $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$. If these points are all contained in a set Ω, then we can define a measure of the density of these points with respect to Ω by $\sup_{t \in \Omega} \inf_{a \in \mathcal{A}} |t - a|$. Let h denote this quantity. In [11], Duchon showed that, for $p \geq 2$,

$$\|f - u\|_{p,\Omega} = \mathcal{O}(h^{k - \frac{n}{2} + \frac{n}{p}}).$$

Therefore, as the set $\mathcal{A}$ 'fills' out the set Ω, the interpolant gradually imitates the L^p nature of f closer and closer. This approach to the error analysis is well-suited to scattered data problems, as we will see.

However, the approach to the problem was based on spaces of functions, each having some common smoothness property. If a function was selected from one of these spaces, and a valid set of nodes provided, then an interpolant with known, desirable properties could always be constructed. In the univariate cases, the surface splines coincide with the natural splines, and the variational theories which generate each viewpoint are then seen to be equivalent.

In contrast with the surface splines, Hardy published in 1971 his work on multivariate interpolation using the multiquadrics [16], whose basis function is given by

$$\Psi(x) = \sqrt{1 + |x|^2}, \qquad x \in \mathbb{R}^n.$$

The theory surrounding them was not easily seen to arise from a 'natural' space of functions, until the work of Madych and Nelson [21, 22], and Wu and Schaback [39]. We turn therefore to the second approach to radial basis function interpolation.

Conditionally positive definite functions

The second approach to multivariate interpolation centres on the choice of basis function, rather than that which naturally arises from a space of functions when combined with a variational theory. The questions that were asked in the 1980s were, which functions should be used, and how good are they? We will answer the former question first, using a definition which is now common throughout the literature [23, 29, 21].

1.0.1 Definition *A function $\phi \in C[0,\infty)$ is conditionally positive definite of order k on $\mathbb{R}^n$ if and only if, for any set of distinct points $\{x_1,\ldots,x_m\}$ in $\mathbb{R}^n$, the inequality*

$$\sum_{r,s=1}^{m} \alpha_r\alpha_s\phi(|x_r - x_s|) \geq 0,$$

holds for all sets of coefficients $\{\alpha_r\}$ which satisfy

$$\sum_{r=1}^{m} \alpha_r p(x_r) = 0,$$

for all p in π_{k-1}. If the inequality holds for any choice of the coefficients $\{\alpha_r\}$, then this corresponds to the case when $k = 0$, and ϕ is referred to as positive definite.

Using these types of functions, Micchelli gives, in [23], sufficient conditions for the non-singularity of the interpolation matrices which arise from the conditions in (1.1). Moreover, he gives a fundamental characterisation of the conditionally positive definite functions, which we include here.

1.0.2 Definition *A function $f \in C^\infty(0,\infty)$ is completely monotone if, and only if, for all $j \geq 0$, $(-1)^j f^{(j)}$ is non-negative.*

1.0.3 Theorem (Micchelli [23]) *A function $\phi \in C[0,\infty) \cap C^\infty(0,\infty)$ is conditionally positive definite of order k on $\mathbb{R}^n$ if, and only if, $(-1)^k\phi^{(k)}$ is completely monotone.*

This characterisation extended a result of Schoenberg [31] which discussed the case of positive definite functions, but nevertheless, provides considerable motivation for the further study of those radial basis functions which are conditionally positive definite. It was then found that the degree of conditionally positive definiteness dictated the degree of polynomials which should be added to the linear combination of basis function. Moreover, associated with each basis function, a 'native space' could be constructed and an error analysis given, as seen in the work of Madych and Nelson [21, 22], Wu and Schaback [39], and Schaback [29].

Examples of the immediate applications of this alternative approach included Hardy's multiquadrics, inverse multiquadrics, Gaussians and surface splines, and therefore, genuinely represented a theory of multivariate interpolation. This work, however, returns to the first approach, and defines a parameterised family of spaces of distributions in which interpolation problems are well-posed. An error analysis is provided under general assumptions, and applications are studied. The theory here is also sufficiently general as to capture many of the more popular theories mentioned above. However, the conditions are fairly relaxed and will admit many more examples — it is undoubtably true to say that radial basis functions provided the motivation for this work, but as will be seen, the interpolants that arise do not necessarily exhibit this property.

Any work of this type has a foundation in an abstract setting, and indeed, the results herein owe much to functional analysis — a subject which has grown steadily since the turn of the century, both in terms of its position as a core mathematical subject, and in its influence in other areas, such as approximation theory, and partial differential equations. Within functional analysis, two centuries of deliberation on the subject of generalised functions culminated in the work of Schwartz [33, 34] in 1950. The resulting *Theorie des Distributions* and its treatment of generalised Fourier analysis, has had a huge impact in many diverse areas, and certainly plays a major role in this work.

Chapter 2 therefore deals with the functional analytic representation of interpolation problems in Hilbert spaces. The history of the method stretches back to the beginning of the century, with the work of Peano, through Hilbert, Riesz and Sz-Nagy, but the presentation here is a brief review of the work of Golomb and Weinberger [14], whose seminal work in 1959 sketches out the framework for the calculation of the interpolant and the error estimates.

Chapter 3 then takes the abstract setting and, under some fairly general assumptions, transforms the Golomb and Weinberger theory into an error analysis for the interpolation problems we will be dealing with. In the process, we will see the development of the 'power function' as our error estimator — something that can be seen in the alternative theories of Wu and Schaback [39], Schaback [29] and Powell [25]. The assumptions required for the error analysis also highlight, quite clearly, the necessary steps through which a theory must evolve for the analysis to be applicable.

With this blueprint, Chapter 4 presses ahead with the construction of a Hilbert space of tempered distributions, supplying a sufficiently rich set of results to describe the space, and the resulting interpolants in some detail. Furthermore, it pulls together many of the common threads between the different approaches to multivariate interpolation, whilst at the same time, uncovering some new insights into the basis functions.

Chapter 5 then takes on board the analysis of Chapter 3, with the benefit of the insights of Chapter 4. Different examples of the parameterised space are given, along with the familiar theories they generate. An alternative error estimate, reminiscent of that seen in [21], is given, and compared with those that arise from Chapter 3 through a more rudimentary means of analysis.

Chapter 2

Hilbert Space Theory

This chapter serves as a brief introduction to the variational theory which lies at the heart of our work. We will attempt to be concise, highlighting only those aspects which will become relevant at later stages. The history of this theory goes certainly as far back as the foundations of functional analysis, but culminates in the seminal paper by Golomb and Weinberger [14], on which the following is based.

2.1 Foundations of the variational theory

We begin with a linear space X, and a collection of m linear 'information' functionals, $\gamma_1, \ldots, \gamma_m$. The intention is that, for a given f from X, the information $\gamma_r(f)$, $r = 1, \ldots, m$ is known. From this information, it is desired to compute the action on f of a further linear functional, γ. However, with only the above 'information', we will have to settle for an approximation to $\gamma(f)$. This will be done by introducing a particular subset of X, from which an element u will be selected with sufficiently nice behaviour, so as to make $\gamma(u)$ a reasonable approximation to $\gamma(f)$.

In what follows, we will consider interpolation problems, namely where the information functionals are point evaluations based on a set of nodes, $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$.

The element u will then be chosen so that $\gamma_r(u) = \gamma_r(f)$, $r = 1, \ldots, m$. Having made this choice, we will determine the interpolant u and the pointwise error estimate $|\gamma(f) - \gamma(u)| = |f(x) - u(x)|$, where x is some fixed point in $\mathbb{R}^n$. Pointwise errors between derivatives will also be discussed, when it can be shown that they exist.

Two considerations are pertinent to the argument. First, we assume that $\gamma, \gamma_1, \ldots, \gamma_m$ form a linearly independent collection of linear functionals. If a dependence exists between $\gamma_1, \ldots, \gamma_m$, then one or more values $\gamma_i(f)$ contributes no additional information, and so can be discarded without altering the problem. If γ is a linear combination of $\gamma_1, \ldots, \gamma_m$, then $\gamma(v)$ can be computed exactly, for all v in X and so our problem becomes trivial.

Our second point centres on the fact that, without any restrictions on the class of functions, there is no hope of controlling $\gamma(f)$ simply by knowing the values of $\gamma_1(f), \ldots, \gamma_m(f)$. This reasoning follows from the fact that the linear independence of $\gamma, \gamma_1, \ldots, \gamma_m$ implies the existence of an element g in X such that $\gamma(g) = 1$, but $\gamma_r(g) = 0$, $r = 1, \ldots, m$. Consequently, the function $v = f + \alpha g$, where α is real, is such that $\gamma_r(v) = \gamma_r(f)$, $r = 1, \ldots, m$ as required, but $\gamma(v) = \gamma(f) + \alpha$, and hence, may be unbounded. Classical restrictions on functions, so as to rule out this phenomenon, usually consist of bounding certain derivatives, as will be seen in later chapters. For the moment, we will assume that a semi-inner product $\langle \cdot, \cdot \rangle$ is defined on X for this purpose. Thus, $\langle \cdot, \cdot \rangle$ is a complex-valued, quadratic form satisfying all of the properties of the usual inner product, except that $\langle v, v \rangle = 0$ does not necessarily imply that $v = 0$. In addition, we shall assume that if the data is small, then $\gamma(v)$ is small in comparison. More precisely, we require the existence of a constant C such that, whenever v lies in X, and $\gamma_r(v) = 0$, $r = 1, \ldots, m$, then $|\gamma(v)| \leq C\sqrt{\langle v, v \rangle}$. An important consequence of this assumption is that, if $\langle v, v \rangle$ and $\gamma_r(v)$ are zero, $r = 1, \ldots, m$, then $\gamma(v) = 0$.

Now consider the set X_0 defined by $X_0 = \{v \in X : \langle v, v \rangle = \gamma_1(v) = \cdots = \gamma_m(v) = 0\}$. The addition of any element in X_0 to an element of X changes nothing in terms of our problem, so it is clear that X_0 should be factored out of the discussion. We will go one step further and choose $\gamma_1, \dots, \gamma_m$ so that X_0 is reduced to the trivial set. In terms of interpolation problems, this often involves ensuring the nodes form a unisolvent set with respect to the kernel of $\langle \cdot, \cdot \rangle$.

It is also possible to factor X in such a way that $\langle \cdot, \cdot \rangle$ becomes a genuine inner product. Indeed, if K is the kernel of the quadratic form, and P is a projection from X onto K, then by writing $v = (v - Pv) + Pv$, one achieves a suitable factorisation. An alternative is to modify $\langle \cdot, \cdot \rangle$ in such a way that it becomes a genuine inner product — this is the method proposed by Golomb and Weinberger, and one that introduces our first restriction on m. Let $K = \{v \in X : \langle v, v \rangle = 0\}$. Assuming the dimension of K is ℓ, and $m \geq \ell$, then at most ℓ of $\gamma_1, \dots, \gamma_m$ are linearly independent over K. If necessary, we will re-order $\gamma_1, \dots, \gamma_m$ so that $\gamma_1, \dots, \gamma_\ell$ are linearly independent over K. The latter can then be used to define a bilinear form $(\cdot, \cdot)$ on X. For all u and v in X, let

$$(u, v) = \langle u, v \rangle + \sum_{i=1}^{\ell} \gamma_i(u)\overline{\gamma_i(v)}. \tag{2.1}$$

Suppose now, that $(u, u) = 0$, for some u in V. Then $\langle u, u \rangle$ and $\gamma_1(u), \dots, \gamma_\ell(u)$ are all zero, so that $u \in K$. Furthermore, since $\gamma_{\ell+1}, \dots, \gamma_m$ can be expressed in terms of $\gamma_1, \dots, \gamma_\ell$ on K, it follows that $\gamma_r(u) = 0$, $r = 1, \dots, m$. Thus u is the zero element in X and consequently, $(\cdot, \cdot)$ is a genuine inner product on X. We can then complete X to form a Hilbert space, H whose norm, induced from the inner product, will be denoted by $\| \cdot \|$. It should be noted that any N functionals from $\gamma_1, \dots, \gamma_m$ can be chosen to build an inner product of the form

$$(u, v) = \langle u, v \rangle + \sum_{i=1}^{N} \gamma_i(u)\overline{\gamma_i(v)},$$

as long as $\{\gamma_1, \dots, \gamma_N\}$ is linearly independent over K. A common choice is therefore to use all m information functionals.

At this point, we want to further constrain $\gamma_1, \ldots, \gamma_m$ and γ to be bounded linear functionals on the Hilbert space H. Consequently, by the Riesz representation theorem, there exist elements q_r in H, $r = 1, \ldots, m$ such that, for all v in H, $\gamma_r(v) = (v, q_r)$, $r = 1, \ldots, m$. In what follows, we will use q_r to denote these representers — q_0 will also be used, where appropriate, to denote the representer for γ.

Recall now, that we introduced the quadratic form $\langle \cdot, \cdot \rangle$ to impose some restriction on $|\gamma(v)|$ when v interpolates f. This idea can now be formulated in our Hilbert space as the set of admissible interpolants C_f, defined by

$$C_f = \{v \in H : (v, v) \leq \tau^2 \text{ and } \gamma_r(v) = \gamma_r(f), r = 1, \ldots, m\},$$

where

$$\tau^2 = \langle f, f \rangle + \sum_{i=1}^{\ell} |\gamma_i(f)|^2.$$

Writing

$$C_f = \{v \in H : (v, v) \leq \tau^2\} \cap \left[\bigcap_{r=1}^{m} \{v \in H : \gamma_r(v) = \gamma_r(f)\} \right],$$

we see that C_f is the intersection of a closed ball in H with finitely many hyperplanes $\{v \in H : \gamma_r(v) = \gamma_r(f)\}$, $r = 1, \ldots, m$. Therefore, C_f is closed. It is also convex since, if u and v are chosen from C_f, then for $0 \leq \theta \leq 1$,

$$\begin{aligned} \gamma_r(\theta u + (1 - \theta)v) &= \theta \gamma_r(u) + \gamma_r(v) - \theta \gamma_r(v) \\ &= \gamma_r(v) \\ &= \gamma_r(f), \qquad r = 1, \ldots, m, \end{aligned}$$

and

$$
\begin{aligned}
(\theta u + (1-\theta)v, \theta u + (1-\theta)v) &= \theta^2(u,u) + 2\theta(1-\theta)(u,v) + (1-\theta)^2(v,v) \\
&\leq \tau^2\{\theta^2 + 2\theta(1-\theta) + (1-\theta)^2\} \\
&= \tau^2.
\end{aligned}
$$

Hence, C_f possesses a unique element u of minimal norm, that is, $u \in C_f$ and

$$(u,u) = \inf\{(v,v) : v \in H \text{ and } \gamma_r(v) = \gamma_r(f), r = 1, \ldots, m\}.$$

However, this characterisation may be enhanced to provide us with a method of calculating u. Set

$$G = \{v \in H : \gamma_r(v) = 0, r = 1, \ldots, m\}.$$

Now, for any element h in C_f we can find an element v in G and a real number λ such that $h = u + \lambda v$. Since u is the minimal norm interpolant

$$0 \leq \|u + \lambda v\|^2 - \|u\|^2 = 2\lambda(u,v) + \lambda^2\|v\|^2,$$

and, similarly,

$$0 \leq -2\lambda(u,v) + \lambda^2\|v\|^2.$$

Combining these, we have,

$$2|\lambda||(u,v)| \leq \lambda^2\|v\|^2.$$

As $|\lambda| \to 0$, this inequality becomes untenable unless $(u,v) = 0$. Furthermore, if $(u,v) = 0$, then

$$\|u + \lambda v\|^2 = \|u\|^2 + \lambda^2\|v\|^2 \geq \|u\|^2,$$

with equality only if λ or v are zero. Therefore, $u \in C_f$ and $(u, v) = 0$ for all v in G completely characterises the element u.

Writing G as $\{v \in H : (v, q_r) = 0, r = 1, \ldots, m\}$ we then see that

$$G^\perp = \left\{ \sum_{r=1}^{m} c_r q_r : c_r \in \mathbb{C}, r = 1, \ldots, m \right\}.$$

Since u lies in $G^\perp$, we can therefore find coefficients $\lambda_1, \ldots, \lambda_m$ in $\mathbb{R}$ such that $u = \sum_{r=1}^{m} \lambda_r q_r$. The interpolation conditions can now be formulated as

$$\sum_{r=1}^{m} \lambda_r \gamma_s(q_r) = \gamma_s(f), \qquad s = 1, \ldots, m,$$

which, on using the representers again, yields the system

$$\sum_{r=1}^{m} \lambda_r (q_s, q_r) = \gamma_s(f), \qquad s = 1, \ldots, m.$$

This gives us a method of calculating the interpolant. Expressing the conditions as a matrix equation, we have $Ax = d$, where $A_{ij} = (q_i, q_j)$, $x = (\lambda_1, \ldots, \lambda_m)^T$ and d is the data, $(\gamma_1(f), \ldots, \gamma_m(f))^T$. Finally, let $c_1, \ldots, c_m$ be any set of real numbers. Then,

$$\begin{aligned} \sum_{r,s=1}^{m} c_r c_s A_{rs} &= \sum_{r,s=1}^{m} c_r c_s (q_r, q_s) \\ &= \left(\sum_{r=1}^{m} c_r q_r, \sum_{s=1}^{m} c_s q_s \right) \\ &\geq 0. \end{aligned}$$

Hence, the interpolation matrix is positive definite. Furthermore, it is the Gramian of the linearly independent representers, and therefore, non-singular (c.f. Davis [7] **8.7.2**).

Having described the minimal norm interpolant u, we are going to use $\gamma(u)$ as our estimate for $\gamma(f)$. Thus we aim, in the next section, to bound $|\gamma(f) - \gamma(u)|$.

2.2 Deriving error estimates

The basic error estimate of the theory depends on a certain element of H which we describe in the next two lemmas.

2.2.1 Lemma *Let* w *be any element of unit norm in* G *satisfying*

$$\gamma(\mathrm{w}) = \sup\{|\gamma(v)| : v \in G \text{ and } \|v\| = 1\},$$

and let R *denote the representer for* γ *in* G*, that is, the element of* G *which satisfies, for all* v *in* G, $\gamma(v) = (v, R)$*. Then* $\gamma(\mathrm{w}) = \|R\|$ *and this defines* w *uniquely as* $R/\|R\|$.

Proof. On one hand,

$$\sup_{v\in G}\{|\gamma(v)| : \|v\| = 1\} = \sup_{v\in G}\{|(v, R)| : \|v\| = 1\} \leq \|R\|,$$

by Schwarz's inequality. However, since $R/\|R\| \in G$,

$$\sup_{v\in G}\{|\gamma(v)| : \|v\| = 1\} \geq \gamma\Big(\frac{R}{\|R\|}\Big) = \frac{(R,R)}{\|R\|} = \|R\|.$$

From this we also see that there is at least one candidate for w, namely $R/\|R\|$. To see that w is unique, suppose now that u and w can be found in G, both having unit norm and satisfying

$$\gamma(u) = \gamma(\mathrm{w}) = \sup_{v\in G}\{|\gamma(v)| : \|v\| = 1\} = \|R\|.$$

Then $\|u + \mathrm{w}\| \leq \|u\| + \|\mathrm{w}\| = 2$, but $2\|R\| = \gamma(u + \mathrm{w}) = (u + \mathrm{w}, R) \leq \|u + \mathrm{w}\|\|R\|$. Hence, $\|u+\mathrm{w}\| = 2$. From the underlying theory of Hilbert spaces, we know the norm $\|\cdot\|$ is strictly convex (c.f. Friedman [12] **6.1.3**). Therefore, $\|u\| = \|\mathrm{w}\|$ and the uniqueness follows directly. ∎

2.2.2 Lemma *Let* w *be the unique element of unit norm in* G *which satisfies* $\gamma(\mathrm{w}) = \sup\{|\gamma(v)| : v \in G \text{ and } \|v\| = 1\}$, *and let* $q_0, q_1, \ldots, q_m$ *denote the representers for the linear functionals* $\gamma, \gamma_1, \ldots, \gamma_m$. *Then there exist coefficients* $\omega_0, \ldots, \omega_m \in \mathbb{C}$ *such that* $\mathrm{w} = \sum_{r=0}^{m} \omega_r q_r$.

Proof. Let G_0 be the set

$$\{v \in H : \gamma_r(v) = 0, r = 1, \ldots, m \text{ and } \gamma(v) = 0\},$$

and let $P : G \longrightarrow G_0$ be the orthogonal projection of G onto G_0. Then $\gamma(\mathrm{w} - P\mathrm{w}) = \gamma(\mathrm{w})$ and since, by the previous lemma, w is unique, we conclude that $P\mathrm{w} \equiv 0$. Thus, $\mathrm{w} \in G \cap G_0^{\perp}$. Writing G_0 as $\{v \in H : (v, q_r) = 0, r = 0, \ldots, m\}$, we see that

$$G_0^{\perp} = \left\{ \sum_{r=0}^{m} \lambda_r q_r : \lambda_r \in \mathbb{C}, r = 0, \ldots, m \right\},$$

and the assertion of the lemma follows immediately. ∎

Noting that G is a subspace of H having co-dimension m, whilst G_0 is a subspace of G having co-dimension $m+1$ in H, it follows that $G \cap G_0^{\perp}$ has only one dimension. Therefore, the conditions $\mathrm{w} \in G \cap G_0^{\perp}$ and $\|\mathrm{w}\| = 1$ specify w uniquely up to a factor ± 1.

Returning to the error estimate, the difference between f and its minimal norm interpolant u is an element of G. Thus,

$$\gamma(\mathrm{w}) \geq \left| \frac{\gamma(f-u)}{\|f-u\|} \right|,$$

and so we have the estimate: $|\gamma(f-u)|^2 \leq \gamma(\mathrm{w})^2 (f-u, f-u)$. Now,

$$(f, f) \quad = \quad (f - u + u, f - u + u)$$

$$= \quad (f-u, f-u) + (u,u) + 2(u, f-u),$$

but, as we noted earlier, u lies in $G^{\perp}$. Therefore, $(f-u, f-u) = (f,f) - (u,u)$. Moreover, since $\gamma_r(f) = \gamma_r(u)$, $r = 1, \ldots, m$, $(f,f) - (u,u) = \langle f,f \rangle - \langle u,u \rangle$ and the error estimate reduces to

$$\begin{aligned} |\gamma(f-u)|^2 &\leq \gamma(\mathrm{w})^2 \{\langle f,f \rangle - \langle u,u \rangle\} \\ &\leq \gamma(\mathrm{w})^2 \langle f,f \rangle. \end{aligned}$$

This bound is optimal in the sense that, if f lies in $G \cap G_0^{\perp}$, then $u \equiv 0$ and $f = \|f\|\mathrm{w}$, yielding equality in the estimate. Of course, the calculation of $\gamma(\mathrm{w})$ may be difficult, but as we saw in the proof of **2.2.1**, $\gamma(\mathrm{w}) = \|R\|$, where R is the representer for γ in G. It is easily seen that this can then be relaxed in the following way. Let G_1 be a subspace of H containing G, and let R_1 be the representer for γ in G_1. Then

$$\|R\| = \sup_{v \in G}\{|\gamma(v)| : \|v\| = 1\} \leq \sup_{v \in G_1}\{|\gamma(v)| : \|v\| = 1\} = \|R_1\|,$$

using similar arguments to those seen in **2.2.1**. We may therefore bound $\gamma(\mathrm{w})$ by $\|R_1\|$, and indeed, this idea will be pursued later in section 3.2. However, we now turn to an alternative estimate of the error.

2.3 An alternative error estimate

The following theorem illustrates an alternative error estimate which yields yet more information about minimal norm interpolation, and extends the results of Golomb and Weinberger.

2.3.1 Theorem *Given any f from H, let u denote the interpolant based on the information $\gamma_1(f), \ldots, \gamma_m(f)$, and let v denote the interpolant based on $\gamma(f), \gamma_1(f), \ldots, \gamma_m(f)$. Let w be the unique element of unit norm in G for which*

$$\gamma(\mathrm{w}) = \sup\{|\gamma(v)| : v \in G \text{ and } \|v\| = 1\}.$$

Then, expressing v as $\sum_{r=0}^{m} \alpha_r q_r$,

$$|\gamma(f) - \gamma(u)| = |\alpha_0|\gamma(w)^2.$$

Proof. Let $z = \sum_{r=0}^{m} \sigma_r q_r$ denote the cardinal interpolant for which $\gamma(z) = 1$, and $\gamma_r(z) = 0$, $r = 1, \ldots, m$. In addition, let v_1 be the interpolant satisfying $\gamma(v_1) = \gamma(f - u)$, and $\gamma_r(v_1) = 0$, $r = 1, \ldots, m$. Then $v_1 = v - u = \alpha_0 q_0 + \sum_{r=1}^{m} \beta_r q_r$, for some $\beta_1, \ldots, \beta_m$. Moreover, $v_1 = \gamma(f - u)z$, so that, using the linear independence of the representers, $\alpha_0 = \gamma(f - u)\sigma_0$. To evaluate σ_0, we examine two relations.

On the one hand, **2.2.2** allows us to write w as $\sum_{r=0}^{m} \omega_r q_r$. Therefore, observing that $\mathrm{w} = \gamma(\mathrm{w})z$, we know $\sigma_0 = \omega_0/\gamma(\mathrm{w})$. On the other, recalling that $\|\mathrm{w}\| = 1$ and $\gamma_r(\mathrm{w}) = 0$, $r = 1, \ldots, m$,

$$\begin{aligned} 1 = \|\mathrm{w}\|^2 &= \left(\mathrm{w}, \sum_{r=0}^{m} \omega_r q_r\right) \\ &= \sum_{r=1}^{m} \omega_r \gamma_r(\mathrm{w}) + \omega_0 \gamma(\mathrm{w}) \\ &= \omega_0 \gamma(\mathrm{w}). \end{aligned}$$

Hence, $\omega_0 = \gamma(\mathrm{w})^{-1}$ and overall, we have $\sigma_0 = \gamma(\mathrm{w})^{-2}$. Therefore, $\alpha_0 = \gamma(f - u)\gamma(\mathrm{w})^{-2}$ and so, $|\gamma(f - u)| = |\alpha_0|\gamma(\mathrm{w})^2$. ∎

An immediate consequence of this estimate is that it relates the error to the stability of the interpolation matrices. Small perturbations in the data $\gamma_1(f), \ldots, \gamma_m(f)$ create perturbations in $|\alpha_0|$ dependent on the size of the condition numbers of the matrix. A thorough analysis of condition numbers arising in the more common applications in interpolation may be found in the work of Ball, Sivakumar, and Ward [5], Narcowich and Ward [24], and Sun [38], but we refer to the work of Schaback [30] for a specific analysis of this phenomenon. However, we will remark here that, in the context of interpolation problems, a large perturbation in a single point evaluation $\gamma_i(f)$ may cause large variations in the error, even in the case of well-conditioned matrices. For interpolation problems, where point evaluations are the main concern, this provides further motivation for using $\langle \cdot, \cdot \rangle$ to control the size of the derivatives of elements in the Hilbert space.

As a means of estimating the error, this formulation has some attractive qualities, although matrix theory does not lend itself well to a thorough analysis. Using Cramer's rule, we can, however, write α_0 as

$$\frac{\begin{vmatrix} (f,q_0) & (q_1,q_0) & \cdots & (q_m,q_0) \\ (f,q_1) & (q_1,q_1) & \cdots & (q_m,q_1) \\ \vdots & \vdots & \ddots & \vdots \\ (f,q_m) & (q_1,q_m) & \cdots & (q_m,q_m) \end{vmatrix}}{\begin{vmatrix} (q_0,q_0) & (q_1,q_0) & \cdots & (q_m,q_0) \\ (q_0,q_1) & (q_1,q_1) & \cdots & (q_m,q_1) \\ \vdots & \vdots & \ddots & \vdots \\ (q_0,q_m) & (q_1,q_m) & \cdots & (q_m,q_m) \end{vmatrix}}$$

showing the stark similarity α_0 shares with divided differences in the classical approximation theory of polynomial interpolation (c.f. Davis [7]).

Throughout this chapter, we have treated the representers of the functionals γ, and $\gamma_1, \ldots, \gamma_m$ as known quantities — quite naively at times. As we shall see in the remaining chapters, more respect is, indeed, deserved, but we conclude the Hilbert space theory with a method of calculating the representers for the ℓ functionals which make up the inner product (2.1).

2.3.2 Lemma *Suppose elements $p_1, \ldots, p_\ell$ can be found in K such that, together with $\gamma_1, \ldots, \gamma_\ell$, they form a biorthonormal set, that is, $\gamma_i(p_j) = \delta_{ij}$, $1 \le i, j \le \ell$, where δ_{ij} is the Kronecker delta. Then p_i is the representer for γ_i in H, $1 \le i \le \ell$.*

Proof. For any v in H, we have

$$\begin{aligned} (v, p_j) &= \langle v, p_j \rangle + \sum_{i=1}^{\ell} \gamma_i(v)\gamma_i(p_j) \\ &= \langle v, p_j \rangle + \gamma_j(v). \end{aligned}$$

However, $|\langle v, p_j \rangle| \le \sqrt{\langle v, v \rangle \langle p_j, p_j \rangle} = 0$. Thus, $(v, p_j) = \gamma_j(v)$ as required. ∎

Since the representers $p_1, \ldots, p_\ell$ span K, any interpolation process that uses $a_1, \ldots, a_\ell$ as nodes will preserve elements in K.

Chapter 3

Error Analysis

We devote this chapter to a method for analysing the error estimate

$$|\gamma(f) - \gamma(u)| \leq \gamma(\mathrm{w})\sqrt{\langle f, f\rangle}, \tag{3.1}$$

in the case of a specific interpolation problem, where the abstract Hilbert space H satisfies a small set of conditions. These assumptions will be our starting point for the analysis, and moreover, our guide through the next chapter, where we describe the general problem of multivariate interpolation.

3.1 Minimal assumptions for error analysis

Our attention is now firmly concentrated on interpolation problems — the 'information' functionals will be point evaluations based on a set of nodes $\{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$. Thus, $\gamma_r(f)$ will be the value $f(a_r)$, $r = 1, \ldots, m$.

3.1.1 Assumption *The assumptions on the Hilbert space H now comprise the following:*

- *(i) There exists a non-negative integer j such that $H \subset C^j(\mathbb{R}^n)$. This ensures point evaluations exist.*

(ii) Given a multi-index β satisfying $|\beta| \le j$, and a fixed point x in $\mathbb{R}^n$, $\gamma(f)$ will be the value $(D^\beta f)(x)$. We will therefore assume, for all multi-indices α satisfying $|\alpha| \le j$, the existence of a constant C, dependent on α and x, but not on f, such that $|(D^\alpha f)(x)| \le C\|f\|$.

(iii) The points $a_1, \ldots, a_\ell$ should be unisolvent with respect to K, and $p_1, \ldots, p_\ell$ should be chosen from K so that they satisfy $p_i(a_j) = \delta_{ij}$, $1 \le i, j \le \ell$.

(iv) There exists a function ϕ in $C^{2j}(\mathbb{R}^n)$ such that, for each x in $\mathbb{R}^n$, $\phi(x) = \overline{\phi(-x)}$, and for each multi-index β satisfying $|\beta| \le j$,

$$\begin{aligned} R_x^\beta(y) &= (-1)^{|\beta|}(D^\beta\phi)(y-x) - \sum_{r=1}^{\ell}(D^\beta p_r)(x)\phi(y-a_r) + \sum_{r=1}^{\ell}(D^\beta p_r)(x)p_r(y) \\ &\quad -(-1)^{|\beta|}\sum_{r=1}^{\ell} p_r(y)(D^\beta\phi)(a_r - x) + \sum_{r,s=1}^{\ell} p_r(y)(D^\beta p_s)(x)\phi(a_r - a_s), \end{aligned}$$

defines a function in H which is the representer for the functional γ (the superscript will be dropped when $\beta = 0$).

Some remarks should be made at this stage. First, assumptions (i) and (ii) would be immediate if we assumed that H were continuously embeddable in C^j. However, in our case, this division of labour illustrates the path of future arguments: given a space H, show that the chosen linear functional has a well-defined action, that it is bounded, and then derive the representer.

Secondly, using **2.3.2**, assumption (iii) confirms that $p_1, \ldots, p_\ell$ are the representers for the point evaluations at $a_1, \ldots, a_\ell$, although the unisolvency of these points also forces X_0 to be the trivial subspace. Finally, it should be noted that, when $\beta = 0$,

$$R_x(y) = \phi(y-x) - \sum_{r=1}^{\ell} p_r(x)\phi(y-a_r) - \sum_{r=1}^{\ell} p_r(y)\phi(a_r - x)$$

$$+\sum_{r,s=1}^{\ell} p_r(y)p_s(x)\phi(a_r - a_s) + \sum_{r=1}^{\ell} p_r(x)p_r(y),$$

defines the *reproducing kernel* for H, that is, for any admissable point x in $\mathbb{R}^n$, the element R_x in H which satisfies the relation $(f, R_x) = f(x)$, for all f in H. (c.f. Shapiro [35] **6.2**). Clearly $R_x(y) = \overline{R_y(x)}$, but this Hermitian symmetry does not occur when $|\beta| > 0$.

3.2 Power functions

Some additional notation is required. For $k \geq \ell$, let G^k denote the space

$$\{v \in H : \gamma_1(v) = \cdots = \gamma_k(v) = 0\},$$

and let G_0^k denote the space

$$\{v \in H : \gamma_1(v) = \cdots = \gamma_k(v) = 0 \text{ and } \gamma(v) = 0\}.$$

For convenience, we will continue to use $q_0, q_1, \ldots, q_m$ to denote the representers for $\gamma, \gamma_1, \ldots, \gamma_m$.

3.2.1 Lemma *Let k be any integer, greater than or equal to ℓ. Let q denote the representer for γ in G^k. Then there exist coefficients $\alpha_0, \ldots, \alpha_k$ such that $q = \sum_{r=0}^{k} \alpha_r q_r$. If $q \neq 0$, the conditions $\alpha_0 = 1$ and $q \in G^k$ then characterise q completely.*

Proof. Since $G_0^k \subset G^k$, we have, for all v in G_0^k, $\gamma(v) = (v, q) = 0$. However, this is precisely the statement that q lies in $(G_0^k)^{\perp}$, and therefore, q can be written as a linear combination of $q_0, \ldots, q_k$. To complete the proof, we examine two alternatives for the value $\gamma(q)$. On the one hand, $\gamma(q) = (q, q) = \|q\|^2$, whilst on the other,

$$\gamma(q) = \gamma(q) - \alpha_0\gamma(q) + \alpha_0\gamma(q)$$

$$
\begin{aligned}
&= \gamma(q) - \alpha_0\gamma(q) + (q, \alpha_0 q_0) \\
&= \left(q, \alpha_0 q_0 + \sum_{r=1}^{k} \alpha_r q_r\right) + \gamma(q)(1-\alpha_0) \\
&= (q,q) + \gamma(q)(1-\alpha_0),
\end{aligned}
$$

recalling that $(q, q_r) = \gamma_r(q) = 0,\ r = 1, \ldots, m$. Combining the two, we have either $\gamma(q) = 0$, or $\alpha_0 = 1$. However, since q lies in $G^k \cap (G_0^k)^\perp$, $\gamma(q)$ can only be zero if q is the zero element. Hence, we conclude that, if $q \neq 0$, $\alpha_0 = 1$. ∎

3.2.2 Lemma *Let k be any integer, greater than or equal to ℓ, and let $\alpha_1, \ldots, \alpha_k$ satisfy the conditions*

$$
\sum_{r=1}^{k} \alpha_r p_s(a_r) = -(D^\beta p_s)(x), \qquad s = 1, \ldots, \ell. \tag{3.2}
$$

If Q denotes the element $\sum_{r=1}^{k} \alpha_r R_{a_r} + R_x^\beta$, then $\|Q\|^2$ is given by either of the expressions,

$$
\begin{aligned}
(-1)^{|\beta|}(D^{2\beta}\phi)(0) &- (-1)^{|\beta|}\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta\phi)(a_r - x) \\
&- \sum_{r=1}^{k}\alpha_r\sum_{s=1}^{\ell}(D^\beta p_s)(x)\phi(a_s - a_r) + \sum_{r=1}^{k}\alpha_r(D^\beta\phi)(x - a_r),
\end{aligned}
$$

or,

$$
(-1)^{|\beta|}(D^{2\beta}\phi)(0) + \sum_{r,s=1}^{k}\alpha_r\alpha_s\phi(a_s - a_r) + \sum_{r=1}^{k}\alpha_r\{(D^\beta\phi)(x-a_r) + (-1)^{|\beta|}(D^\beta\phi)(a_r - x)\}.
$$

Proof. Using assumption (iv), we can write

$$
\begin{aligned}
Q(y) = \sum_{r=1}^{k}\alpha_r R_{a_r}(y) + R_x^\beta(y) = \\
\sum_{r=1}^{k}\alpha_r\Bigg\{\phi(y-a_r) &- \sum_{s=1}^{\ell}p_s(a_r)\phi(y-a_s) - \sum_{s=1}^{\ell}p_s(y)\phi(a_s - a_r) \\
&+ \sum_{s,t=1}^{\ell}p_s(y)p_t(a_r)\phi(a_s - a_t) + \sum_{s=1}^{\ell}p_s(a_r)p_t(y)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&+(-1)^{|\beta|}(D^\beta\phi)(y-x)-\sum_{r=1}^{\ell}(D^\beta p_r)(x)\phi(y-a_r)\\
&\qquad+\sum_{r=1}^{\ell}(D^\beta p_r)(x)p_r(y)-(-1)^{|\beta|}\sum_{r=1}^{\ell}p_r(y)(D^\beta\phi)(a_r-x)\\
&\qquad\qquad+\sum_{r,s=1}^{\ell}p_r(y)(D^\beta p_s)(x)\phi(a_r-a_s)\\
=\ &\sum_{r=1}^{k}\alpha_r\phi(y-a_r)-\sum_{r=1}^{k}\alpha_r\sum_{s=1}^{\ell}p_s(y)\phi(a_s-a_r)-(-1)^{|\beta|}\sum_{s=1}^{\ell}p_s(y)(D^\beta\phi)(a_s-x)\\
&+(-1)^{|\beta|}(D^\beta\phi)(y-x)-\sum_{s=1}^{\ell}\phi(y-a_s)\Big\{\sum_{r=1}^{k}\alpha_r p_s(a_r)+(D^\beta p_s)(x)\Big\}\\
&\qquad+\sum_{s,t=1}^{\ell}p_s(y)\phi(a_s-a_t)\Big\{\sum_{r=1}^{k}\alpha_r p_t(a_r)+(D^\beta p_t)(x)\Big\}\\
&\qquad\qquad+\sum_{s=1}^{\ell}p_s(y)\Big\{\sum_{r=1}^{k}\alpha_r p_s(a_r)+(D^\beta p_s)(x)\Big\}.
\end{aligned}
$$

Using the conditions on $\alpha_1,\dots,\alpha_k$ in the hypothesis (3.2), it is easily seen that the last three terms in Q above vanish to leave,

$$
\begin{aligned}
Q(y)\ &=\ (-1)^{|\beta|}(D^\beta\phi)(y-x)+\sum_{r=1}^{k}\alpha_r\phi(y-a_r)\\
&\quad-\sum_{s=1}^{\ell}p_s(y)\Big\{\sum_{r=1}^{k}\alpha_r\phi(a_s-a_r)+(-1)^{|\beta|}(D^\beta\phi)(a_s-x)\Big\}.
\end{aligned}
\tag{3.3}
$$

Thus, $\|Q\|^2=(Q,Q)=\gamma(Q)=(D^\beta Q)(x)$, and

$$
\begin{aligned}
(D^\beta Q)(x)\ &=\ (-1)^{|\beta|}(D^{2\beta}\phi)(0)+\sum_{r=1}^{k}\alpha_r(D^\beta\phi)(x-a_r)\\
&\quad-\sum_{s=1}^{\ell}(D^\beta p_s)(x)\sum_{r=1}^{k}\alpha_r\phi(a_s-a_r)-(-1)^{|\beta|}\sum_{s=1}^{\ell}(D^\beta p_s)(x)(D^\beta\phi)(a_s-x).
\end{aligned}
$$

To obtain the second form of $\|Q\|$, we write

$$
\|Q\|^2\ =\ \Big(Q,\sum_{s=1}^{k}\alpha_r R_{a_s}+R_x^\beta\Big)
$$

$$
\begin{aligned}
&= \sum_{s=1}^{k} \alpha_s Q(a_s) + (D^\beta Q)(x) \\
&= \sum_{s=1}^{k} \alpha_s \Big\{ (-1)^{|\beta|}(D^\beta \phi)(a_s - x) + \sum_{r=1}^{k} \alpha_r \phi(a_s - a_r) \\
&\quad - \sum_{t=1}^{\ell} p_t(a_s) \Big\{ \sum_{r=1}^{k} \alpha_r \phi(a_t - a_r) + (-1)^{|\beta|}(D^\beta \phi)(a_t - x) \Big\} \Big\} \qquad (3.4) \\
&\quad + (-1)^{|\beta|}(D^{2\beta}\phi)(0) + \sum_{r=1}^{k} \alpha_r (D^\beta \phi)(x - a_r) \\
&\quad - \sum_{t=1}^{\ell} (D^\beta p_t)(x) \Big\{ \sum_{r=1}^{k} \alpha_r \phi(a_t - a_r) + (-1)^{|\beta|}(D^\beta \phi)(a_t - x) \Big\}, \qquad (3.5)
\end{aligned}
$$

and use (3.2) to cancel the terms in (3.4) and (3.5). This yields the given form, and completes the proof. ■

3.2.3 Lemma *For any integer $k \geq \ell$, let q denote the representer for γ in G^k. Then there exist coefficients $\alpha_1, \ldots, \alpha_k$ with which $\|q\|^2$ may be written as*

$$
(-1)^{|\beta|}(D^{2\beta}\phi)(0) + \sum_{r,s=1}^{k} \alpha_r \alpha_s \phi(a_s - a_r) + \sum_{r=1}^{k} \alpha_r \{ (D^\beta \phi)(x - a_r) + (-1)^{|\beta|}(D^\beta \phi)(a_r - x) \}.
$$

Proof. From **3.2.1**, we know there exist coefficients $\alpha_1, \ldots, \alpha_k$ such that

$$
q = \sum_{r=1}^{k} \alpha_r R_{a_r} + R_x^\beta .
$$

Then, since q lies in G^k, it is straightforward to see that, for $1 \leq s \leq \ell$,

$$
\begin{aligned}
0 = q(a_s) &= \sum_{r=1}^{k} \alpha_r R_{a_r}(a_s) + R_x^\beta(a_s) \\
&= \sum_{r=1}^{k} \alpha_r (R_{a_r}, p_s) + (R_x^\beta, p_s) \\
&= \sum_{r=1}^{k} \alpha_r p_s(a_r) + (D^\beta p_s)(x),
\end{aligned}
$$

and the result follows from **3.2.2**. ■

Let us momentarily consider the case when $\beta = 0$, so that $\gamma(g) = g(x)$, and ϕ is real-valued. Then, from assumption (iv), $\phi(x) = \phi(-x)$, and the second form of $\|q\|$ above then tells us that our error estimator $\gamma(\mathrm{w})$ satisfies

$$\gamma(\mathrm{w})^2 = [\mathrm{w}(x)]^2 \leq \|q\|^2 = \phi(0) + \sum_{r,s=1}^{k} \alpha_r \alpha_s \phi(a_s - a_r) + 2\sum_{r=1}^{k} \alpha_s \phi(a_s - x),$$

recalling our comments at the end of section 2.2. Treated as a function of x, w is more commonly known as the 'power function', as referred to in the work of Schaback [29, 30] and Powell [25].

3.2.4 Lemma *Given any f from H, let u be the interpolant based on the information $f(a_1), \ldots, f(a_m)$. Then there exist coefficients $\alpha_1, \ldots, \alpha_m$ satisfying*

$$\sum_{r=1}^{m} \alpha_r R_{a_r}(a_s) = -R_x^{\beta}(a_s), \qquad s = 1, \ldots, m,$$

such that

$$\begin{aligned} |(D^{\beta} f)(x) - (D^{\beta} u)(x)| \;\; &\leq \;\; \sqrt{\langle f, f\rangle} \Big|(-1)^{|\beta|}(D^{2\beta}\phi)(0) + \sum_{r=1}^{m} \alpha_r (D^{\beta}\phi)(x - a_r) \\ &- \sum_{s=1}^{\ell} (D^{\beta} p_s)(x) \sum_{r=1}^{m} \alpha_r \phi(a_s - a_r) - (-1)^{|\beta|} \sum_{r=1}^{\ell} (D^{\beta} p_r)(x)(D^{\beta}\phi)(a_s - x)\Big|^{\frac{1}{2}}. \end{aligned}$$

Proof. With $\alpha_1, \ldots, \alpha_m$ so defined, **3.2.1** tells us that $\sum_{r=1}^{m} \alpha_r R_{a_r} + R_x^{\beta}$ is the representer for γ in G^m, which we will denote by q. The above error estimate now follows by taking the basic error estimate from the Hilbert space theory (3.1), applying **2.2.1** to obtain $|(D^{\beta} f)(x) - (D^{\beta} u)(x)| \leq \|q\| \sqrt{\langle f, f\rangle}$ and then applying **3.2.3** to find the appropriate form for $\|q\|$. ■

However, this estimate carries a price — the question of the indeterminates $\alpha_1, \ldots, \alpha_m$ and whether $\sum_{r=1}^{m} |\alpha_r|$ is a bounded quantity, for differing sets of interpolation nodes. However, as we remarked at the end of section 2.2, $\gamma(\mathrm{w}) \le \|q\|$ where q is the representer for γ in any subspace of H containing G^m. This leads to the following error estimate, which is far more accessible to analytic techniques, as we will see in the next section.

3.2.5 Theorem *Given any f from H, let u be the interpolant based on the information $f(a_1), \ldots, f(a_m)$. Then, for all multi-indices β satisfying $|\beta| \le j$,*

$$|(D^\beta f)(x) - (D^\beta u)(x)| \le \sqrt{\langle f, f\rangle} \Big| (-1)^{|\beta|}(D^{2\beta}\phi)(0) + \sum_{r,s=1}^{\ell} (D^\beta p_r)(x)(D^\beta p_s)(x)\phi(a_r - a_s)$$
$$- \sum_{r=1}^{\ell} (D^\beta p_r)(x)\Big\{(D^\beta\phi)(x - a_r) + (-1)^{|\beta|}(D^\beta\phi)(a_r - x)\Big\}\Big|^{\frac{1}{2}}.$$

Proof. Let q be the representer for γ in G^ℓ. Then from **3.2.1**, there exist coefficients $\alpha_1, \ldots, \alpha_\ell$ such that

$$q = \sum_{r=1}^{\ell} \alpha_r R_{a_r} + R_x^\beta = \sum_{r=1}^{\ell} \alpha_r p_r + R_x^\beta.$$

Since $q(a_s) = 0$, $s = 1, \ldots, \ell$, we have

$$\begin{aligned} 0 &= \sum_{r=1}^{\ell} \alpha_r p_r(a_s) + R_x^\beta(a_s) \\ &= \alpha_s + (R_x^\beta, p_s) \\ &= \alpha_s + (D^\beta p_s)(x). \end{aligned}$$

The form of $\|q\|$ can now be immediately deduced using **3.2.3**, and the observation that $G^m \subset G^\ell$, $m \ge \ell$, then verifies the error estimate. ■

We will also illustrate an alternative error estimate which, under certain circumstances, provides greater flexibility than the rigid form of error outlined above. The approach, at its most basic, follows Section 7 of Golomb and Weinberger [14]. Even though the rewards will not be seen until Chapter 5, the results which follow belong here, in the general case, where they hint at the strength of this Hilbert space approach to interpolation.

In essence, we will introduce a new linear functional L and calculate the size of $|L(g)|$ for some g in H. In particular, we will choose an integer N satisfying $\ell \le N \le m$, coefficients $\beta_1, \ldots, \beta_N$, and set $L(v) = \gamma(v) + \sum_{r=1}^{N} \beta_r \gamma_r(v)$, $v \in H$. In our problem, if u denotes the minimal norm interpolant constructed from the information $\gamma_1(f), \ldots, \gamma_m(f)$, for some f in H, then $|L(f-u)| = |\gamma(f-u)|$. Denoting the representer for L by Q, we can therefore write

$$|\gamma(f-u)| = |(f-u, Q)| \le \|Q\| \, \|f-u\|,$$

and proceed as before. Clearly, the optimal error bound can be recovered by choosing $N = m$ and ensuring that $q_0(a_s) + \sum_{r=1}^{m} \beta_r (q_r, q_s) = 0$, $s = 1, \ldots, m$, since, by **3.2.1**, this is enough to make Q the representer for γ in G^m. We therefore have a powerful method of error estimation if we can capitalise on our freedom in the choice of $\beta_1, \ldots, \beta_m$. We highlight such a choice in the final lemma of this section.

3.2.6 Lemma *Under the assumptions outlined in* **3.1.1**, *let N be chosen so that $\ell \le N \le m$, and let $\beta_1, \ldots, \beta_N$ satisfy the conditions*

$$\sum_{r=1}^{N} \beta_r p_s(a_r) = -(D^\beta p_s)(x), \qquad s = 1, \ldots, \ell.$$

Let L denote the functional, defined for all v in H by $L(v) = \gamma(v) + \sum_{r=1}^{N} \beta_r \gamma_r(v)$, and let Q denote the representer for L. Then,

$$\begin{aligned} \|Q\|^2 \;=\; & \Big\{(-1)^{|\beta|}(D^{2\beta}\phi)(0) \\ & + \sum_{r,s=1}^{N} \beta_r \beta_s \phi(a_s - a_r) + \sum_{r=1}^{N} \beta_r \Big\{(D^\beta \phi)(x - a_r) + (-1)^{|\beta|}(D^\beta \phi)(a_r - x)\Big\}\Big\}. \end{aligned}$$

Proof. The form of $\|Q\|$ follows immediately upon application of **3.2.2** to Q. ∎

3.3 Polynomial kernels

We now explore the error estimates in more detail in the case when the kernel, K, is π_k, the space of polynomials of degree, at most, k. Let ℓ denote the dimension of π_k, and assume that the set of interpolation nodes is unisolvent with respect to π_k.

Under these assumptions, the representers for the point evaluations at $a_1, \ldots, a_\ell$ form a cardinal basis for π_k and hence, the work ties in with the large body of material on Lagrange polynomial interpolation. We begin with some straightforward results which may be found in the theory of finite elements.

3.3.1 Definition *Let $\ell = \dim \pi_k$, and let $b = \{b_1, \ldots, b_\ell\}$ define a set of points in $\mathbb{R}^n$ which is unisolvent with respect to π_k. Then $L_b : C(\mathbb{R}^n) \longrightarrow \pi_k$ will denote the Lagrange interpolation operator defined by $(L_b f)(b_r) = f(b_r)$, $r = 1, \ldots, \ell$.*

3.3.2 Lemma *Let Ω be a closed, bounded subset of $\mathbb{R}^n$, and let $b_1, \ldots, b_\ell$ be a set of points in Ω, unisolvent with respect to the polynomial space π_k. Let $p_1, \ldots, p_\ell$ be the cardinal functions in π_k, based on $b_1, \ldots, b_\ell$. Suppose $L_b : C(\Omega) \longrightarrow \pi_k$ is as defined in* **3.3.1**. *Then, for all multi-indices β, $L_b f = \sum_{r=1}^{\ell} f(b_r) p_r$, and*

$$\|L_b\|_\beta = \sup_{x \in \Omega} \sum_{r=1}^{\ell} |(D^\beta p_r)(x)|.$$

Proof. Since $p_r(a_s) = \delta_{rs}$, $1 \le r, s \le \ell$, it is clear that the given form of $L_b f$ satisfies the interpolation conditions in **3.3.1**. Moreover, as a linear combination of $p_1, \ldots, p_\ell$, $L_b f$ is uniquely defined, because the unisolvency of $b_1, \ldots, b_\ell$ ensures that the matrix associated with the interpolation conditions is non-singular. Therefore, $\|L_b f\|_\beta$ which is, by definition, $\sup\{\|D^\beta L_b g\|_{\infty,\Omega} : \|g\|_{\infty,\Omega} = 1\}$, satisfies

$$\begin{aligned}
\|L_b\|_\beta &= \sup_{\|g\|_{\infty,\Omega}=1} \sup_{x\in\Omega} \left\{ \left| \sum_{r=1}^{\ell} g(b_r)(D^\beta p_r)(x) \right| \right\} \\
&= \sup_{x\in\Omega} \sup_{\|g\|_{\infty,\Omega}=1} \left\{ \left| \sum_{r=1}^{\ell} g(b_r)(D^\beta p_r)(x) \right| \right\} \\
&\le \sup_{x\in\Omega} \sup_{\|g\|_{\infty,\Omega}=1} \left\{ \sum_{r=1}^{\ell} |(D^\beta p_r)(x)| \|g\|_{\infty,\Omega} \right\} \\
&\le \sup_{x\in\Omega} \left\{ \sum_{r=1}^{\ell} |(D^\beta p_r)(x)| \right\}.
\end{aligned}$$

Since $b_1, \ldots, b_\ell$ are distinct points in $\mathbb{R}^n$, we can construct a continuous function χ, whose value in a neighbourhood of each b_r, $r = 1, \ldots, \ell$ is $(D^\beta p_r)(x)/|(D^\beta p_r)(x)|$, and throughout Ω, $-1 \le \chi \le 1$. Then,

$$\|L_b\|_\beta \ge \sup_{x\in\Omega} \left\{ \left| \sum_{r=1}^{\ell} \chi(b_r)(D^\beta p_r)(x) \right| \right\} = \sup_{x\in\Omega} \left\{ \sum_{r=1}^{\ell} |(D^\beta p_r)(x)| \right\},$$

which completes the proof. ∎

The following is a multivariate analogue of Markov's inequality for algebraic polynomials.

3.3.3 Theorem (Ditzian [9]) *Let k and q be chosen so that $k > 0$ and $0 < q \le \infty$, and let Ω be any bounded, convex set in $\mathbb{R}^n$. Then, for any direction ξ, and for all polynomials p from π_k, we can find a constant C, dependent on q and Ω, such that*

$$\left\| \frac{\partial p}{\partial \xi} \right\|_{q,\Omega} \le Ck^2 \|p\|_{q,\Omega}.$$

3.3.4 Lemma *Let $B(x,r)$ denote the ball $\{y \in \mathbb{R}^n : |y - x| \le r\}$ and let $b = (b_1, \ldots, b_\ell)$ define an ℓ-tuple of points in $\mathbb{R}^n$, which is unisolvent with respect to π_k. Then there exists a constant $\delta > 0$ such that, if $c \in B(b_1, \delta) \times \cdots \times B(b_\ell, \delta)$, then c is also a unisolvent ℓ-tuple.*

Furthermore, if Ω is a closed, bounded, convex set containing each $B(b_i, \delta)$, $i = 1, \ldots, \ell$, and if β is any multi-index, then there exists a constant $K = K(k, \beta, \Omega)$ such that $\|L_c\|_\beta \le K$, for all c in $B(b_1, \delta) \times \cdots \times B(b_\ell, \delta)$.

Proof. The set of all unisolvent ℓ-tuples in $(\mathbb{R}^n)^\ell$ is open (its complement describes an algebraic surface in $(\mathbb{R}^n)^\ell$). This establishes the first part of the Lemma.

Next, for each f in $C(\Omega)$, the mapping $c \mapsto \|L_c f\|_{\infty,\Omega}$ is a continuous mapping from $B(b_1, \delta) \times \cdots \times B(b_\ell, \delta)$ into $\mathbb{R}$. Since the domain of this mapping is compact,

$$\sup\{\|L_c f\|_{\infty,\Omega} : c \in B(b_1, \delta), \ldots, B(b_\ell, \delta)\} < \infty.$$

Therefore, for each f in $C(\Omega)$, the set $\{L_c f\}$ is bounded in $C(\Omega)$. Applying the principle of uniform boundedness (c.f. Friedman [12] **4.5.1**) then tells us that $\{\|L_c\|\}$ is bounded, and so the result holds for $|\beta| = 0$. Repeated application of the Markov inequality detailed in **3.3.3** yields the result for $|\beta| > 0$. ■

3.3.5 Lemma *Let Ω be a closed, bounded subset of $\mathbb{R}^n$ and suppose $L_b : C(\Omega) \longrightarrow \pi_k$ is a Lagrange interpolation operator based on the set of points $\{b_1, \ldots, b_\ell\}$. Let σ be a dilation operator of the form $\sigma(y) = \lambda y$, $\lambda \in \mathbb{R}$, $y \in \mathbb{R}^n$. Then the operator $L_{\sigma(b)} : C(\sigma(\Omega)) \longrightarrow \pi_k$ satisfies $\|L_{\sigma(b)}\|_\beta = \lambda^{-|\beta|}\|L_b\|_\beta$, for all multi-indices β.*

Proof. Suppose $p_1, \ldots, p_\ell$ are the cardinal basis functions for π_k, based on $b_1, \ldots, b_\ell$. Then $(p_r \circ \sigma^{-1})(\sigma(b_s)) = \delta_{rs}$, $r, s = 1, \ldots, \ell$. Since each $p_r \circ \sigma^{-1}$ belongs to π_k, these must be the cardinal functions for $L_{\sigma(b)}$. Now,

$$\begin{aligned} \|L_{\sigma(b)}\|_\beta &= \sup_{x \in \sigma(\Omega)} \left\{ \sum_{r=1}^{\ell} |(D^\beta(p_r \circ \sigma^{-1}))(x)| \right\} \\ &= \lambda^{-|\beta|} \sup_{x \in \sigma(\Omega)} \left\{ \sum_{r=1}^{\ell} |((D^\beta p_r) \circ \sigma^{-1})(x)| \right\} \end{aligned}$$

$$
\begin{aligned}
&= \lambda^{-|\beta|} \sup_{y\in\Omega} \left\{ \sum_{r=1}^{\ell} |((D^\beta p_r) \circ \sigma^{-1})(\sigma(y))| \right\} \\
&= \lambda^{-|\beta|} \sup_{y\in\Omega} \left\{ \sum_{r=1}^{\ell} |(D^\beta p_r)(y)| \right\} \\
&= \lambda^{-|\beta|} \|L_b\|_\beta. \qquad \blacksquare
\end{aligned}
$$

3.3.6 Definition (Adams [2] 4.3) *A domain Ω has the cone property if there exists a finite cone C such that each point $x \in \Omega$ is the vertex of a finite cone C_x, contained in Ω, and congruent to C.*

3.3.7 Lemma (Duchon [11]) *Let Ω be an open subset of $\mathbb{R}^n$ having the cone property. Then there exist constants M, M_1 and $\epsilon_0 > 0$ such that, to each $0 < \epsilon < \epsilon_0$, there corresponds a set T_ϵ within Ω such that*

(i) $B(t,\epsilon) \subset \Omega$, for all t in T_ϵ,

(ii) $\Omega \subset \bigcup_{t\in T_\epsilon} B(t, M\epsilon)$, and

(iii) $\sum_{t\in T_\epsilon} \chi_t \le M_1$, where χ_t is the characteristic function for the ball $B(t, M\epsilon)$.

These tools allow us to prove the following theorem.

3.3.8 Theorem *Let Ω be any open, connected subset of $\mathbb{R}^n$ having the cone property, let $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$ define a π_k-unisolvent subset of Ω, and let h be defined by $\sup_{t\in\Omega} \inf_{a\in\mathcal{A}} |t-a|$.*

Given any f from H, let u denote the minimal norm interpolant based on the data $f(a_1), \ldots, f(a_m)$. Then there exist positive constants h_0, K and C, dependent on Ω, but not on h, such that, for all multi-indices β satisfying $|\beta| \le j$,

$$
|(D^\beta f)(x) - (D^\beta u)(x)| \le \\
\left\{ |(D^{2\beta}\phi)(0)| + 2Kh^{-|\beta|} \max_{0\le|y|\le Ch} \{|(D^\beta\phi)(y)|\} + K^2h^{-2|\beta|} \max_{0\le|y|\le Ch} \{|\phi(y)|\} \right\}^{\frac{1}{2}} \sqrt{\langle f, f\rangle},
$$

whenever $h < h_0$.

Proof. We begin by taking a π_k-unisolvent set of points $\{v_1, \ldots, v_\ell\}$ from $\mathbb{R}^n$. By **3.3.4**, there exists $\delta > 0$ such that every choice of ℓ-tuple from $B(v_1, \delta) \times \cdots \times B(v_\ell, \delta)$ is π_k-unisolvent. Dilation by a factor δ^{-1} creates a new set of points $\{x_1, \ldots, x_\ell\}$ such that the set $B(x_1, 1) \times \cdots \times B(x_\ell, 1)$ also generates unisolvent ℓ-tuples from $(\mathbb{R}^n)^\ell$. Choose $R > 0$ such that $B(x_r, 1) \subset B(0, R)$, $r = 1, \ldots, \ell$.

Now, applying **3.3.7** to Ω yields two constants, ϵ_0 and M, with which the following properties are associated. Firstly, to each $0 < h < \epsilon_0/R$ there corresponds a set of centres T_{Rh} such that, for all t in T_{Rh}, $B(t, Rh) \subset \Omega$, and, secondly, $\Omega \subset \bigcup_{t \in T_{Rh}} B(t, MRh)$. We therefore set $h_0 = \epsilon_0/R$.

Now suppose x lies in Ω. Then x lies in $B(t, MRh)$, for some t in T_{Rh}. Define $\sigma : B(t, MRh) \longrightarrow B(0, MR)$ by $\sigma(y) = h^{-1}(y - t)$, where $y \in B(t, MRh)$. Each ball $B(x_r, 1)$ must contain at least one image under σ of a point in $\mathcal{A}$. Hence, we can select $a_1, \ldots, a_\ell$ in $B(t, Rh)$ such that $\sigma(a_i) \in B(x_r, 1)$, $r = 1, \ldots, \ell$. We will use the point evaluations at $a_1, \ldots, a_\ell$ as the keystone in our Hilbert space theory. They will then be used in the inner product and, furthermore, form the basis for assumption (iii). Correspondingly, let $p_1, \ldots, p_\ell$ be the cardinal basis for π_k based on $a_1, \ldots, a_\ell$. Let $L_a : C(B(t, MRh)) \longrightarrow \pi_k$ be the Lagrange interpolation operator associated with $a = \{a_1, \ldots, a_\ell\}$. By **3.3.5**, $\|L_a\|_\beta = h^{-|\beta|}\|L_{\sigma(a)}\|_\beta$. However, by **3.3.4**, there exists a constant K such that $\|L_{\sigma(a)}\|_\beta \leq K$, independent of the particular selection of $a_1, \ldots, a_\ell$. Now apply **3.2.5** for x in $B(t, MRh)$. Then, setting $C = MR$,

$$
\begin{aligned}
&|(D^\beta f)(x) - (D^\beta u)(x)|^2 \leq \langle f, f\rangle \\
&\qquad \left|(-1)^{|\beta|}(D^{2\beta}\phi)(0) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(D^\beta\phi)(x - a_r) + \sum_{r,s=1}^{\ell}\phi(a_s - a_r)(D^\beta p_r)(D^\beta p_s)(x)\right| \\
&\leq \;\; \langle f, f\rangle \left\{|(D^{2\beta}\phi)(0)| + 2\max_{0 \leq |y| \leq Ch}\{|(D^\beta\phi)(y)|\}\sum_{r=1}^{\ell}|(D^\beta p_r)(x)|\right.
\end{aligned}
$$

$$
\begin{aligned}
& \qquad + \max_{0 \le |y| \le Ch} \{|\phi(y)|\} \left\{ \sum_{r=1}^{\ell} |(D^\beta p_r)(x)| \right\}^2 \Bigg\} \\
\le \ & \langle f, f \rangle \left\{ |(D^{2\beta}\phi)(0)| + 2Kh^{-|\beta|} \max_{0 \le |y| \le Ch} \{|(D^\beta \phi)(y)|\} \right. \\
& \qquad \left. + K^2 h^{-2|\beta|} \max_{0 \le |y| \le Ch} \{|\phi(y)|\} \right\}. \qquad \blacksquare
\end{aligned}
$$

The strength of this error estimate thus depends on the local behaviour of the basis function ϕ — an observation made in the work of Madych and Nelson [21]. We will return to these considerations when we discuss applications in Chapter 5.

Chapter 4

Spaces of Distributions

This chapter introduces several spaces of tempered distributions, with a view to applying the theory of previous chapters. We begin by considering questions of completeness and density of certain subspaces — this then provides us with the framework on which we can build the necessary Hilbert space structure. Once the structure is in place, we turn to establishing the validity of the assumptions required for the error analysis laid out in the previous chapter, and an examination of the resulting basis functions.

4.1 Weighted spaces of tempered distributions

Some aspects of notation and distribution theory should be remarked upon before we begin, so as to give an idea of the nature of the spaces we will be dealing with. The letter $\mathscr{D}$ will denote the Schwartz space of infinitely differentiable, compactly supported test functions, whose topological dual is the space of Schwartz distributions, $\mathscr{D}'$. Similarly, $\mathscr{S}$ will be the space of infinitely differentiable, rapidly decreasing functions whose dual is the space of tempered distributions, $\mathscr{S}'$. In the case of the compactly supported test functions, the topology is derived from the Fréchet space topology induced by the separating family of semi-norms,

$$|\phi|_N = \max\{|(D^\alpha \phi)(x)| : x \in \mathbb{R}^n, |\alpha| \leq N\}, \qquad \phi \in \mathscr{D}.$$

The topology on $\mathscr{S}$ is defined in a similar manner using the norms

$$\|\psi\|_N = \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^n} (1 + |x|^2)^N |(D^\alpha \psi)(x)|, \qquad \psi \in \mathscr{S}. \tag{4.1}$$

More details of the topology in these spaces of distributions can be found in the two volumes by Schwartz [33, 34], or the concise presentation of Hörmander [18], but we take the above definitions from Rudin [27]. However, it is important to note that the tempered distributions are those distributions which have a continuous extension to $\mathscr{S}$ (c.f. Rudin [27] **7.11**). Throughout the remainder of this work, the action of a distribution f on a test function ψ will be denoted by $[f, \psi]$.

It will be convenient to have some notion to describe when a function is bounded above by a finite degree polynomial for large values of its argument. We will therefore say that f has polynomial growth at infinity, if for some polynomial P, $|f(x)| \leq |P(x)|$, as $|x| \to \infty$.

When we indicate that a distribution is locally integrable, we are not implying that the continuous linear functional is in some sense 'measurable', and 'integrable' over compact sets; rather, we are using a locally integrable function to define the distribution, and making the conventional identification between functional and function. For example, given a locally integrable function f, we can define a unique distribution via the mapping given by $\psi \mapsto \int_{\mathbb{R}^n} f(y)\psi(y)dy$, $\psi \in \mathscr{D}$. It is then convention which allows us to label the resulting distribution f.

On the other hand, if we wish to assert that a distribution, Λ, is locally integrable, we must prove the existence of a locally integrable function g such that, for all ψ in $\mathscr{D}$, $[\Lambda, \psi] = \int_{\mathbb{R}^n} g(y)\psi(y)dy$. As above, we may identify g with the distribution $\psi \mapsto \int_{\mathbb{R}^n} g(y)\psi(y)dy$. It is important to stress, though, that $\Lambda : \mathscr{S} \longrightarrow \mathbb{C}$ and $g : \mathbb{R}^n \longrightarrow \mathbb{C}$.

They are not the same object, but we will, nevertheless, write $\Lambda = g$, remarking always that this is in the above distributional sense. The justification for this lies in the choice of test functions — the integrals $\int g\psi$, $\psi \in \mathscr{D}$ determine g almost everywhere and hence, uniquely define a distribution.

Why then, do we choose spaces of distributions over spaces of functions? The answer to this lies in another part of Schwartz's theory which, in turn, was motivated by the theory of partial differential equations — Fourier analysis. The extension by Schwartz, of Fourier transforms to tempered distributions allows us to manipulate large spaces of objects in creative ways, using concise and well-defined operations. Fourier transforms thus play a large role in what follows and so, notationally, we will let $\widehat{f}$ be the Fourier transform of f, and let $\widetilde{f}$ be the reflection of f defined by $\widetilde{f}(x) = f(-x)$. In distributional terms, if $f \in \mathscr{D}'$, then $\widetilde{f}$ is defined, for all ψ in $\mathscr{D}$, by $[\widetilde{f}, \psi] = [f, \widetilde{\psi}]$. Finally, the functions $v_\beta : \mathbb{R}^n \longrightarrow \mathbb{C}$ will be defined for any multi-index β, by $v_\beta(y) = (iy)^\beta$, $y \in \mathbb{R}^n$.

4.1.1 Definition *A weight function is any extended real-valued, measurable function on $\mathbb{R}^n$, which, in the complement of the origin, is continuous and positive.*

4.1.2 Definition *Let w be a weight function. Then Y^w will denote the space defined by*

$$\{f \in \mathscr{S}' : \widehat{f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n) \text{ and } \|f\|_w < \infty\},$$

where

$$\|f\|_w^2 = \int_{\mathbb{R}^n} |\widehat{f}(y)|^2 w(y)\, dy. \tag{4.2}$$

It should be noted that, given a weight function w, the reciprocal $1/w$ is also a valid weight function. If we assume a little more, namely that $1/w$ is locally integrable, then Y^w can be re-written in a useful way.

4.1.3 Lemma *Suppose w is a weight function which additionally satisfies the condition, $1/w \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Then*

$$Y^w = \{f \in \mathscr{S}' : \widehat{f} \text{ is measurable and } \|f\|_w < \infty\}.$$

Proof. Clearly, $Y^w \subset \{f \in \mathscr{S}' : \widehat{f} \text{ is measurable and } \|f\|_w < \infty\}$. Take any f from $\mathscr{S}'$ for which $\widehat{f}$ is measurable and $\|f\|_w < \infty$. Letting K be any compact subset of $\mathbb{R}^n$,

$$\int_K |\widehat{f}(y)|\, dy \le \left\{ \int_K |\widehat{f}(y)|^2 w(y)\, dy \right\}^{\frac{1}{2}} \left\{ \int_K \frac{1}{w(y)}\, dy \right\}^{\frac{1}{2}},$$

since $1/w \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Hence, $\widehat{f}$ is locally integrable and so, f lies in Y^w. ■

Suppose f and g are chosen from Y^w so that $\|f-g\|_w = 0$. Then, since $\widehat{f}$ and $\widehat{g}$ are locally integrable, and $w > 0$ everywhere except possibly the origin, the integral in (4.2) tells us that $\widehat{f} - \widehat{g}$ is defined by a function which is zero almost everywhere. Hence, $\widehat{f} = \widehat{g}$ distributionally. Since the Fourier transform is an isomorphism of $\mathscr{S}'$ onto $\mathscr{S}'$ (c.f. Rudin [27] **7.4**), it follows that $f = g$. The positivity of w, and linearity of the integral (4.2) under scalar multiplication then confirm that $\|\cdot\|_w$ is a norm on Y^w. We might then ask whether Y^w endowed with the norm $\|\cdot\|_w$ is a Banach space. Answers to this question are found in the relationship between Y^w and $L^2(\mathbb{R}^n)$.

4.1.4 Lemma *Let w be a weight function. Then Y^w is isometrically isomorphic to a subspace of $L^2(\mathbb{R}^n)$. If, in addition, $1/w$ is locally integrable and has polynomial growth at infinity, then Y^w is isometrically isomorphic to $L^2(\mathbb{R}^n)$.*

Proof. Given any f from Y^w, let I be defined by $If = \sqrt{w}\widehat{f}$. It follows that $\|If\|_2^2 = \int_{\mathbb{R}^n} w(y)|\widehat{f}(y)|^2\, dy$, and so, I is the required isometry from Y^w to $L^2(\mathbb{R}^n)$. To see that I is an isomorphism, suppose f and g are chosen from Y^w so that $If = Ig$. Then $\|I(f-g)\|_2 = \|f-g\|_w = 0$, and so $f = g$.

Now assume $1/w$ belongs to $L^1_{\text{loc}}(\mathbb{R}^n)$ and has polynomial growth at infinity. Take any f from $L^2(\mathbb{R}^n)$. Letting K represent any compact set in $\mathbb{R}^n$, we then note that $w^{-\frac{1}{2}}f$ is measurable and

$$\int_K \left|\frac{f(y)}{\sqrt{w(y)}}\right| dy \leq \left\{ \int_K \frac{1}{w(y)}\, dy \right\}^{\frac{1}{2}} \left\{ \int_K |\widehat{f}(y)|^2\, dy \right\}^{\frac{1}{2}}.$$

The local integrability of $1/w$ then tells us that $w^{-\frac{1}{2}}f$ is locally integrable. Moreover, since f is square-integrable, and $1/w$ has polynomial growth at infinity, we remark that, for all ψ in $\mathscr{S}$, we can find a polynomial P such that

$$\left| \int_{\mathbb{R}^n} w^{-\frac{1}{2}}(y) f(y) \psi(y) dy \right| \leq \left| \int_{\mathbb{R}^n} P(y) f(y) \psi(y) dy \right| \leq \|P\psi\|_2 \|f\|_2. \tag{4.3}$$

We now utilise several facts from distribution theory. Firstly, tempered distributions may be constructed from locally integrable functions (with appropriate growth restrictions at infinity) by integrating against the test function. This ensures that they are extensions of the corresponding regular distributions to $\mathscr{S}$. When the density of $\mathscr{D}$ in $\mathscr{S}$ is taken into account, this condition becomes absolute, since to determine whether two tempered distributions are the same, it is enough to test their respective actions on members of $\mathscr{D}$. Secondly, the continuity of a linear functional on $\mathscr{S}$ may be deduced through sequential continuity. Here, a sequence of test functions $\{\psi_j\}$ are taken which converge to zero in the topology on $\mathscr{S}$. If, for a linear functional Λ, the sequence $|[\Lambda, \psi_j]|$ tends to zero as j tends to infinity, then Λ is tempered.

With this in mind, take any sequence $\{\psi_j\}$ in $\mathscr{S}$ which converges to zero in the topology on $\mathscr{S}$ as $j \to \infty$. Examining the norms on $\mathscr{S}$ mentioned earlier in (4.1), we see that $P\psi_j$ will pointwise tend to zero as $j \to \infty$. Therefore we can apply the above inequality (4.3) to the sequence $\{\psi_j\}$ and conclude that $w^{-\frac{1}{2}}f$ is tempered. Consequently, the mapping $J : L^2(\mathbb{R}^n) \longrightarrow Y^w$, given by $Jf = (w^{-\frac{1}{2}}f)^{\widehat{}}$, is well defined. Moreover,

$$\widehat{Jf} = (w^{-\frac{1}{2}}f)^{\widehat{\widehat{}}} = (w^{-\frac{1}{2}}f)^{\widetilde{}} = (w^{-\frac{1}{2}}f).$$

Using that fact that $w^{-\frac{1}{2}}f$ is measurable, we can write

$$\|Jf\|_w^2 = \int_{\mathbb{R}^n} |(\widehat{Jf})(y)|^2 w(y)\, dy = \int_{\mathbb{R}^n} \left|\frac{\widehat{f}(y)}{\sqrt{w(y)}}\right|^2 w(y)\, dy = \|f\|_2^2.$$

Therefore, J is an isometric isomorphism of $L^2(\mathbb{R}^n)$ onto Y^w. ∎

4.1.5 Corollary *Let w be a weight function. Then the space Y^w is an inner product space. If, in addition, $1/w$ is locally integrable and has polynomial growth at infinity, then Y^w is a Hilbert space. In either case, the inner product is given, for all f and g in Y^w, by*

$$(f, g) = \int_{\mathbb{R}^n} \widehat{f}(y)\bar{\widehat{g}}(y) w(y)\, dy.$$

We now detail several results concerning nice subsets of Y^w which can be utilised to enhance our knowledge of Y^w and provide us with some structure on which later work may capitalise.

4.1.6 Lemma *Let w be a weight function. Then the set $\{\sqrt{w}f : f \in C_0(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$.*

Proof. Recall that $C_0(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$ (c.f. Adams [2] **2.13**). Therefore, given f from $L^2(\mathbb{R}^n)$ and $\epsilon > 0$, there exists a function g in $C_0(\mathbb{R}^n)$ for which $\|f - g\|_2 < \epsilon/\sqrt{3}$. Choose ρ from $C_0^\infty(\mathbb{R}^n)$ so that $\rho(x) = 1$ if $|x| \leq 1$, $\rho(x) = 0$ when $|x| > 2$, and $0 \leq \rho(x) \leq 1$ when $1 < |x| < 2$. Define ρ_h for all x in $\mathbb{R}^n$ and $h > 0$ by $\rho_h(x) = \rho(x/h)$, and set $u_h = (1 - \rho_h)g/\sqrt{w}$. Note that, since w is non-zero in the complement of the origin, $u_h \in C(\mathbb{R}^n \setminus \{0\})$. However, $u_h(x) = 0$ for all x such that $|x| < h$, and $x \neq 0$.

By defining $u_h(0) = 0$, we obtain a function which is continuous everywhere. Setting $B_{2h} = \{x \in \mathbb{R}^n : |x| \leq 2h\}$, we have

$$\begin{aligned}
\|f - \sqrt{w}u_h\|_2^2 &= \|f - (1-\rho_h)g\|_2^2 \\
&= \int_{\mathbb{R}^n \setminus B_{2h}} |f(y) - (1-\rho_h(y))g(y)|^2\,dy + \int_{B_{2h}} |f(y) - (1-\rho_h(y))g(y)|^2\,dy \\
&= \int_{\mathbb{R}^n \setminus B_{2h}} |f(y) - g(y)|^2\,dy + \int_{B_{2h}} |f(y) - (1-\rho_h(y))g(y)|^2\,dy \\
&\leq \|f-g\|_2^2 + \left[\left\{\int_{B_{2h}} |f(y)|^2\,dy\right\}^{\frac{1}{2}} + \left\{\int_{B_{2h}} |(1-\rho_h(y))g(y)|^2\,dy\right\}^{\frac{1}{2}}\right]^2 \\
&\leq \|f-g\|_2^2 + \left[\left\{\int_{B_{2h}} |f(y)|^2\,dy\right\}^{\frac{1}{2}} + \left\{\int_{B_{2h}} |g(y)|^2\,dy\right\}^{\frac{1}{2}}\right]^2.
\end{aligned}$$

Now h can be chosen sufficiently small so that

$$\int_{B_{2h}} |f(y)|^2\,dy < \frac{1}{6}\epsilon^2 \text{ and } \int_{B_{2h}} |g(y)|^2\,dy < \frac{1}{6}\epsilon^2.$$

Consequently, $\|f - \sqrt{w}u_h\|_2 < \epsilon$. ■

An immediate consequence of this result is the following.

4.1.7 Corollary *Let w be a weight function. Then Y^w is isometrically isomorphic to a dense subset of $L^2(\mathbb{R}^n)$.*

Proof. From **4.1.4**, Y^w is isometrically isomorphic to a subset of $L^2(\mathbb{R}^n)$. Let f be chosen from $C_0(\mathbb{R}^n)$ so that $\sqrt{w}f \in L^2(\mathbb{R}^n)$. Then f is tempered and is the Fourier transform of some element in Y^w. Let $I : Y^w \longrightarrow L^2(\mathbb{R}^n)$ be defined, as in **4.1.4**, by $If = \sqrt{w}\widehat{f}$, and let $\mathcal{A} = \{g \in Y^w : \widehat{g} \in C_0(\mathbb{R}^n)\}$. Then

$$\begin{aligned}
I(\mathcal{A}) &= \{\sqrt{w}\widehat{g} : g \in Y^w \text{ and } \widehat{g} \in C_0(\mathbb{R}^n)\} \\
&= \{\sqrt{w}\widehat{g} : \widehat{g} \in C_0(\mathbb{R}^n) \text{ and } \sqrt{w}\widehat{g} \in L^2(\mathbb{R}^n)\} \\
&= \{\sqrt{w}g : g \in C_0(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n).
\end{aligned}$$

The latter is dense in $L^2(\mathbb{R}^n)$ by **4.1.6**. ∎

A useful extension of these ideas can be found in the next lemma.

4.1.8 Lemma *Let w be a weight function. Then the set $\{\sqrt{w}f : f \in C_0^\infty(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$.*

Proof. We will begin by proving the density of $\{\sqrt{w}g : g \in C_0^\infty(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$ in $\{\sqrt{w}g : g \in C_0(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$ with respect to the L^2-norm.

Fix $\epsilon > 0$, and let f be any element from $\{\sqrt{w}g : g \in C_0(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$. Let ρ and ρ_h be as defined in **4.1.6**, and choose h sufficiently small so that $\|\sqrt{w}\rho_h f\|_2 = \|\sqrt{w}f - \sqrt{w}(1-\rho_h)f\|_2 < \epsilon/2$. Next, since $C_0(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$, and $C_0^\infty(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$, we can find a function g in $C_0^\infty(\mathbb{R}^n)$ such that $\operatorname{supp}\{g\}$ lies in the ball of radius R and

$$\sup_{h<|x|<R} \left\{\sqrt{w(x)}|(1-\rho_h(x))|\right\}\|f-g\|_2 < \epsilon/2.$$

Then, $(1-\rho_h)g \in C_0^\infty(\mathbb{R}^n)$ and,

$$\begin{aligned}
\|\sqrt{w}f - \sqrt{w}(1-\rho_h)g\|_2 &\leq \|\sqrt{w}(f-(1-\rho_h)f)\|_2 + \|\sqrt{w}(1-\rho_h)(f-g)\|_2 \\
&< \frac{1}{2}\epsilon + \left\{\int_{h<|x|<R} w(y)(1-\rho_h(y))^2|f(y)-g(y)|^2\,dy\right\}^{\frac{1}{2}} \\
&< \frac{1}{2}\epsilon + \sup_{h<|x|<R}\left\{\sqrt{w(x)}|1-\rho_h(x)|\right\}\|f-g\|_2 \\
&< \epsilon.
\end{aligned}$$

The density of $\{\sqrt{w}f : f \in C_0(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$ now completes the proof. ∎

4.1.9 Corollary *Let w be a weight function. Then the set $\{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap Y^w$ is dense in Y^w.*

Proof. Let $\mathcal{A} = \{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap Y^w$, and let $I : Y^w \longrightarrow L^2(\mathbb{R}^n)$ be defined by

$If = \sqrt{w}\widehat{f}$. Then, as seen in **4.1.4**, I is an isometry of Y^w into $L^2(\mathbb{R}^n)$. Furthermore,

$$\begin{aligned} I(\mathcal{A}) &= \{\sqrt{w}\widehat{f} : f \in Y^w \text{ and } \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \\ &= \{\sqrt{w}\widehat{f} : \widehat{f} \in C_0^\infty(\mathbb{R}^n) \text{ and } \sqrt{w}\widehat{f} \in L^2(\mathbb{R}^n)\} \\ &= \{\sqrt{w}f : f \in C_0^\infty(\mathbb{R}^n)\} \cap L^2(\mathbb{R}^n). \end{aligned}$$

The latter is dense in $L^2(\mathbb{R}^n)$ by **4.1.8** and so, $\mathcal{A}$ must be dense in Y^w. ■

4.1.10 Corollary *Let w be a weight function. Then the set $\mathscr{S} \cap Y^w$ is dense in Y^w.*

Proof. The result is immediate from the fact that $\{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap Y^w \subset \mathscr{S} \cap Y^w$, and the former is dense in Y^w by **4.1.9**. ■

From the nature of the norm,

$$\|f\|_w = \left\{ \int_{\mathbb{R}^n} |\widehat{f}(y)|^2 w(y)\, dy \right\}^{\frac{1}{2}}, \qquad f \in Y^w,$$

it is clear that, if w grows exponentially, then $\mathscr{S} \cap Y^w$ will shrink to $\mathscr{D} \cap Y^w$. On the other hand, a large singularity in w at the origin will reduce the size of $\mathscr{S} \cap Y^w$ regardless of the behaviour of the weight at infinity. It is perhaps of interest, therefore, to ascertain when $\mathscr{S} \subset Y^w$.

4.1.11 Theorem *Let w be a weight function which has polynomial growth at infinity and is, itself, locally integrable. Then $\mathscr{S} \subset Y^w$.*

Proof. By hypothesis, there exist constants C and R such that, for some real number μ and for all $|x| > R$, $w(x) \leq C|x|^\mu$. Take any ψ in $\mathscr{S}$. Then,

$$\int_{\mathbb{R}^n} |\widehat{\psi}(y)|^2 w(y)\, dy = \int_{|x| \leq R} |\widehat{\psi}(y)|^2 w(y)\, dy + \int_{|x| > R} |\widehat{\psi}(y)|^2 w(y)\, dy.$$

Since w is locally integrable, the first integral on the right hand side is finite. The second integral is also finite because $\widehat{\psi}$ lies in $\mathscr{S}$ and therefore has sufficiently rapid decay at infinity to counter the polynomial growth of w. Hence ψ belongs to Y^w. ■

With this information, we now introduce a new space which is closely related to Y^w, and acts as a useful prelude to the space in which we will be predominantly interested.

4.1.12 Definition *Let α be a non-negative multi-index, and let w be a weight function. Then Y_α^w will denote the space defined by*

$$\{f \in \mathscr{S}' : \widehat{D^\alpha f} \in L^1_{\text{loc}}(\mathbb{R}^n) \text{ and } |f|_{\alpha,w} < \infty\},$$

where the semi-norm $|\cdot|_{\alpha,w}$ is given by

$$|f|^2_{\alpha,w} = \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 w(y)\, dy, \qquad f \in Y_\alpha^w.$$

4.1.13 Theorem *Let α be any multi-index and let w be a weight function. Then the operator D^α maps Y_α^w onto Y^w, and for all f in Y_α^w, $\|D^\alpha f\|_w = |f|_{\alpha,w}$.*

Proof. Choose any f from Y_α^w. Since f is tempered, $D^\alpha f$ is also tempered. Furthermore, $\widehat{D^\alpha f}$ is locally integrable and $\|D^\alpha f\|_w^2 = \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 w(y)\, dy = |f|^2_{\alpha,w}$. Hence D^α maps Y_α^w into Y^w. To see that D^α maps Y_α^w onto Y^w, we let g be any element of Y^w and remark that the work of Hörmander [19] admits a tempered solution of the distributional partial differential equation, $D^\alpha f = g$. We assert that all such solutions lie in Y_α^w.

Since $\widehat{D^\alpha f} = \widehat{g}$, and $\widehat{g}$ lies in $L^1_{\text{loc}}(\mathbb{R}^n)$ by definition, we have the required integrability condition on f. Observing, once again, that $|f|_{\alpha,w} = \|D^\alpha f\|_w = \|g\|_w$ completes the proof. ■

When $|\alpha| > 0$, it should be noted that $|\cdot|_{\alpha,w}$ is a semi-norm. It is desirable to know more about the completeness of Y_α^w with respect to this semi-norm, but this conventionally involves factoring out the kernel, and then showing completeness of the resulting normed linear space. We, however, adopt an alternative route. If $(U, |\cdot|)$ is a linear space equipped with a semi-norm, we will refer to U as being complete if to each Cauchy sequence $\{u_j\} \subset U$, there corresponds an element u in U such that $|u_j - u| \to 0$ as $j \to \infty$. With this definition, though, u is not uniquely defined by the sequence $\{u_j\}$.

4.1.14 Corollary *Let α be any multi-index, and let w be a weight function whose reciprocal, $1/w$, is locally integrable and has polynomial growth at infinity. Then Y_α^w is complete.*

Proof. The conditions on w ensure that Y^w is complete, as shown in **4.1.5**. From **4.1.13**, the fact that D^α is an isometry from Y_α^w to Y^w means that, if $\{f_j\}$ is a Cauchy sequence in Y_α^w, then $\{D^\alpha f_j\}$ is a Cauchy sequence in Y^w. The completeness of Y^w then implies that we can find an element g in Y^w such that $D^\alpha f_j \to g$ in the Y^w topology as $j \to \infty$. However, the fact that $D^\alpha : Y_\alpha^w \longrightarrow Y^w$ is surjective shows that there is an element f in Y_α^w such that $D^\alpha f = g$. Then the observation that $|f_j - f|_{\alpha,w} = \|D^\alpha f_j - g\|_w \to 0$ completes the proof. ■

We are now at the point where we can bring all the previous ideas together and define a space of distributions which will become the centre of attention for the remainder of our work.

4.1.15 Definition *Let w be a weight function, and let k be a non-negative integer. Then the space $X^{k,w}$ will be defined as the intersection of all the spaces Y_α^w, for which $|\alpha| = k$. One possible semi-norm on $X^{k,w}$ is given by*

$$|f|_{k,w}^2 = \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 w(y)\, dy, \tag{4.4}$$

where the numbers c_α are binomial coefficients from the identity, $|x|^{2k} = \sum_{|\alpha|=k} c_\alpha x^{2\alpha}$, $x \in \mathbb{R}^n$.

It should perhaps be noted that if w has spherical symmetry, then the choice of coefficients $\{c_\alpha\}$ makes the semi-norm rotation invariant. Another point of interest is that the semi-norm may be re-written as

$$\begin{aligned} |f|^2_{k,w} &= \int_{\mathbb{R}^n} \sum_{|\alpha|=k} c_\alpha |v_\alpha|^2 |\widehat{f}(y)|^2 w(y)\, dy \\ &= \int_{\mathbb{R}^n} w(y)|y|^{2k}|\widehat{f}(y)|^2\, dy. \end{aligned} \tag{4.5}$$

However, this reformulation is inherently misleading without the other restrictions imposed on $X^{k,w}$, namely that for all α satisfying $|\alpha| = k$, $\widehat{D^\alpha f}$ must be locally integrable, in the sense described at the beginning of the chapter.

For example, let p be a polynomial from π_{k-1}. Then p is tempered and for all $|\alpha| = k$, $D^\alpha p = 0$. Hence $|p|_{k,w}$ is easily seen to be zero from (4.4). However, the alternative form of the semi-norm (4.5) appears to be 'integrating' delta distributions and their derivatives, which, by definition, are not 'locally integrable', or even 'measurable'. The advantages of the second form then stem mainly from the case when $\widehat{f}$ is locally integrable. We will therefore endeavour to avoid any such misconceptions in future work, by paying careful attention to the afore-mentioned restrictions.

4.1.16 Theorem *Let k be a non-negative integer, and let w be a weight function whose reciprocal, $1/w$, is locally integrable and has polynomial growth at infinity. Then the semi-norm on $X^{k,w}$ has kernel π_{k-1}, and with respect to this semi-norm, $X^{k,w}$ is complete.*

Proof. If f is chosen from $X^{k,w}$ so that $|f|_{k,w} = 0$, then for all multi-indices α satisfying $|\alpha| = k$, $D^\alpha f \in Y^w$ and $\|D^\alpha f\|_w = 0$. Since Y^w is a normed linear space, we conclude that $D^\alpha f = 0$ for all $\alpha \in \mathbb{Z}^n_+$ satisfying $|\alpha| = k$, and hence, f lies in π_{k-1}. With the extra conditions on the weight function, **4.1.14** shows that each Y_α^w is complete and so, $X^{k,w}$, being the intersection of finitely many complete spaces, is also complete. ∎

We conclude this section with three useful results, each revealing a little more of the nature of the spaces $X^{k,w}$.

4.1.17 Lemma *Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ within a neighbourhood, N, of the origin, and let ρ_h be defined for all x in $\mathbb{R}^n$ and $h > 0$, by $\rho_h(x) = \rho(x/h)$. Let f be any tempered distribution, for which it is known that, for some multi-index α, $\widehat{D^\alpha f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Then, for all $h > 0$, $\widehat{f}(1 - \rho_h)$ is locally integrable.*

Proof. Suppose we are given an element f from $\mathscr{S}'$ and a multi-index α for which $\widehat{D^\alpha f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Let k denote the order of α. Since $\widehat{D^\alpha f} = v_\alpha \widehat{f}$, and multiplication by $\overline{v}_\alpha$ preserves the local integrability, we know that $|v_\alpha|^2 \widehat{f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Choosing coefficients $\{c_\alpha\}$ such that

$$\sum_{|\alpha|=k} c_\alpha |v_\alpha(y)|^2 = \sum_{|\alpha|=k} c_\alpha y^{2\alpha} = |y|^{2k}, \qquad y \in \mathbb{R}^n,$$

we then see that $|\cdot|^{2k} \widehat{f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Let η_h be chosen from $C^\infty(\mathbb{R}^n)$ such that, for all $h > 0$, $\eta_h > 0$, and for all y in $\mathbb{R}^n \setminus hN$, $\eta_h(y) = |y|^{2k}$. Then, for all $h > 0$, the set $\{\eta_h \phi : \phi \in \mathscr{D}\}$ is, itself, $\mathscr{D}$. Subsequently, for any ϕ in $\mathscr{D}$, there exists a corresponding function ψ in $\mathscr{D}$ such that

$$\begin{aligned}
[\widehat{f}(1 - \rho_h), \phi] &= [\widehat{f}(1 - \rho_h), \eta_h \psi] \\
&= [\widehat{f}(1 - \rho_h)\eta_h, \psi] \\
&= [\widehat{f}(1 - \rho_h)|\cdot|^{2k}, \psi] \\
&= [|\cdot|^{2k} \widehat{f}, (1 - \rho_h)\psi].
\end{aligned}$$

Since $|\cdot|^{2k}\widehat{f}$ is locally integrable, there exists a locally integrable function F such that

$$\begin{aligned}[|\cdot|^{2k}\widehat{f},(1-\rho_h)\psi] &= \int_{\mathbb{R}^n} F(y)(1-\rho_h(y))\psi(y)\,dy \\ &= \int_{\mathbb{R}^n} F(y)\frac{(1-\rho_h(y))}{\eta_h(y)}\phi(y)\,dy.\end{aligned}$$

Now, $(1-\rho_h)/\eta_h$ is infinitely differentiable and so, the function $G = F(1-\rho_h)/\eta_h$ is also locally integrable. Bringing the two parts together, we have

$$[\widehat{f}(1-\rho_h),\phi] = \int_{\mathbb{R}^n} G(y)\phi(y)\,dy,$$

and the proof is complete. ■

One immediate consequence of this result is that, for all f in $X^{k,w}$, $\widehat{f}$ is integrable over any compact set not containing the origin. Another interpretation might be to assert that the Fourier transforms of elements of $X^{k,w}$ are almost 'regular', that is, defined by locally integrable functions. Our next result continues to pursue this idea, and is fundamental to our treatment of interpolation problems. It is essentially analogous to part of the Sobolev embedding theorems, showing us that $X^{k,w}$ is, under certain conditions, a subset of $C(\mathbb{R}^n)$.

4.1.18 Theorem *Let w be any weight function for which there exists a real number μ such that $[w(x)]^{-1} = \mathcal{O}(|x|^{-2\mu})$ as $|x| \to \infty$, and let k and n be integers, chosen so that $k \geq 0$, $n > 0$ and $k+\mu-n/2 > 0$. Then $X^{k,w} \subset C^j(\mathbb{R}^n)$, for any non-negative integer j satisfying $j < k+\mu-n/2$.*

Proof. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ within a neighbourhood, N, of the origin, and $0 \leq \rho \leq 1$, and let f be any element of $X^{k,w}$. As a special case of **4.1.17**, we know that $(1-\rho)\widehat{f}$ is locally integrable. Hence, for any multi-index β, $v_\beta(1-\rho)\widehat{f} \in$

$L^1_{\mathrm{loc}}(\mathbb{R}^n)$. We may then extend this idea by asking whether $v_\beta(1-\rho)\widehat{f}$ lies in $L^1(\mathbb{R}^n)$. To this end, we write

$$\begin{aligned}
&\int_{\mathbb{R}^n} |v_\beta(y)(1-\rho(y))\widehat{f}(y)|\,dy = \int_{\mathbb{R}^n\setminus N} |v_\beta(y)(1-\rho(y))\widehat{f}(y)|\,dy \\
&= \int_{\mathbb{R}^n\setminus N} \left|(v_\alpha \widehat{f})(y)(1-\rho(y))\,\frac{v_{\beta-\alpha}(y)}{\sqrt{w(y)}}\sqrt{w(y)}\right| dy \\
&\leq \left\{\int_{\mathbb{R}^n\setminus N} (1-\rho(y))^2\,|(v_\alpha \widehat{f})(y)|^2\,w(y)\,dy\right\}^{\frac{1}{2}} \left\{\int_{\mathbb{R}^n\setminus N} \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy\right\}^{\frac{1}{2}} \\
&= \left\{\int_{\mathbb{R}^n\setminus N} (1-\rho(y))^2\,|(\widehat{D^\alpha f})(y)|^2\,w(y)\,dy\right\}^{\frac{1}{2}} \left\{\int_{\mathbb{R}^n\setminus N} \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy\right\}^{\frac{1}{2}} \\
&\leq \left\{\int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2\,w(y)\,dy\right\}^{\frac{1}{2}} \left\{\int_{\mathbb{R}^n\setminus N} \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy\right\}^{\frac{1}{2}}.
\end{aligned}$$

By hypothesis, $|f|_{\alpha,w}$ is finite for all $|\alpha| = k$. Furthermore, because of the assumptions on w, we can find constants C and R such that $[w(x)]^{-1} \leq C|x|^{-2\mu}$, for all $|x| > R$. Now, on the compact set $K = \{y \in \mathbb{R}^n \setminus N : |y| \leq R\}$, w is positive and continuous, and so, there exists a positive number δ such that $w(x) \geq 1/\delta$ for all x in K. Consequently, for all x in K, $[w(x)]^{-1} \leq \delta$ and we can write

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus N} \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy &= \int_K \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy + \int_{\mathbb{R}^n\setminus\{N\cup K\}} \frac{|v_{2\beta-2\alpha}(y)|}{w(y)}\,dy \\
&\leq \delta\int_K |v_{2\beta-2\alpha}(y)|\,dy + C\int_{|y|>R} |y|^{-2\mu}|v_{2\beta-2\alpha}(y)|\,dy.
\end{aligned}$$

The first integral on the right hand side is a fixed finite number. For the second integral, we make the change of variables $y = rt$, so that dt becomes the induced Lebesgue measure on the sphere in $\mathbb{R}^n$, denoted by S^{n-1} (c.f. Stein and Weiss [37]). Hence,

$$\begin{aligned}
\int_{|y|>R} |v_{2\beta-2\alpha}(y)||y|^{-2\mu}\,dy &= \int_R^\infty r^{-2\mu}\int_{S^{n-1}} |v_{2\beta-2\alpha}(rt)|\,dt\,r^{n-1}\,dr \\
&= \int_R^\infty r^{-2\mu+2|\beta-\alpha|+n-1}\int_{S^{n-1}} |v_{2\beta-2\alpha}(t)|\,dt\,dr,
\end{aligned}$$

and the latter is finite whenever $-2\mu+2|\beta-\alpha|+n-1 < -1$. Since $2|\beta-\alpha| = 2|\beta|-2|\alpha| = 2|\beta| - 2k$, we conclude that $v_\beta(1-\rho)\widehat{f} \in L^1(\mathbb{R}^n)$ when $|\beta| < k + \mu - n/2$. Writing $F = v_\beta \widehat{f}(1-\rho) = \widehat{D^\beta f}(1-\rho)$, we may then take Fourier transforms once more to obtain

$$\widehat{F} = (D^\beta f)^{\widetilde{}} - (D^\beta f)^{\widetilde{}} * \widehat{\rho}.$$

Since F lies in $L^1(\mathbb{R}^n)$, $\widehat{F}$ is continuous (c.f. Rudin [27], **7.5**). Furthermore, the convolution of a tempered distribution and a test function is an infinitely differentiable function (c.f. Rudin [27], **7.19**). Hence, $(D^\beta f)^{\widetilde{}} = \widehat{F} + (D^\beta f)^{\widetilde{}} * \widehat{\rho}$ is continuous. ∎

Our final result of this section concerns a question of density, and extends an idea first seen in **4.1.9**.

4.1.19 Theorem *Let k be a non-negative integer, and let w be any weight function. Then the space $\{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap X^{k,w}$ is dense in $X^{k,w}$.*

Proof. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ in some neighbourhood of the origin and $0 \le \rho \le 1$ elsewhere. Then, for all x in $\mathbb{R}^n$ and $h > 0$, define ρ_h by $\rho_h(x) = \rho(x/h)$ and ψ_h by $\widehat{\psi}_h = \rho_h$.

Choose any f from $X^{k,w}$. Since, for all $h > 0$, $f * \psi_h$ is tempered, the same is true of $f - f * \psi_h$. Letting α denote any multi-index of order k, we then make the observation that $\{D^\alpha(f - f * \psi_h)\}^{\widehat{}} = \widehat{D^\alpha f}(1-\rho_h)$. Since, by definition, $\widehat{D^\alpha f} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $(1-\rho_h) \in C^\infty(\mathbb{R}^n)$, we see that $\{D^\alpha(f - f * \psi_h)\}^{\widehat{}} \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Finally,

$$\begin{aligned} |f - f * \psi_h|^2_{k,w} &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |\{D^\alpha(f - f * \psi_h)\}^{\widehat{}}(y)|^2 w(y)\, dy \\ &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 (1-\rho_h(y))^2 w(y)\, dy \\ &\le |f|^2_{k,w}, \end{aligned}$$

which confirms that, for $h > 0$, $f - f * \psi_h \in X^{k,w}$. Alternatively, noting that $(f - f * \psi_h)\widehat{} = \widehat{f}(1 - \rho_h)$, and the latter is locally integrable by **4.1.17**, we can use the second form of $|\cdot|_{k,w}$ mentioned earlier (4.5) and write

$$|f - f * \psi_h|_{k,w}^2 = \int_{\mathbb{R}^n} |\widehat{f}(y)(1 - \rho_h(y))|^2 |y|^{2k} w(y)\, dy.$$

In this sense, $f - f * \psi_h$ lies in Y^τ, where τ is the weight function $|\cdot|^{2k} w$. Given any $\epsilon > 0$, we can therefore find, from **4.1.9**, a function ϕ_h in $\{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap Y^\tau$ such that

$$\|f - f * \psi_h - \phi_h\|_\tau < \frac{1}{2}\epsilon.$$

Since, $|\phi_h|_{k,w} = \|\phi_h\|_\tau$, we know that ϕ_h also lies in $\{f \in \mathscr{S} : \widehat{f} \in C_0^\infty(\mathbb{R}^n)\} \cap X^{k,w}$. Therefore,

$$\begin{aligned} |f - \phi_h|_{k,w} &\le |f - f * \psi_h - \phi_h|_{k,w} + |f * \psi_h|_{k,w} \\ &= \|f - f * \psi_h - \phi_h\|_\tau + |f * \psi_h|_{k,w} \\ &< \frac{1}{2}\epsilon + |f * \psi_h|_{k,w}. \end{aligned}$$

Now, letting Ω denote the support of ρ, we note that $\operatorname{supp}\{\rho_h\} = h\Omega$ and so,

$$\begin{aligned} |f * \psi_h|_{k,w}^2 &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)\rho_h(y)|^2 w(y)\, dy \\ &= \sum_{|\alpha|=k} c_\alpha \int_{h\Omega} |(\widehat{D^\alpha f})(y)\rho_h(y)|^2 w(y)\, dy \\ &\le \sum_{|\alpha|=k} c_\alpha \int_{h\Omega} |(\widehat{D^\alpha f})(y)|^2 w(y)\, dy. \end{aligned}$$

Hence, there exists a value $h_0 > 0$ such that, for all $h < h_0$, $|f * \psi_h|_{k,w} < \epsilon/2$. For such values of h, $|f - \phi_h|_{k,w} < \epsilon$, as required. ■

Our final remark of this section will be to say that, as in **4.1.10**, it is a straightforward result to deduce from the previous theorem that $\mathscr{S} \cap X^{k,w}$ is dense in $X^{k,w}$.

4.2 Reproducing kernel Hilbert spaces

Where has our analysis led us? Defining a quadratic bilinear form $\langle\cdot,\cdot\rangle$ on $X^{k,w}$ by

$$\langle f,g\rangle = \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} (\widehat{D^\alpha f})(y)(\overline{\widehat{D^\alpha g}})(y)w(y)\,dy, \qquad f,g \in X^{k,w},$$

we quickly see that $\langle\cdot,\cdot\rangle = |\cdot|^2_{k,w}$ and hence $(X^{k,w}, \langle\cdot,\cdot\rangle)$ is a semi-inner product space. Ideally though, we would like to have in our possession, a Hilbert space in which the interpolation problems discussed in earlier chapters, are well posed.

Construction of a Hilbert space from $(X^{k,w}, \langle\cdot,\cdot\rangle)$ is a straightforward consequence of our previous work, as we shall see, but the interpolation problems, and subsequent error analysis, require a proof of validity for each of the assumptions outlined at the beginning of Chapter 3. Some of these have been shown already, but the boundedness of point evaluations in the Hilbert space norm, and construction of their representers still requires quite some machinery, and this forms the bulk of this section.

We begin, however, by revising our definition of a weight function, so as to take advantage of the main ideas in the previous section.

4.2.1 Definition *A weight function will be an extended real-valued function w which satisfies the following properties.*

(i) $w \in C(\mathbb{R}^n \setminus \{0\})$,

(ii) $w(x) > 0$, *whenever* $x \neq 0$,

(iii) $1/w \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, *and*

(iv) there exists a real number μ such that $[w(x)]^{-1} = \mathcal{O}(|x|^{-2\mu})$, as $|x| \to \infty$.

When the dependence on the value of μ in (iv) is important, we will denote the weight function by w_μ.

4.2.2 Theorem *Let k be any non-negative integer, and let w be a weight function. Let the dimension of π_{k-1} be denoted by ℓ, and suppose $\{a_1, \ldots, a_\ell\}$ is a set of points in $\mathbb{R}^n$ which is unisolvent with respect to π_{k-1}. Then the bilinear form defined on $X^{k,w}$ by*

$$(f, g) = \langle f, g \rangle + \sum_{r=1}^{\ell} f(a_r)\overline{g(a_r)}, \qquad f, g \in X^{k,w},$$

is an inner product with respect to which, $X^{k,w}$ is a Hilbert space.

Proof. From **4.1.18**, we know the point evaluations in the bilinear form make sense, and are finite. The Hilbert space theory of Chapter 2 then confirms that $(\cdot,\cdot)$ is a genuine inner product on $X^{k,w}$, and so, it only remains to show that $X^{k,w}$ is complete with respect to the norm $\|\cdot\|_{k,w}$ induced by the inner product.

Let $\{f_j\}$ be a Cauchy sequence with respect to $\|\cdot\|_{k,w}$. Then, given any $\epsilon > 0$, we can find a threshold N such that, whenever $s, t \geq N$,

$$\|f_s - f_t\|_{k,w}^2 = |f_s - f_t|_{k,w}^2 + \sum_{r=1}^{\ell} |f_s(a_r) - f_t(a_r)|^2 < \epsilon. \tag{4.6}$$

Therefore, $\{f_j\}$ is a Cauchy sequence with respect to $|\cdot|_{k,w}$ and by **4.1.16**, has a limit f which lies in $X^{k,w}$, but is not uniquely defined by $\{f_j\}$ — we can add any polynomial q from π_{k-1} to f without changing the nature of the convergence.

With this in mind, we also notice from (4.6) that for each a_r, $r = 1, \ldots, \ell$, $|f_s(a_r) - f_t(a_r)| < \epsilon$. Hence, for each $r = 1, \ldots, \ell$, $\{f_j(a_r)\}$ is a Cauchy sequence of complex numbers, which converges to some finite value d_r. Since $a_1, \ldots, a_\ell$ are unisolvent with respect to π_{k-1}, we may now choose q from π_{k-1} such that $q(a_r) = d_r + f(a_r)$, $r = 1, \ldots, \ell$.

Then,

$$\begin{aligned}
\|f_j - (f+q)\|_{k,w}^2 &= |f_j - f|_{k,w}^2 + \sum_{r=1}^{\ell} |f_j(a_r) - f(a_r) - q(a_r)|^2 \\
&= |f_j - f|_{k,w}^2 + \sum_{r=1}^{\ell} |f_j(a_r) - d_r|^2.
\end{aligned}$$

The right hand side then tends to zero as $j \to \infty$ and so, observing that $f+q$ lies in $X^{k,w}$ completes the proof. ■

4.2.3 Remark Throughout this section, the letter ℓ will represent the dimension of π_{k-1}, $\{a_1, \ldots, a_\ell\}$ will denote a π_{k-1} unisolvent set of points in $\mathbb{R}^n$, and $p_1, \ldots, p_\ell$ will be the corresponding cardinal basis for π_{k-1}. With respect to these definitions, P will denote the mapping from $C(\mathbb{R}^n)$ to π_{k-1} defined by $Pf = \sum_{r=1}^{\ell} f(a_r)p_r$. Finally, we will assume that $a_1, \ldots, a_\ell$ are the points used in the inner product, and so, by **2.3.2**, p_r is the representer for the point evaluation at a_r, $r = 1, \ldots, \ell$. Assumption (iii) for the error analysis of Chapter 3 is now satisfied.

Our next aim is the boundedness of the point evaluations, but this requires some diverse results.

4.2.4 Lemma *Let w_μ be a weight function. Then $1/w_\mu$ is a tempered distribution.*

Proof. Let N be the neighbourhood of the origin, outside of which, for some constant C, the relation $[w_\mu(y)]^{-1} \le C|y|^{-2\mu}$ holds. Next, choose a positive integer λ such that $-2\lambda - 2\mu < -n$. Then, for all ψ in $\mathscr{S}$, we can write

$$\begin{aligned}
\left| \int_{\mathbb{R}^n} \frac{1}{w_\mu(y)} \psi(y)\, dy \right| &\le C \int_{\mathbb{R}^n \setminus N} |y|^{-2\mu} |\psi(y)|\, dy + \int_N \frac{1}{w_\mu(y)} |\psi(y)|\, dy \\
&\le C \sup_{y \in \mathbb{R}^n} \{|y|^{2\lambda} |\psi(y)|\} \int_{\mathbb{R}^n \setminus N} |y|^{-2\lambda - 2\mu}\, dy + \|\psi\|_\infty \int_N \frac{1}{w_\mu(y)} dy.
\end{aligned}$$

Both of the integrals on the right hand side are finite because of the conditions on w_μ and λ. In a similar manner to **4.1.4**, we now take a sequence $\{\psi_j\}$ from $\mathscr{S}$ which converges to zero in the topology on $\mathscr{S}$. Then, the above inequality shows that the sequence $|[1/w_\mu, \psi_j]|$ tends to zero. Hence, we can deduce that $1/w_\mu$ is a tempered distribution (c.f. Stein and Weiss [37]). ■

Let k be a non-negative integer. Then in conjunction with **4.2.4**, the work of Hörmander, in [19], verifies the existence of tempered distributions Λ which satisfy $|\cdot|^{2k}\widehat{\Lambda} = 1/w_\mu$. These solutions will play a central role in the interpolation problem, so we will spend some time describing them.

4.2.5 Lemma *Let w_μ be a weight function, and let k and n be integers, chosen so that $k \geq 0$, $n > 0$, and $2k + 2\mu - n > 0$. Subsequently, let j be the largest integer less than $2k+2\mu-n$. Then all tempered distributions Λ satisfying $|\cdot|^{2k}\widehat{\Lambda} = 1/w_\mu$, belong to $C^j(\mathbb{R}^n)$.*

Proof. Let N be the neighbourhood of the origin, outside which, for some constant C, the relation $[w_\mu(y)]^{-1} \leq C|y|^{-2\mu}$ holds. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ such that $\rho = 1$ on N and $0 \leq \rho \leq 1$ elsewhere. Then, since $(1-\rho)$ is zero in a neighbourhood of the origin, $(1-\rho)/|\cdot|^{2k}$ is continuous. Furthermore, since $|\cdot|^{2k}\widehat{\Lambda}$ is locally integrable, we can then quite reasonably form the product

$$\frac{(1-\rho)}{|\cdot|^{2k}}|\cdot|^{2k}\widehat{\Lambda} = (1-\rho)\widehat{\Lambda} = \frac{(1-\rho)}{|\cdot|^{2k}w_\mu}.$$

Let F be the function $(1-\rho)/(|\cdot|^{2k}w_\mu)$. Then F is measurable, and for all multi-indices β,

$$\begin{aligned}\int_{\mathbb{R}^n} |v_\beta(y)F(y)|\,dy &= \int_{\mathbb{R}^n}\left|\frac{v_\beta(y)(1-\rho(y))}{|y|^{2k}w_\mu(y)}\right| dy \\ &= \int_{\mathbb{R}^n\setminus N}\left|\frac{v_\beta(y)(1-\rho(y))}{|y|^{2k}w_\mu(y)}\right| dy\end{aligned}$$

$$\begin{aligned} &\leq C \int_{\mathbb{R}^n \setminus N} |(1-\rho(y))||y|^{-2k-2\mu+|\beta|}\, dy \\ &\leq C \int_{\mathbb{R}^n \setminus N} |y|^{-2k-2\mu+|\beta|}\, dy. \end{aligned}$$

The integral on the right is finite whenever $-2k - 2\mu + |\beta| + n - 1 < -1$, or $0 \leq |\beta| < 2k + 2\mu - n$, and for these values, therefore, $v_\beta F$ is an integrable function. Moreover, using the fact that products of distributions with infinitely differentiable functions are well defined (c.f. Rudin [27] **6.15**),

$$\widehat{D^\beta \Lambda}(1-\rho) = v_\beta (1-\rho)\widehat{\Lambda} = v_\beta F,$$

in the distributional sense. Taking Fourier transforms of both sides, we have

$$(D^\beta \Lambda)\tilde{} - (D^\beta \Lambda)\tilde{} * \widehat{\rho} = (-1)^{|\beta|} D^\beta \widehat{F},$$

and, re-arranging,

$$(-1)^{|\beta|} D^\beta \widetilde{\Lambda} - (-1)^{|\beta|} (D^\beta \widetilde{\Lambda}) * \widehat{\rho} = (-1)^{|\beta|} D^\beta \widehat{F},$$

or

$$D^\beta \widetilde{\Lambda} = D^\beta \widehat{F} + D^\beta (\widetilde{\Lambda} * \widehat{\rho}).$$

Now, the first term on the right is the Fourier transform of an integrable function, which is known to be continuous and to vanish at infinity. The second term, however, contains the convolution of a tempered distribution and a test function. As such, it is infinitely differentiable and has polynomial growth at infinity. Therefore, no matter which solution Λ we select, $\Lambda * \widehat{\widehat{\rho}} + \widehat{\widehat{F}}$ lies in $C^j(\mathbb{R}^n)$. Consequently, Λ lies in $C^j(\mathbb{R}^n)$. ∎

4.2.6 Lemma *Let E be the real-valued function, defined on $\mathbb{R}^n$ by $E(y) = |y|^{2\lambda}$, $y =$*

$(y_1, \ldots, y_n)$*, for some real* λ*. Then*

$$\frac{\partial^\kappa E}{\partial y_j^\kappa} = \sum_{r=\sigma}^{\kappa} b_r |y|^{2\lambda-2r} y_j^{2r-\kappa},$$

where σ *is the largest integer less than* $(\kappa+1)/2$*, and the coefficients* $\{b_r\}$ *depend only on* λ *and* κ*.*

Proof. The proof is by induction, stepping through intervals of two derivatives. Suppose that the formula holds for $\kappa = d$, $d \geq 1$, and consider the case when $\kappa = d+2$. Letting σ denote the largest integer less than $(d+1)/2$, we have

$$\begin{aligned}
\frac{\partial^{d+2}E}{\partial y_j^{d+2}} &= \frac{\partial^2}{\partial y_j^2}\left\{\frac{\partial^d E}{\partial y_j^d}\right\} \\
&= \frac{\partial}{\partial y_j}\left\{\sum_{r=\sigma}^{d} 2b_r|y|^{2\lambda-2r-2}y_j^{2r-d+1} + \sum_{r=\sigma}^{d}(2r-d)b_r|y|^{2\lambda-2r}y_j^{2r-d-1}\right\} \\
&= \sum_{r=\sigma}^{d} 4b_r|y|^{2\lambda-2r-4}y_j^{2r-d+2} + \sum_{r=\sigma}^{d} 2b_r(2r-d+1)|y|^{2\lambda-2r-2}y_j^{2r-d} \\
&\quad + \sum_{r=\sigma}^{d} 2b_r(2r-d)|y|^{2\lambda-2r-2}y_j^{2r-d} + \sum_{r=\tau}^{d} b_r(2r-d)(2r-d-1)|y|^{2\lambda-2r}y_j^{2r-d-2}.
\end{aligned}$$

Here, the appearance of τ in the range of summation in the last term is due to the fact that the act of differentiation eventually annihilates y_j in some of the terms. When d is even, $\tau = \sigma+2$ and when d is odd, $\tau = \sigma+1$. We now re-write the four terms as

$$\begin{aligned}
&\sum_{r=\sigma+2}^{d+2} 4b_{r-2}|y|^{2\lambda-2r}y_j^{2r-d-2} + \sum_{r=\sigma+1}^{d+1} 2b_{r-1}(2r-d-1)|y|^{2\lambda-2r}y_j^{2r-d-2} \\
&+ \sum_{r=\sigma+1}^{d+1} 2b_{r-1}(2r-d-2)|y|^{2\lambda-2r}y_j^{2r-d-2} + \sum_{r=\tau}^{d} 2b_r(2r-d)(2r-d-1)|y|^{2\lambda-2r}y_j^{2r-d-2},
\end{aligned}$$

and collect them together to obtain the desired form $\sum_{r=\sigma+1}^{d+2} z_i|y|^{2\lambda-2r}y_j^{2r-(d+2)}$. Observing that

$$\frac{\partial E}{\partial y_j} = 2\lambda y_j |y|^{2\lambda-2} \qquad \text{and} \qquad \frac{\partial^2 E}{\partial y_j^2} = 2\lambda |y|^{2\lambda-2} + 4\lambda(\lambda-1) y_j^2 |y|^{2\lambda-4},$$

we see that we have initial cases for both odd and even d, and hence, the proof is complete. ■

4.2.7 Corollary *Let γ be an arbitrary multi-index. Then, for all $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$,*

$$(D^\gamma E)(y) = \sum_{j_1=\sigma_1}^{\gamma_1} b_{j_1} \sum_{j_2=\sigma_2}^{\gamma_2} b_{j_2} \cdots \sum_{j_n=\sigma_n}^{\gamma_n} b_{j_n} |y|^{2\lambda - 2(j_1+\cdots+j_n)} \prod_{r=1}^{n} y_r^{2j_r - \gamma_r},$$

where σ_r is the largest integer less than $(\gamma_r + 1)/2$ and the coefficients $\{b_{j_r}\}$ depend only on λ and γ.

Proof. Having established the form of the κ^{th} partial derivative of E, we note that it contains no products $y_r y_t$ outside the modulus. Hence, if we were then to differentiate with respect to another variable, we can treat it as just the differentiation of a diminished power of the modulus function. ■

4.2.8 Lemma *Let k be a non-negative integer. If Λ is a solution of the distributional equation $|\cdot|^{2k}\Lambda = 0$, then there exist coefficients $\{z_\gamma\}$ such that $\Lambda = \sum_{|\gamma|<2k} z_\gamma D^\gamma \delta$.*

Proof. Let Ω be any open set containing the origin. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ within a neighbourhood, N, of the origin, and $\operatorname{supp}\{\rho\} = \Omega$. Define ρ_h for all x in $\mathbb{R}^n$ and $h > 0$ by $\rho_h(x) = \rho(x/h)$. Then $\operatorname{supp}\{\rho_h\} = h\Omega$.

Let η_h be chosen now from $C_0^\infty(\mathbb{R}^n)$ such that, for all $h > 0$, $\eta_h > 0$, and for all y in $\mathbb{R}^n \setminus hN$, $\eta_h(y) = |y|^{2k}$. Consequently, for all ψ in $\mathscr{D}$ and $h > 0$,

$$[\Lambda, \psi] - [\rho_h \Lambda, \psi] \quad = \quad [\Lambda(1-\rho_h), \psi]$$

$$\begin{aligned} &= [\Lambda, (1-\rho_h)\psi] \\ &= \left[\Lambda, |\cdot|^{2k}\frac{(1-\rho_h)}{\eta_h}\psi\right] \\ &= \left[|\cdot|^{2k}\Lambda, \frac{(1-\rho_h)}{\eta_h}\psi\right] \\ &= 0. \end{aligned}$$

Therefore, $\Lambda = \rho_h\Lambda$ and furthermore, $\operatorname{supp}\{\Lambda\} = \operatorname{supp}\{\rho_h\Lambda\}$. This equality can only be maintained for all $h > 0$, if $\operatorname{supp}\{\Lambda\} = \{0\}$. Since we have deduced that $\operatorname{supp}\{\Lambda\}$ is compact, the theory of distributions (c.f. Rudin [27], **6.24**) tells us that Λ has finite order κ and so, $\Lambda = \sum_{|\gamma|\le\kappa} z_\gamma D^\gamma\delta$ (c.f. Rudin [27], **6.25**).

In order that Λ satisfies the equation $|\cdot|^{2k}\Lambda = 0$, we now require that, for all ψ in $\mathscr{D}$,

$$0 = [|\cdot|^{2k}\Lambda, \psi] = [\Lambda, |\cdot|^{2k}\psi] = \sum_{|\gamma|\le\kappa}(-1)^{|\gamma|}z_\gamma(D^\gamma\{|\cdot|^{2k}\psi\})(0).$$

Using Leibniz's formula, we can find constants $\{\xi_{\beta\gamma}\}_{\beta\le\gamma}$ such that

$$(D^\gamma\{|\cdot|^{2k}\psi\})(0) = \sum_{\beta\le\gamma}\xi_{\beta\gamma}(D^\beta\{|\cdot|^{2k}\})(0)(D^{\gamma-\beta}\psi)(0). \tag{4.7}$$

From **4.2.7**, we know that $(D^\beta\{|\cdot|^{2k}\})(0) = 0$, whenever $|\beta| < 2k$. Therefore, $|\cdot|^{2k}\Lambda = 0$ if, and only if, $\Lambda = \sum_{|\gamma|<2k} z_\gamma D^\gamma\delta$. ∎

4.2.9 Corollary *Given any weight function w and non-negative integer k, all tempered solutions Λ of $|\cdot|^{2k}\widehat{\Lambda} = 1/w$, differ from one another by polynomials of degree less than $2k$.*

Proof. From **4.2.8**, $\widehat{\Lambda}$ is a linear combination of the delta distribution, and its derivatives, up to, but not including order $2k$. Inverting the Fourier transform yields a polynomial from π_{2k-1}. ∎

Schwartz also introduced in his *Theorie des Distributions* a method for defining the distributional solutions of division problems. This requires some technical experience of Hademard's finite part of a divergent integral and the resulting definition of a pseudo-function. Rather than follow this path, we will introduce an equivalent concept, with fewer tools, tailored to the very specific type of singularity we are dealing with.

Let $\mathscr{S}_0$ be the subspace of $\mathscr{S}$ defined by

$$\mathscr{S}_0 = \{\psi \in \mathscr{S} : (D^\alpha \psi)(0) = 0, \text{ for all } \alpha \text{ satisfying } |\alpha| < 2k\}.$$

Clearly, examining the action of a tempered distribution on $\mathscr{S}_0$ does not define the distribution uniquely. If we denote the equivalence classes of restrictions to $\mathscr{S}_0$ of elements in $\mathscr{S}'$ by $\mathscr{S}_0{}^*$, then $\mathscr{S}_0{}^*$ is of finite co-dimension in $\mathscr{S}'$. The distributions $\{D^\beta \delta : \beta \in \mathbb{Z}_+^n \text{ and } |\beta| < 2k\}$ span the complement of $\mathscr{S}_0^*$ in $\mathscr{S}'$, and so, determining the action of a tempered distribution on $\mathscr{S}_0$ specifies that distribution except for a distribution of the form $\sum_{|\beta|<2k} z_\beta D^\beta \delta$.

We are going to partially describe a functional χ as follows. Given any non-negative integer k and weight function w, we define the action of χ on any ψ in $\mathscr{S}_0$ by

$$[\chi, \psi] = \int_{\mathbb{R}^n} \frac{1}{|y|^{2k} w(y)} \psi(y)\, dy. \tag{4.8}$$

Our next result shows that these integrals exist, and furthermore, that $\chi \in \mathscr{S}_0'$. As such, it has many extensions to the whole of $\mathscr{S}$, which by the above argument differ by linear combinations of delta distributions, and their derivatives.

4.2.10 Lemma *Let k be a non-negative integer, and let w_μ be a weight function. For each ψ in $\mathscr{S}_0$, let the action of χ on ψ be defined by (4.8). Then $\chi \in \mathscr{S}_0'$.*

Proof. We begin by showing that the action of χ on elements in $\mathscr{S}_0$ is well-defined. To this end, we split the integral in (4.8) as follows:

$$\int_{\mathbb{R}^n} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy = \int_{|y|<1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy + \int_{|y|\geq 1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy.$$

Applying a Taylor's series argument to the first integral gives

$$\int_{|y|<1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy = \int_{|y|<1} \frac{1}{|y|^{2k} w_\mu(y)} 2k \int_0^1 (1-t)^{2k-1} \sum_{|\alpha|=2k} y^\alpha \frac{(D^\alpha \psi)(ty)}{\alpha!}\, dt\, dy.$$

Hence,

$$\begin{aligned} \left| \int_{|y|<1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy \right| &\leq \int_{|y|<1} \frac{1}{|y|^{2k} w_\mu(y)} |y|^{2k} \max_{|\alpha|=2k} \sup_{|u|<1} \frac{|(D^\alpha \psi)(u)|}{\alpha!}\, dy \\ &\leq \max_{|\alpha|=2k} \|D^\alpha \psi\|_\infty \int_{|y|<1} \frac{1}{w_\mu(y)}\, dy, \end{aligned}$$

which is finite, recalling the properties of the weight function.

Returning to the second integral, and recalling now the growth properties of $1/w_\mu$, and the continuity of w_μ outside of the unit ball, we can find a constant $A > 0$ such that $[w_\mu(y)]^{-1} \leq A|y|^{-2\mu}$, for all y in $\mathbb{R}^n$ satisfying $|y| \geq 1$. Hence,

$$\left| \int_{|y|\geq 1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy \right| \leq A \left| \int_{|y|\geq 1} |y|^{-2k-2\mu} \psi(y)\, dy \right|.$$

Now choose a multi-index γ such that $|\gamma| > n - 2\mu - 2k$. Then,

$$\left| \int_{|y|\geq 1} \frac{\psi(y)}{|y|^{2k} w_\mu(y)}\, dy \right| \leq A \sup_{y\in\mathbb{R}^n} \{|y^\gamma \psi(y)|\} \int_{|y|\geq 1} |y|^{-2k-2\mu-|\gamma|}\, dy.$$

The integral is finite because of the conditions on γ. Overall then, we conclude that there exists a constant $B > 0$ such that

$$\left| \int_{\mathbb{R}^n} \frac{\psi(y)}{|y|^{2k} w_\mu(y)} \, dy \right| \le B \left(\sup_{|\alpha|=2k} \{\|D^\alpha \psi\|_\infty\} + \sup_{y \in \mathbb{R}^n} \{|y^\gamma \psi(y)|\} \right).$$

Once more, applying the above inequality in a similar manner to that seen in **4.1.4**, to a sequence from $\mathscr{S}_0$ which converges to zero in the topology on $\mathscr{S}$, we see that $\chi \in \mathscr{S}_0'$ by sequential continuity. ■

Our next result highlights the relationship between χ and the solutions of the equation $|\cdot|^{2k}\widehat{\Lambda} = 1/w$.

4.2.11 Lemma *Let w be a weight function, and let k be a non-negative integer. Let Λ be any tempered distribution satisfying $|\cdot|^{2k}\widehat{\Lambda} = 1/w$, and let χ be as defined in* **4.2.10**. *Then $\widehat{\Lambda}$ is an extension of χ to the whole of $\mathscr{S}$.*

Proof. We need only reassure ourselves that, for any ψ_0 in $\mathscr{S}_0$, $[\widehat{\Lambda}, \psi_0] = [\chi, \psi_0]$. However, for all multi-indices γ satisfying $|\gamma| < 2k$, we know from (4.7), that for any ψ in $\mathscr{S}$,

$$(D^\gamma\{|\cdot|^{2k}\psi\})(0) = \sum_{\beta \le \gamma} \xi_{\beta\gamma} (D^\beta\{|\cdot|^{2k}\})(0)(D^{\gamma-\beta}\psi)(0) = 0.$$

Hence, $|\cdot|^{2k}\psi \in \mathscr{S}_0$, and so, letting χ_1 denote any extension of χ to the whole of $\mathscr{S}$,

$$[|\cdot|^{2k}\chi_1, \psi] = [\chi_1, |\cdot|^{2k}\psi] = [\chi, |\cdot|^{2k}\psi] = \int_{\mathbb{R}^n} \frac{|y|^{2k}\psi(y)}{|y|^{2k}w(y)} \, dy = \left[\frac{1}{w}, \psi\right].$$

Hence, χ_1 is a solution of $|\cdot|^{2k}\Delta = 1/w$. From the theory of distributions, these solutions differ by solutions of the homogeneous equation $|\cdot|^{2k}\Delta = 0$, and by **4.2.8**, the latter take the form $\sum_{|\gamma|<2k} z_\gamma D^\gamma \delta$. Therefore, for any ψ in $\mathscr{S}$,

$$[\widehat{\Lambda} - \chi_1, \psi] = \sum_{|\gamma|<2k} (-1)^{|\gamma|} z_\gamma (D^\gamma \psi)(0).$$

In particular, $[\widehat{\Lambda} - \chi, \psi_0] = [\widehat{\Lambda} - \chi_1, \psi_0] = 0$, which completes the proof. ■

Summarising these results, given a weight function w_μ and a non-negative integer k, we can find many solutions of the equation $|\cdot|^{2k}\widehat{\Lambda} = 1/w_\mu$, all of which are continuous functions of degree j, where j is the integer part of $2k+2\mu-n$, and differ from one another by polynomials of degree less than $2k$. Moreover, given any ψ in $\mathscr{S}_0$, the action of $\widehat{\Lambda}$ on ψ is given by the integral

$$\int_{\mathbb{R}^n} \frac{1}{|y|^{2k} w_\mu(y)} \psi(y)\, dy.$$

Since Λ may contain polynomials which clearly do not lie in the space, it is unreasonable to expect Λ to lie in $X^{k,w}$. However, the assembled facts about solutions of $|\cdot|^{2k}\widehat{\Lambda} = 1/w$ will help us prove the following result, in which it will be convenient to let T_x denote the shift operator, so that $(T_x f)(y) = f(y-x)$. In terms of distributions, if $f \in \mathscr{D}'$, $T_x f \in \mathscr{D}'$ and $[T_x f, \psi] = [f, T_{-x}\psi]$.

4.2.12 Theorem *Let w_μ be a weight function, and let k and n be integers, chosen so that $k \geq 0$, $n > 0$, and $2k + 2\mu - n > 0$. Next, let ϕ be any solution of the equation $|\cdot|^{2k}\widehat{\phi} = 1/w_\mu$, and let $a_1, \ldots, a_\ell$ and $p_1, \ldots, p_\ell$ be as defined in* **4.2.3**. *Then, for all multi-indices β satisfying $|\beta| < k + \mu - n/2$, and for each point x in $\mathbb{R}^n$, the tempered distribution q defined by*

$$q = (-1)^{|\beta|} T_x D^\beta \phi - \sum_{s=1}^{\ell} (D^\beta p_s)(x) T_{a_s} \phi,$$

belongs to $X^{k,w}$.

Proof. Fix x in $\mathbb{R}^n$. Being a linear combination of shifts of derivatives of ϕ, q defines a tempered distribution which can be written as the convolution product $\rho * \phi$, where

$$\rho = (-1)^{|\beta|} D^\beta T_x \delta - \sum_{s=1}^{\ell} (D^\beta p_s)(x) T_{a_s} \delta.$$

Since ρ is compactly supported, the theory of distributions tells us that $\widehat{\rho} \in C^{\infty}(\mathbb{R}^n)$ (c.f. Rudin [27] **7.23**), as can be seen from either of the two representations,

$$\widehat{\rho}(y) = (-1)^{|\beta|} e^{-ixy} (iy)^{\beta} - \sum_{s=1}^{\ell} (D^{\beta} p_s)(x) e^{-ia_s y},$$

or

$$\widehat{\rho}(y) = D_x^{\beta} \Big\{ e^{-ixy} - \sum_{s=1}^{\ell} p_s(x) e^{-ia_s y} \Big\}. \tag{4.9}$$

Here, the subscript in the derivative notation in the second form of $\widehat{\rho}$ indicates that the derivatives should be taken with respect to x. Now, for any $\epsilon > 0$, consider $|\widehat{\rho}(y)|$ for $|y| > \epsilon$. The first form of $\widehat{\rho}$ above can be manipulated to reveal that

$$\begin{aligned}
|\widehat{\rho}(y)| &= \Big| (-1)^{|\beta|} (iy)^{\beta} e^{-ixy} - \sum_{s=1}^{\ell} (D^{\beta} p_s)(x) e^{-ia_s y} \Big| \\
&\leq |y^{\beta}| + \sum_{s=1}^{\ell} |(D^{\beta} p_s)(x)| \\
&\leq |y|^{|\beta|} \Big\{ 1 + |y|^{-|\beta|} \sum_{s=1}^{\ell} |(D^{\beta} p_s)(x)| \Big\} \\
&\leq |y|^{|\beta|} \Big\{ 1 + |\epsilon|^{-|\beta|} \sum_{s=1}^{\ell} |(D^{\beta} p_s)(x)| \Big\} \\
&= C_1 |y|^{|\beta|}.
\end{aligned}$$

Now suppose $|y| \leq \epsilon < 1$. Then,

$$e^{-ixy} - \sum_{s=1}^{\ell} p_s(x) e^{-ia_s y} = \sum_{j=0}^{\infty} \Big\{ \frac{(-ixy)^j}{j!} - \sum_{s=1}^{\ell} p_s(x) \frac{(-ia_s y)^j}{j!} \Big\}.$$

Since, for any f in $C(\mathbb{R}^n)$, $\sum_{s=1}^{\ell} f(a_s) p_s$ is the Lagrange interpolatory polynomial for f

on $a_1, \ldots, a_\ell$, it follows that

$$e^{-ixy} - \sum_{s=1}^{\ell} p_s(x) e^{-ia_s y} = \sum_{j=k}^{\infty} \left\{ \frac{(-ixy)^j}{j!} - \sum_{s=1}^{\ell} p_s(x) \frac{(-ia_s y)^j}{j!} \right\}.$$

Therefore, the second form of $\widehat{\rho}$ in (4.9) above, satisfies

$$D_x^\beta \left\{ e^{-ixy} - \sum_{s=1}^{\ell} p_s(x) e^{-ia_s y} \right\} =$$
$$\sum_{j=k+|\beta|}^{\infty} \left\{ (-iy)^\beta \frac{(-ixy)^{j-|\beta|}}{(j-|\beta|)!} \right\} - \sum_{j=k}^{\infty} \left\{ \sum_{s=1}^{\ell} (D^\beta p_s)(x) \frac{(-ia_s y)^j}{j!} \right\}.$$

Subsequently, for all multi-indices τ,

$$\begin{aligned}
(D^\tau \widehat{\rho})(y) &= \sum_{j=k+|\beta|}^{\infty} \left\{ (-i)^{|\beta|} y^{\beta-\tau} \frac{(-ixy)^{j-|\beta|}}{(j-|\beta|)!} \right\} \\
&\quad + \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ (-iy)^\beta (-ix)^\tau \frac{(-ixy)^{j-|\beta|-|\tau|}}{(j-|\beta|-|\tau|)!} \right\} \\
&\quad - \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} (D^\beta p_s)(x) (-ia_s)^\tau \frac{(-ia_s y)^{j-|\tau|}}{(j-|\tau|)!} \right\}.
\end{aligned}$$

For $|\tau| \le |\beta|$, we have

$$\begin{aligned}
|(D^\tau \widehat{\rho})(y)| &\le \sum_{j=k+|\beta|}^{\infty} \left\{ |y|^{|\beta|-|\tau|} \frac{|xy|^{j-|\beta|}}{(j-|\beta|)!} \right\} \\
&\quad + \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ |y|^{|\beta|} |x|^{|\tau|} \frac{|xy|^{j-|\beta|-|\tau|}}{(j-|\beta|-|\tau|)!} \right\} \\
&\quad + \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} |(D^\beta p_s)(x)| |a_s|^{|\tau|} \frac{|a_s y|^{j-|\tau|}}{(j-|\tau|)!} \right\} \\
&\le |y|^{-|\tau|} \left[\sum_{j=k+|\beta|}^{\infty} \left\{ |y|^j \frac{|x|^{j-|\beta|}}{(j-|\beta|)!} \right\} \right. \\
&\quad + \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ |y|^j \frac{|x|^{j-|\beta|}}{(j-|\beta|-|\tau|)!} \right\}
\end{aligned}$$

$$
\begin{aligned}
&\qquad + \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} |(D^{\beta}p_s)(x)||y|^{j}\frac{|a_s|^{j-|\tau|}}{(j-|\tau|)!} \right\} \Bigg] \\
&\le |y|^{k-|\tau|} \Bigg[\sum_{j=k+|\beta|}^{\infty} \left\{ |y|^{j-k}\frac{|x|^{j-|\beta|}}{(j-|\beta|)!} \right\} \\
&\qquad + \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ |y|^{j-k}\frac{|x|^{j-|\beta|}}{(j-|\beta|-|\tau|)!} \right\} \\
&\qquad\qquad + \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} |(D^{\beta}p_s)(x)||y|^{j-k}\frac{|a_s|^{j-|\tau|}}{(j-|\tau|)!} \right\} \Bigg] \\
&\le |y|^{k-|\tau|} \Bigg[\sum_{j=k+|\beta|}^{\infty} \left\{ \frac{|x|^{j-|\beta|}}{(j-|\beta|)!} \right\} \\
&\qquad + \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ \frac{|x|^{j-|\beta|}}{(j-|\beta|-|\tau|)!} \right\} \\
&\qquad\qquad + \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} |(D^{\beta}p_s)(x)|\frac{|a_s|^{j-|\tau|}}{(j-|\tau|)!} \right\} \Bigg] \\
&\le C_2|y|^{k-|\tau|},
\end{aligned}
$$

For $|\tau| > |\beta|$, the first term in (4.10) disappears leaving

$$
\begin{aligned}
|(D^{\tau}\widehat{\rho})(y)| \le & \sum_{j=k+|\beta|+|\tau|}^{\infty} \left\{ |y|^{|\beta|}|x|^{|\tau|}\frac{|xy|^{j-|\beta|-|\tau|}}{(j-|\beta|-|\tau|)!} \right\} \\
& + \sum_{j=k+|\tau|}^{\infty} \left\{ \sum_{s=1}^{\ell} |(D^{\beta}p_s)(x)||a_s|^{|\tau|}\frac{|a_s y|^{j-|\tau|}}{(j-|\tau|)!} \right\},
\end{aligned}
$$

and using the same style of argument, we find that $|(D^{\tau}\widehat{\rho})(y)| \le C_3|y|^{k-|\tau|}$.

Let α and γ be multi-indices satisfying $|\alpha| = k$ and $|\gamma| < 2k$, and let ψ be any test function from $\mathscr{S}$. Then, having established the behaviour of $\widehat{\rho}$, Leibniz's formula for repeated differentiation yields coefficients $\{\xi_{\lambda\gamma}\}$ and $\{\xi'_{\sigma\lambda}\}$, $\sigma \le \lambda \le \gamma$, for which

$$
(D^{\gamma}\{v_{\alpha}\,\widehat{\rho}\,\psi\})(y) = \sum_{\lambda\le\gamma} \xi_{\lambda\gamma}(i)^{k}y^{\alpha-\gamma+\lambda}(D^{\lambda}\{\widehat{\rho}\psi\})(y)
$$

$$= \sum_{\lambda \le \gamma} \xi_{\lambda\gamma} (i)^k y^{\alpha-\gamma+\lambda} \sum_{\sigma \le \lambda} \xi'_{\sigma\lambda} (D^\sigma \widehat{\rho})(y)(D^{\lambda-\sigma}\psi)(y),$$

so that, for $|y| < 1$,

$$|(D^\gamma \{v_\alpha \widehat{\rho}\, \psi\})(y)| \le \sum_{\lambda \le \gamma} \xi_{\lambda\gamma} \sum_{\sigma \le \lambda} \xi'_{\sigma\lambda} |(D^{\lambda-\sigma}\psi)(y)| C_4 |y|^{2k-|\gamma|+|\beta|-|\lambda|}.$$

The right hand side tends to zero as $|y| \to 0$, and so, we conclude that $v_\alpha \widehat{\rho}\, \psi$ lies in $\mathscr{S}_0$. We now write $\widehat{D^\alpha q}$ as $(D^\alpha(\rho * \phi))\widehat{\ } = v_\alpha \widehat{\rho}\, \widehat{\phi}$. Then, for any ψ in $\mathscr{S}$,

$$[v_\alpha \widehat{\rho}\, \widehat{\phi}, \psi] = [\widehat{\phi}, v_\alpha \widehat{\rho}\, \psi],$$

but, by **4.2.11**, $\widehat{\phi}$ is an extension of χ so that

$$\begin{aligned} [\widehat{\phi}, v_\alpha \widehat{\rho}\, \psi] &= [\chi, v_\alpha \widehat{\rho}\, \psi] \\ &= \int_{\mathbb{R}^n} \frac{v_\alpha(y)\widehat{\rho}(y)}{|y|^{2k} w_\mu(y)} \psi(y)\, dy. \end{aligned}$$

Let K be any compact set in $\mathbb{R}^n$. Then, letting B_t denote the ball of radius t centred at the origin, and recalling that we are still considering $\epsilon < 1$,

$$\int_K \left| \frac{v_\alpha(y)\widehat{\rho}(y)}{|y|^{2k} w_\mu(y)} \right| dy \le \int_{K \setminus B_\epsilon} \left| \frac{v_\alpha(y)\widehat{\rho}(y)}{|y|^{2k} w_\mu(y)} \right| dy + C_2 \int_{B_\epsilon} \frac{1}{w_\mu(y)}\, dy,$$

both of which are finite. Therefore, we have found a locally integrable function F equal almost everywhere to $(v_\alpha \widehat{\rho})/(|\cdot|^{2k} w_\mu)$ for which

$$[(D^\alpha q)\widehat{\ }, \psi] = \int_{\mathbb{R}^n} F(y)\psi(y)\, dy, \qquad \psi \in \mathscr{S}.$$

It therefore, only remains to show that $|q|_{k,w}$ is finite. From the definition of w_μ, we can find a constant C_5 such that for all $|y| > \epsilon$, $\{w_\mu(y)\}^{-1} \leq C_5|y|^{-2\mu}$. Therefore,

$$\begin{aligned}
|q|^2_{k,w} &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |(D^\alpha q)\widehat{\ }(y)|^2 w(y)\, dy \\
&= \int_{\mathbb{R}^n} \frac{\sum_{|\alpha|=k} c_\alpha y^{2\alpha} |\widehat{\rho}(y)|^2}{|y|^{4k} w_\mu(y)^2} w_\mu(y)\, dy \\
&= \int_{\mathbb{R}^n} \frac{|\widehat{\rho}(y)|^2}{|y|^{2k} w_\mu(y)}\, dy \\
&\leq \int_{|y|>\epsilon} C_1 C_5 |y|^{2|\beta|-2k-2\mu}\, dy + C_2 \int_{|y|\leq\epsilon} \frac{1}{w_\mu(y)}\, dy.
\end{aligned}$$

The second integral is easily seen to be finite — the first term is finite when $2|\beta| - 2k - 2\mu + n - 1 < -1$, or $|\beta| < k + \mu - n/2$. ∎

Our next result is fundamental to our interpolation problems. On one hand, it may be viewed as an extension of **4.1.18**, showing that $X^{k,w}$ can be continuously embedded in $C^j(\mathbb{R}^n)$, for some non-negative integer j. For our purposes, we shall view it as simply the statement that point evaluation functionals are bounded over the Hilbert space. This then assures us of the existence of the representers with which we hope to recover the reproducing kernel.

4.2.13 Theorem *Let ℓ, $\{a_1, \dots, a_\ell\}$ and $\{p_1, \dots, p_\ell\}$ be as defined in* **4.2.3**, *along with the mapping P. Let w_μ be a weight function, and let k and n be positive integers, chosen so that $k+\mu-n/2 > 0$. Let β be any multi-index for which $|\beta| < k+\mu-n/2$. Then there exists a constant $K > 0$ such that, for all f in $X_{k,w}$ satisfying $Pf = 0$, $|(D^\beta f)(x)| \leq K|f|_{k,w}$.*

Proof. Choose ϕ from $\mathscr{S}'$ so that $|\cdot|^{2k}\widehat{\phi} = 1/w_\mu$. Since $1/w_\mu$ is locally integrable, $|\cdot|^{2k}\widehat{\phi}$ is a distribution of order zero, and so, multiplication by continuous functions is well-defined. Let ψ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\psi(y)$ is a decreasing function of $|y|$ outside of some neighbourhood of the origin, yet inside, $\psi = 1$. Define ψ_j for all y in $\mathbb{R}^n$ and j in $\mathbb{N}$, by

$\psi_j(y) = 1 - \psi(jy)$. Then, $w_\mu \psi_j \in C(\mathbb{R}^n)$, for all j, and so,

$$w_\mu \psi_j |\cdot|^{2k}\widehat{\phi} = w_\mu \psi_j \frac{1}{w_\mu} = \psi_j.$$

Let ρ be a compactly supported distribution and suppose f is chosen arbitrarily from $X^{k,w}$. Since f is tempered, $\rho * f$ is tempered and has a Fourier transform which satisfies

$$\begin{aligned}
(\rho * f)\widehat{\ }\psi_j &= \widehat{\rho}\widehat{f}\psi_j \\
&= \widehat{\rho}\widehat{f}\psi_j w_\mu |\cdot|^{2k}\widehat{\phi} \\
&= \psi_j \widehat{\rho}\widehat{f} w_\mu (-1)^k \sum_{|\alpha|=k} c_\alpha |v_\alpha|^2 \widehat{\phi} \\
&= (-1)^k \psi_j \sum_{|\alpha|=k} c_\alpha (v_\alpha \widehat{\rho}\widehat{\phi})(\overline{v}_\alpha \widehat{f}) w_\mu \\
&= \psi_j \sum_{|\alpha|=k} c_\alpha \{D^\alpha(\rho * \phi)\}\widehat{\ }\{D^\alpha f\}\widehat{\ } w_\mu.
\end{aligned}$$

Now fix x in $\mathbb{R}^n$, and take

$$\rho = (-1)^{|\beta|} T_x D^\beta \delta - \sum_{s=1}^{\ell} (D^\beta p_s)(x) T_{a_s} \delta.$$

Since $f \in X^{k,w}$, and by **4.1.18**, $X^{k,w} \subset C^{|\beta|}(\mathbb{R}^n)$,

$$(\rho * f)(y) = (-1)^{|\beta|}(D^\beta f)(y - x) - \sum_{s=1}^{\ell} (D^\beta p_s)(x) f(y - a_s), \qquad y \in \mathbb{R}^n.$$

Now, from **4.2.12**, we know that $\rho * \phi \in X^{k,w}$, which means that $\{D^\alpha(\rho * \phi)\}\widehat{\ }\sqrt{w_\mu}$ belongs to $L^2(\mathbb{R}^n)$. Similarly, since $f \in X^{k,w}$, the same is true of $\{D^\alpha f\}\widehat{\ }\sqrt{w_\mu}$. Hence, $(\rho * f)\widehat{\ }\psi_j$ is measurable, and

$$\int_{\mathbb{R}^n} |(\rho * f)\widehat{\ }(y)\psi_j(y)|\, dy = \int_{\mathbb{R}^n} \left| \psi_j(y) \sum_{|\alpha|=k} c_\alpha \{D^\alpha(\rho * \phi)\}\widehat{\ }(y)\{D^\alpha f\}\widehat{\ }(y) w_\mu(y) \right| dy$$

$$
\begin{aligned}
&\le \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} \left|\{D^\alpha(\rho * f)\}^\frown(y)\{D^\alpha f\}^\frown(y) w_\mu\right| dy \\
&\le \sum_{|\alpha|=k} c_\alpha \left\{ \int_{\mathbb{R}^n} |\{D^\alpha(\rho * \phi)\}^\frown(y)|^2 w_\mu(y)\, dy \right\}^{\frac{1}{2}} \left\{ \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 w_\mu(y)\, dy \right\}^{\frac{1}{2}} \\
&\le \left\{ \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |\{D^\alpha(\rho * \phi)\}^\frown(y)|^2 w_\mu(y)\, dy \right\}^{\frac{1}{2}} \left\{ \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} |(\widehat{D^\alpha f})(y)|^2 w_\mu(y)\, dy \right\}^{\frac{1}{2}} \\
&= |\rho * \phi|_{k,w} |f|_{k,w}.
\end{aligned}
$$

This shows that $(\rho * f)^\frown \psi_j \in L^1(\mathbb{R}^n)$, and

$$\|(\rho * f)^\frown \psi_j\|_1 \le |\rho * \phi|_{k,w}|f|_{k,w}.$$

Furthermore, for each y in $\mathbb{R}^n$, $\{|(\rho * f)^\frown(y)\psi_j(y)|\}$ is an increasing sequence of numbers which converges to $|(\rho * f)^\frown(y)|$ as $j \to \infty$, providing we adjust $(\rho * f)^\frown$ (on a set of Lesbegue measure zero) by setting $(\rho * f)^\frown(0) = 0$. Lesbegue's monotone convergence theorem (c.f. Rudin [28]) then shows that $(\rho * f)^\frown \in L^1(\mathbb{R}^n)$.

Taking the Fourier transform once more, we now see that $(\rho * f)^\smile$ is a continuous function which vanishes at infinity. Moreover,

$$
\begin{aligned}
|(\rho * f)^\smile(y)| &\le \|(\rho * f)^\smile\|_\infty \\
&\le (2\pi)^{-\frac{n}{2}} \|(\rho * f)^\frown\|_1 \\
&\le (2\pi)^{-\frac{n}{2}} |\rho * \phi|_{k,w}|f|_{k,w}.
\end{aligned}
$$

Finally, if $Pf = 0$, then $f(a_s) = 0$, $s = 1, \dots, \ell$, and therefore,

$$|(D^\beta f)(x)| = |(\rho * f)^\smile(0)| \le (2\pi)^{-\frac{n}{2}} |\rho * \phi|_{k,w}|f|_{k,w}. \qquad \blacksquare$$

4.2.14 Corollary *Let k, n, β, and w_μ be chosen as in* **4.2.13,** *so that* $0 \le |\beta| < k + \mu -$

$n/2$. *Then, given any x in $\mathbb{R}^n$, there exists a constant $C > 0$ such that, for all f in $X^{k,w}$, $|(D^\beta f)(x)| \leq C\|f\|_{k,w}$.*

Proof. Suppose ℓ, $\{a_1, \ldots, a_\ell\}$, $\{p_1, \ldots, p_\ell\}$, and P are as defined in **4.2.3**, and that $x \in \mathbb{R}^n$ and $f \in X^{k,w}$. By **4.2.13**, there exists a constant $K > 0$ such that $|(D^\beta(f - Pf))(x)| \leq K|f|_{k,w}$. Hence,

$$\begin{aligned}
|(D^\beta f)(x)| &\leq |(D^\beta(f - Pf))(x)| + |(D^\beta Pf)(x)| \\
&\leq K|f|_{k,w} + \left|\sum_{s=1}^{\ell}(D^\beta p_s)(x)f(a_s)\right| \\
&\leq \max\{K, |(D^\beta p_1)(x)|, \ldots, |(D^\beta p_\ell)(x)|\}\left\{|f|_{k,w} + \sum_{s=1}^{\ell}|f(a_s)|\right\} \\
&\leq \sqrt{\ell+1}\max\{K, |(D^\beta p_1)(x)|, \ldots, |(D^\beta p_\ell)(x)|\}\left\{|f|_{k,w}^2 + \sum_{s=1}^{\ell}|f(a_s)|^2\right\}^{\frac{1}{2}} \\
&= \sqrt{\ell+1}\max\{K, |(D^\beta p_1)(x)|, \ldots, |(D^\beta p_\ell)(x)|\}\,\|f\|_{k,w},
\end{aligned}$$

and the result follows, setting $C = \sqrt{\ell+1}\max\{K, |(D^\beta p_1)(x)|, \ldots, |(D^\beta p_\ell)(x)|\}$. ∎

We now turn to the only remaining problem — that of the representers for the point evaluation functionals and the validity of assumption (iv) in Chapter 3. We begin, however, with an extension of **4.2.8**.

4.2.15 Lemma *Let k be a non-negative integer, and let w be a weight function. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ within a neighbourhood of the origin, and $0 \leq \rho \leq 1$ elsewhere. For all $h > 0$ and x in $\mathbb{R}^n$, define ρ_h by $\rho_h(x) = \rho(x/h)$.*

If, for some f in $X^{k,w}$, it is known that $w(1 - \rho_h)|\cdot|^{2k}\widehat{f} = 0$, for all $h > 0$, then $f \in \pi_{k-1}$.

Proof. Recall from the definition of $X^{k,w}$ that $\widehat{D^\alpha f}$ is locally integrable for all α satisfying

$|\alpha| = k$. It then follows that

$$|\cdot|^{2k}\widehat{f} = \sum_{|\alpha|=k} c_\alpha |v_\alpha|^2 \widehat{f} = \sum_{|\alpha|=k} c_\alpha \overline{v}_\alpha \widehat{D^\alpha f},$$

and so, $|\cdot|^{2k}\widehat{f} \in L^1_{\text{loc}}(\mathbb{R}^n)$. Hence, there exists a locally integrable function G such that, for all ψ in $\mathscr{D}$,

$$[|\cdot|^{2k}\widehat{f}, \psi] = \int_{\mathbb{R}^n} G(y)\psi(y)\,dy.$$

Since $w(1-\rho_h) \in C(\mathbb{R}^n)$ for all $h > 0$, it follows that $w(1-\rho_h)|\cdot|^{2k}\widehat{f}$ is locally integrable and for all ψ in $\mathscr{D}$,

$$[w(1-\rho_h)|\cdot|^{2k}\widehat{f}, \psi] = \int_{\mathbb{R}^n} w(y)(1-\rho_h(y))G(y)\psi(y)\,dy.$$

If $w(1-\rho_h)|\cdot|^{2k}\widehat{f}$ is the zero distribution, then it follows that $w(y)(1-\rho_h(y))G(y) = 0$ almost everywhere. Since this must hold for all $h > 0$, we conclude that G is zero almost everywhere, and that $|\cdot|^{2k}\widehat{f} = 0$. By **4.2.8**, there exist coefficients $\{d_\gamma\}$ such that $\widehat{f} = \sum_{|\gamma|<2k} d_\gamma D^\gamma \delta$. On this evidence alone, f lies in π_{2k-1}, but we must remember the restriction mentioned above, namely that $\widehat{D^\alpha f} \in L^1_{\text{loc}}(\mathbb{R}^n)$ for all α satisfying $|\alpha| = k$. This restricts f to π_{k-1}. ■

4.2.16 Theorem *Let w_μ be a weight function, and let k and n be integers, chosen so that $k \geq 0$, $n > 0$, and $k + \mu - n/2 > 0$. Suppose ℓ, $\{a_1, \ldots, a_\ell\}$ and $\{p_1, \ldots, p_\ell\}$ are as defined in* **4.2.3**, *and let Z denote the subspace of $\mathscr{S}$ whose Fourier transforms have compact support.*

Then, for all multi-indices β satisfying $|\beta| < k + \mu - n/2$, and for all f in $Z \cap X^{k,w}$, *the element R_x^β in $X^{k,w}$ which satisfies $(f, R_x^\beta) = (D^\beta f)(x)$, is given by*

$$\begin{aligned} R_x^\beta &= (-1)^{|\beta|} T_x D^\beta \phi - \sum_{t=1}^{\ell} (D^\beta p_t)(x) T_{a_t} \phi + \sum_{t=1}^{\ell} (D^\beta p_t)(x) p_t \\ &\quad - \sum_{t=1}^{\ell} p_t \Big\{ (-1)^{|\beta|} (T_x D^\beta \phi)(a_t) - \sum_{j=1}^{\ell} (D^\beta p_j)(x)(T_{a_j}\phi)(a_t) \Big\}, \end{aligned}$$

where ϕ is any solution of $|\cdot|^{2k}\widehat{\phi} = 1/w_\mu$.

Proof. Let ρ be chosen from $C_0^\infty(\mathbb{R}^n)$ so that $\rho = 1$ within a neighbourhood N of the origin, and $0 \le \rho \le 1$ elsewhere. For all $h > 0$ and x in $\mathbb{R}^n$, define ρ_h by $\rho_h(x) = \rho(x/h)$ and ψ_h by $\widehat{\psi}_h = \rho_h$.

Let f be any element in Z and set $g_h = f - f * \psi_h$. Then, as seen in the proof of **4.1.17**, g_h lies in $X^{k,w}$. Fix x in $\mathbb{R}^n$. We will first determine the element r_x in $X^{k,w}$ which satisfies $(g_h - Pg_h, r_x) = (D^\beta(g_h - Pg_h))(x)$, where P is as defined in **4.2.3**. On the one hand, r_x must satisfy,

$$\begin{aligned} (D^\beta g_h)(x) - (D^\beta Pg)(x) &= (g_h - Pg_h, r_x) \\ &= \langle g_h, r_x \rangle \\ &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} (\widehat{D^\alpha g_h})(y)(\overline{\widehat{D^\alpha r_x}})(y) w(y)\, dy \\ &= \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^n} (\widehat{D^\alpha f})(y)(\overline{\widehat{D^\alpha r_x}})(y)(1-\rho_h(y)) w(y)\, dy. \end{aligned} \tag{4.10}$$

Now, for r_x to lie in $X^{k,w}$, $\widehat{D^\alpha r_x}$ must be locally integrable for all α satisfying $|\alpha| = k$. Therefore, since the function $(1-\rho_h)w$ is continuous everywhere, $\overline{\widehat{D^\alpha r_x}}(1-\rho_h)w$ is locally integrable and defines a regular distribution whose action on a test function $\widehat{D^\alpha f}$

is described by the above integral (4.10). Hence,

$$\begin{aligned}(D^\beta g_h)(x) - (D^\beta P g_h)(x) &= \sum_{|\alpha|=k} c_\alpha \left[w(1-\rho_h)\overline{\widehat{D^\alpha r_x}}, \widehat{D^\alpha f}\right] \\ &= \left[w(1-\rho_h) \sum_{|\alpha|=k} c_\alpha |v_\alpha|^2 \bar{\hat{r}}_x, \hat{f}\right] \\ &= \left[w(1-\rho_h)|\cdot|^{2k}\bar{\hat{r}}_x, \hat{f}\right].\end{aligned}$$

On the other hand, letting e_z be defined by $e_z(y) = e^{-izy}$, $y \in \mathbb{R}^n$,

$$\begin{aligned}(D^\beta g_h)(x) - (D^\beta P g_h)(x) &= (D^\beta g_h)(x) - \sum_{s=1}^{\ell}(D^\beta p_s)(x) g_h(a_s) \\ &= \left[(-1)^\beta D^\beta T_x\delta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) T_{a_s}\delta, g_h\right] \\ &= \left[\left\{e_x v_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) e_{a_s}\right\}^{\widehat{}}, g_h\right] \\ &= \left[e_x v_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) e_{a_s}, \hat{g}_h\right] \\ &= \left[(1-\rho_h)\left\{e_x v_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) e_{a_s}\right\}, \hat{f}\right].\end{aligned}$$

Since Z is the set of inverse transforms of elements in $\mathscr{D}$, it then follows that, for all ψ in $\mathscr{D}$,

$$\left[w(1-\rho_h)|\cdot|^{2k}\bar{\hat{r}}_x, \psi\right] = \left[(1-\rho_h)\left\{e_x v_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) e_{a_s}\right\}, \psi\right].$$

We therefore obtain, for all $h > 0$, the distributional equality

$$w(1-\rho_h)|\cdot|^{2k}\bar{\hat{r}}_x = (1-\rho_h)\left\{e_x v_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x) e_{a_s}\right\},$$

or equivalently,

$$w(1-\rho_h)|\cdot|^{2k}\widehat{r}_x = (1-\rho_h)\Big\{e_{-x}\overline{v}_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x)e_{-a_s}\Big\}. \tag{4.11}$$

The next step is to construct a particular solution of this equation. Let ϕ be any solution of the equation $|\cdot|^{2k}\widehat{\phi} = 1/w$. Then, for any z in $\mathbb{R}^n$ and multi-index γ, we may reasonably form the product,

$$w(1-\rho_h)|\cdot|^{2k}\widehat{\phi}e_{-z}\overline{v}_\gamma = (1-\rho_h)e_{-z}\overline{v}_\gamma,$$

or equivalently,

$$w(1-\rho_h)|\cdot|^{2k}((-1)^{|\gamma|}T_zD^\gamma\phi)\widehat{\ } = (1-\rho_h)e_{-z}\overline{v}_\gamma.$$

From this general equation, we deduce that

$$w|\cdot|^{2k}(1-\rho_h)\Big\{(-1)^{|\beta|}T_xD^\beta\phi - \sum_{s=1}^{\ell}(D^\beta p_s)(x)T_{a_s}\phi\Big\}\widehat{\ } = \Big\{e_{-x}\overline{v}_\beta - \sum_{s=1}^{\ell}(D^\beta p_s)(x)e_{-a_s}\Big\}(1-\rho_h),$$

and so, a particular solution of (4.11) is given by

$$(-1)^{|\beta|}T_xD^\beta\phi - \sum_{s=1}^{\ell}(D^\beta p_s)(x)T_{a_s}\phi.$$

We know this solution lies in $X^{k,w}$ by **4.2.12**. The general form of r_x then differs from this particular solution by solutions of the homogeneous equation $w(1-\rho_h)|\cdot|^{2k}\Lambda = 0$, which, by **4.2.15**, lie in π_{k-1}. Hence, a candidate for r_x is given by

$$r_x = (-1)^{|\beta|}T_xD^\beta\phi - \sum_{s=1}^{\ell}(D^\beta p_s)(x)T_{a_s}\phi + q,$$

where $q \in \pi_{k-1}$, and this choice is independent of h.

Now, the operator $(I-P)$ annihilates elements in π_{k-1}, and so, noting that the given form of R_x^β may be written as

$$(I-P)r_x + \sum_{s=1}^{\ell} (D^\beta p_s)(x) p_s = r_x - Pr_x + \sum_{s=1}^{\ell} (D^\beta p_s)(x) p_s,$$

the choice of q in the form of r_x becomes irrelevant. Therefore,

$$\begin{aligned}
(g_h, R_x^\beta) &= (g_h, r_x) - (g_h, Pr_x) + \sum_{s=1}^{\ell} (D^\beta p_s)(x)(g_h, p_s) \\
&= (g_h - Pg_h, r_x) + (Pg_h, r_x) - (g_h, Pr_x) + \sum_{s=1}^{\ell} (D^\beta p_s)(x) g_h(a_s) \\
&= (D^\beta g_h)(x) - \sum_{s=1}^{\ell} (D^\beta p_s)(x) g_h(a_s) + \sum_{s=1}^{\ell} (Pg_h)(a_s) r_x(a_s) \\
&\quad - \sum_{s=1}^{\ell} g_h(a_s)(Pr_x)(a_s) + \sum_{s=1}^{\ell} (D^\beta p_s)(x) g_h(a_s).
\end{aligned}$$

Cancelling like terms, we see that $(g, R_x^\beta) = (D^\beta g_h)(x)$. Now, since $\operatorname{supp} \rho_h = hN$, for all multi-indices γ and for all y in $\mathbb{R}^n$,

$$\begin{aligned}
|(D^\gamma g_h)(y) - (D^\gamma f)(y)| &= |[(D^\gamma f) * \psi_h](y)| \\
&= \left| \int_{\mathbb{R}^n} (D^\gamma f)(t) \psi_h(y-t)\, dt \right| \\
&= \left| \int_{\mathbb{R}^n} (D^\gamma f)(t) (T_y \widetilde{\psi}_h)(t)\, dt \right| \\
&= \left| \int_{\mathbb{R}^n} (D^\gamma f)\widehat{\ }(\xi)\, e^{-iy\xi} \bar{\tilde{\rho}}_h(\xi)\, d\xi \right| \\
&= \left| \int_{hN} (D^\gamma f)\widehat{\ }(\xi)\, e^{-iy\xi} \bar{\tilde{\rho}}_h(\xi)\, d\xi \right| \\
&\le \int_{hN} |(D^\gamma f)\widehat{\ }(\xi)|\, d\xi.
\end{aligned}$$

Therefore, for all y in $\mathbb{R}^n$, $|(D^\gamma g_h)(y) - (D^\gamma f)(y)| \to 0$ as $h \to 0$. Furthermore,

$$(g_h, R_x^\beta) = \langle g_h, R_x^\beta \rangle + \sum_{s=1}^{\ell} g_h(a_s) \overline{R_x^\beta(a_s)}$$

$$= \quad \langle f, R_x^\beta\rangle - \langle f * \psi_h, R_x^\beta\rangle + \sum_{s=1}^{\ell} g_h(a_s)\overline{R_x^\beta(a_s)},$$

whenever f lies in $Z \cap X^{k,w}$. Observe then, that

$$|\langle f * \psi_h, R_x^\beta\rangle| \le |f * \psi_h|_{k,w}|R_x^\beta|_{k,w},$$

and, as we saw in the proof of **4.1.19**, the factor $|f * \psi_h|_{k,w}$ tends to zero as $h \to 0$. Therefore, when f lies in $Z \cap X^{k,w}$,

$$\begin{aligned}\lim_{h\to 0}(g_h, R_x^\beta) &= \langle f, R_x^\beta\rangle + \sum_{s=1}^{\ell} f(a_s)\overline{R_x^\beta(a_s)} \\ &= (f, R_x^\beta).\end{aligned}$$

Hence, for all f in $Z \cap X^{k,w}$, $(D^\beta f)(x) = (f, R_x^\beta)$, as required. ∎

4.2.17 Corollary *Under the same hypothesis as detailed in* **4.2.16**, *let f be any element of $X^{k,w}$. Then, given any $\epsilon > 0$, $|(D^\beta f)(x) - (f, R_x^\beta)| < \epsilon$.*

Proof. From the remarks made in **4.2.3**, we know p_s is the representer for the point evaluation at a_s, $s = 1, \ldots, \ell$. This is indeed consistent with the form of R_x^β given in **4.2.16**, when $\beta = 0$, and $x = a_s$, $s = 1, \ldots, \ell$. Moreover, for each $s = 1, \ldots, \ell$, the form of R_x^β quickly reveals that

$$(p_s, R_x^\beta) = R_x^\beta(a_s) = (D^\beta p_s)(x).$$

Therefore, for any q in π_{k-1}, $(q, R_x^\beta) = (D^\beta q)(x)$, and specifically, given any ψ in $Z \cap X^{k,w}$,

$$(\psi - P\psi, R_x^\beta) = (D^\beta\psi)(x) - (D^\beta P\psi)(x).$$

Choose any $\epsilon > 0$, and let f be any element of $X^{k,w}$. By **4.1.19**, we can find a function

ψ in $Z \cap X^{k,w}$ such that $|f-\psi|_{k,w} < \epsilon$. Now,

$$\begin{aligned} (f - Pf, R_x^\beta) &= (f - Pf - \psi + P\psi, R_x^\beta) + (\psi - P\psi, R_x^\beta) \\ &= \langle f - \psi, R_x^\beta \rangle + (D^\beta \psi)(x) - (D^\beta P\psi)(x), \end{aligned}$$

so that

$$\begin{aligned} |(f, R_x^\beta) - (D^\beta f)(x)| &= |(f - Pf, R_x^\beta) - \{(D^\beta f)(x) - (D^\beta Pf)(x)\}| \\ &\leq |\{D^\beta[(\psi - f) - P(\psi - f)]\}(x)| + |\langle f - \psi, R_x^\beta \rangle| \\ &\leq |\{D^\beta[(\psi - f) - P(\psi - f)]\}(x)| + |f - \psi|_{k,w} |R_x^\beta|_{k,w}. \end{aligned}$$

From **4.2.13**, there exists a constant $K > 0$ such that

$$|\{D^\beta[(\psi - f) - P(\psi - f)]\}(x)| \leq K|f - \psi|_{k,w},$$

and so,

$$\begin{aligned} |(f, R_x^\beta) - (D^\beta f)(x)| &\leq \{K + |R_x^\beta|_{k,w}\}|f - \psi|_{k,w} \\ &\leq \epsilon\{K + |R_x^\beta|_{k,w}\}. \qquad \blacksquare \end{aligned}$$

There are now many pressing questions arising from the theory which require attention. The first, and perhaps foremost, is that in all our work we assume the basis function ϕ is merely a solution of $|\cdot|^{2k}\widehat{\phi} = 1/w$. As such it is undetermined up to polynomials of degree less than $2k$, by **4.2.9**, and this property is transferred all the way through to the representer R_x^β. The next result confirms this.

4.2.18 Lemma *Let k be a positive integer, let β be any multi-index and let ℓ, $\{a_1, \ldots, a_\ell\}$*

and $\{p_1, \ldots, p_\ell\}$ *be as defined in* **4.2.3**. *Then, for all* q *in* π_{2k-1},

$$(-1)^{|\beta|}T_x D^\beta q - \sum_{s=1}^{\ell}(D^\beta p_s)(x)T_{a_s}q - \sum_{s=1}^{\ell} p_s\Big\{(-1)^{|\beta|}(T_x D^\beta q)(a_s) - \sum_{j=1}^{\ell}(D^\beta p_j)(T_{a_j}q)(a_s)\Big\}, \tag{4.12}$$

is zero.

Proof. It will suffice to prove the assertion of the lemma for the specific cases when $q(y) = y^\gamma$, for all γ satisfying $|\gamma| < 2k$. Writing,

$$(-1)^{|\beta|}(T_x D^\beta q)(y) = [D^\beta\{(y - \cdot)^\gamma\}](x),$$

we see that, for all x and y in $\mathbb{R}^n$, equation (4.12) reduces to

$$D^\beta\Big[(y-\cdot)^\gamma - \sum_{s=1}^{\ell}(y-a_s)^\gamma p_s(\cdot) - \sum_{s=1}^{\ell}\Big\{(a_s-\cdot)^\gamma - \sum_{j=1}^{\ell}(a_s-a_j)^\gamma p_j(\cdot)\Big\}p_s(y)\Big](x).$$

Therefore, it is sufficient to show that, for all x and y in $\mathbb{R}^n$, and all multi-indices γ satisfying $|\gamma| < 2k$,

$$(y-x)^\gamma - \sum_{s=1}^{\ell}(y-a_s)^\gamma p_s(x) - \sum_{s=1}^{\ell}\Big\{(a_s-x)^\gamma - \sum_{j=1}^{\ell}(a_s-a_j)^\gamma p_j(x)\Big\}p_s(y), \tag{4.13}$$

is zero. Now, using the binomial theorem, we can find constants b_δ, $0 \le \delta \le \gamma$ such that

$$(y-x)^\gamma = \sum_{0\le\delta\le\gamma} b_\delta y^\delta x^{\gamma-\delta}, \qquad x, y \in \mathbb{R}^n.$$

We can therefore expand (4.13) to obtain the expression

$$\sum_{0\le\delta\le\gamma} b_\delta\Big(y^\delta x^{\gamma-\delta} - \sum_{s=1}^{\ell} y^\delta a_s^{\gamma-\delta}p_s(x) - \sum_{s=1}^{\ell}\Big\{a_s^\delta x^{\gamma-\delta} - \sum_{j=1}^{\ell} a_s^\delta a_j^{\gamma-\delta}p_j(x)\Big\}p_s(y)\Big). \tag{4.14}$$

If $|\gamma - \delta| \le k-1$, then for all u in $\mathbb{R}^n$, the fact that the mapping P defined in **4.2.3** is a projection onto π_{k-1} tells us that

$$\begin{aligned} u^\delta x^{\gamma-\delta} - P(u^\delta x^{\gamma-\delta}) &= u^\delta x^{\gamma-\delta} - \sum_{s=1}^{\ell} u^\delta a_s^{\gamma-\delta} p_s(x) \\ &= u^\delta x^{\gamma-\delta} - u^\delta x^{\gamma-\delta} \\ &= 0, \end{aligned}$$

and so, equation (4.14) is certainly zero if $|\gamma - \delta| \le k-1$. Assume then, that $k \le |\gamma - \delta|$. Then $|\delta| < k-1$ and so,

$$\begin{aligned} \sum_{s=1}^{\ell} \left\{ a_s^\delta x^{\gamma-\delta} - \sum_{j=1}^{\ell} a_s^\delta a_j^{\gamma-\delta} p_j(x) \right\} p_s(y) &= \sum_{s=1}^{\ell} a_s^\delta p_s(y) \left\{ x^{\gamma-\delta} - \sum_{j=1}^{\ell} a_j^{\gamma-\delta} p_j(x) \right\} \\ &= y^\delta \left\{ x^{\gamma-\delta} - \sum_{j=1}^{\ell} a_j^{\gamma-\delta} p_j(x) \right\} \\ &= y^\delta x^{\gamma-\delta} - \sum_{j=1}^{\ell} y^\delta a_j^{\gamma-\delta} p_j(x). \end{aligned}$$

This cancels with the first two terms in (4.14) so that, once again, the expression is zero. ■

Our next point concerns the representer for the point evaluation at some fixed point x in $\mathbb{R}^n$. Our theory defines this representer by the equation

$$\begin{aligned} R_x(y) &= \phi(y-x) - \sum_{s=1}^{\ell} p_s(x)\phi(y-a_s) - \sum_{s=1}^{\ell} \phi(a_s - x)p_s(y) \\ &\quad + \sum_{s,t=1}^{\ell} p_s(x)p_t(y)\phi(a_s - a_t) + \sum_{s=1}^{\ell} p_s(x)p_s(y), \end{aligned} \tag{4.15}$$

and the Hilbert space theory then says that $R(x,y) = R_x(y)$ defines the reproducing kernel for $X^{k,w}$. However, the reproducing kernel should exhibit Hermitian symmetry, namely that $R(x,y) = \overline{R(y,x)}$. On inspection of our equation (4.15), though, this is not immediately obvious — it requires that $\phi(x) = \overline{\phi(-x)}$, or $\phi = \overline{\overline{\phi}}$. The next two results help to enlighten us in this respect.

4.2.19 Lemma *For all f in $\mathscr{S}'$, $\hat{\bar{\tilde{f}}} = \bar{\hat{f}}$.*

Proof. We begin by proving the relation for all ψ in $\mathscr{S}$. From the classical expression for the Fourier transform,

$$\hat{\psi}(\xi) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{-it\xi}\psi(t)\,dt,$$

so that

$$\begin{aligned} \bar{\hat{\psi}}(\xi) &= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{it\xi}\bar{\psi}(t)\,dt \\ &= \hat{\hat{\bar{\psi}}}(\xi) \\ &= \tilde{\hat{\bar{\psi}}}(\xi) \\ &= \hat{\bar{\tilde{\psi}}}(\xi). \end{aligned}$$

Subsequently, for all f in $\mathscr{S}'$, and ψ in $\mathscr{S}$,

$$[\hat{\bar{\tilde{f}}}, \psi] = [\bar{\tilde{f}}, \hat{\psi}] = \overline{[\tilde{f}, \bar{\hat{\psi}}]}.$$

Applying the relation to $\bar{\hat{\psi}}$, we have

$$\overline{[\tilde{f}, \bar{\hat{\psi}}]} = \overline{[\tilde{f}, \hat{\bar{\tilde{\psi}}}]} = \overline{[\hat{\tilde{f}}, \bar{\tilde{\psi}}]} = [\tilde{\bar{\hat{f}}}, \tilde{\psi}] = [\bar{\hat{f}}, \psi],$$

which completes the proof. ∎

4.2.20 Lemma *Let k be a positive integer, and let w be a weight function. Let ϕ be any solution of $|\cdot|^{2k}\hat{\phi} = 1/w$. Then ϕ differs from $\bar{\tilde{\phi}}$ by a polynomial of degree less than $2k$.*

Proof. Let ϕ be any solution of $|\cdot|^{2k}\widehat{\phi} = 1/w$. Taking the complex conjugate of both sides, and noting that w is real-valued, we have $|\cdot|^{2k}\widehat{\overline{\overline{\phi}}} = 1/w$. Hence, $|\cdot|^{2k}\widehat{\phi} = |\cdot|^{2k}\widehat{\overline{\overline{\phi}}}$ and, from **4.2.8**, we then know that $\widehat{\phi} = \widehat{\overline{\overline{\phi}}} + Q$, where Q is a complex linear combination of the delta distribution, and its derivatives up to order $2k$. Using **4.2.19**, we therefore have $\widehat{\phi} = \widehat{\overline{\overline{\phi}}} + Q$, and taking inverse transforms completes the proof. ∎

Now, from our first observation in **4.2.18**, the representer R_x is unaffected by modifications in the basis function by polynomials from π_{2k-1}. Hence, as the above result shows, we can replace ϕ in R_x by $\overline{\overline{\phi}} + q$, $q \in \pi_{2k-1}$, without having an effect on the form of R_x. This recovers the Hermitian symmetry.

We finish this section by making the comment that all of the assumptions of Chapter 3 have been satisfied, and hence the error analysis is immediately applicable. The next chapter will concentrate on specific examples of the weight function which induce familiar theories. However, one interesting, and underlying correlation between our theory and preceding theories (especially, those of Madych and Nelson [20, 21, 22], Schaback [29, 30], and Wu and Schaback [39]) depends on the following definition, due to Gel'fand and Vilenkin [13].

4.2.21 Definition *A tempered distribution F is said to be conditionally positive definite of order $s > 0$, if the inequality $[P\overline{P}F, \psi\overline{\psi}] \geq 0$ holds for all test functions ψ from $\mathscr{S}$ and all homogeneous polynomials P of degree s.*

4.2.22 Theorem *Let w be a weight function, and let k be a positive integer. If ϕ is any tempered solution of the distributional equation $|\cdot|^{2k}\widehat{\phi} = 1/w$, then ϕ is conditionally positive definite of order k.*

Proof. Let P be the generic homogeneous polynomial of degree k, defined by $P(y) = \sum_{|\beta|=k} a_\beta(-iy)^\beta$. We require $[P\overline{P}\widehat{\phi}, \psi\overline{\psi}]$ to be non-negative for all ψ in $\mathscr{S}$. However, for any ψ in $\mathscr{S}$, and any multi-index γ satisfying $|\gamma| < 2k$,

$$|(D^\gamma\{|P|^2\psi\})(y)| \le \sum_{\alpha\le\gamma} c_{\alpha\gamma}|(D^\alpha\{|P|^2\})(y)||(D^{\gamma-\alpha}\psi)(y)|.$$

Now,

$$\begin{aligned}
|(D^\alpha\{|P|^2\})(y)| &= \left|D^\alpha\left\{\sum_{|\beta|=k} a_\beta \sum_{|\lambda|=k} \overline{a}_\lambda y^{\beta+\lambda}\right\}\right| \\
&= \left|\sum_{|\beta|=k}\sum_{|\lambda|=k} a_\beta\overline{a}_\lambda y^{\beta+\lambda-\alpha}\right| \\
&\le \sum_{|\beta|=k}\sum_{|\lambda|=k} |a_\beta||a_\lambda||y|^{2k-|\alpha|}.
\end{aligned}$$

As $|y| \to 0$, the right hand side tends to zero, and so we conclude that $P\overline{P}\psi \in \mathscr{S}_0$. Therefore, recalling the definition of χ from **4.2.10**, we see that for any ψ in $\mathscr{S}$,

$$[P\overline{P}\widehat{\phi}, \psi\overline{\psi}] = [\widehat{\phi}, P\overline{P}\psi\overline{\psi}] = [\chi, P\overline{P}\psi\overline{\psi}] = \int_{\mathbf{R}^n} \frac{|P(y)|^2|\psi(y)|^2}{|y|^{2k}w(y)}\,dy,$$

which is finite and, more importantly, non-negative. ∎

Chapter 5

Applications

Having developed a theory for multivariate interpolation, we devote this section to specific examples of $X^{k,w}$ spaces. In Table 5.1, we highlight some of the (k, w) pairings and the familiar theories they generate. Case studies will then extend the error analysis until the strength of the particular interpolation scheme is clear.

However, due to our remarks made in the last chapter, we can immediately provide an extension of the error estimate in **3.3.8**, which holds for every interpolation problem cast in an $X^{k,w}$ space. In final form, it is reminiscent of the error estimates found in Madych and Nelson's early work [21].

5.0.1 Theorem *Let w_μ be a weight function, and let k and n be integers, chosen so that $k \geq 0$, $n > 0$, and $k + \mu - n/2 > 0$. Let Ω be any open connected subset of $\mathbb{R}^n$, having the cone property, and let $\mathcal{A}$ be any π_{k-1} unisolvent subset of Ω. Let h be defined by $\sup_{t\in\Omega} \inf_{a\in\mathcal{A}} |t - a|$, and let ϕ be any solution of $|\cdot|^{2k}\widehat{\phi} = 1/w_\mu$.*

Given any f from $X^{k,w}$, let u denote the minimal norm interpolant to f on $\mathcal{A}$. Then there exist constants h_0, K, and C, all independent of f and h, such that, for all x in Ω, whenever $h < h_0$,

$$|f(x) - u(x)|^2 \leq K \inf_{p\in\pi_{2k-1}} \sup_{0\leq|y|\leq Ch} \{|\phi(y) - p(y)|\} |f|^2_{k,w}.$$

Proof. In view of the remarks made in the last chapter, and especially **4.2.18**, the power function, being built up of the representers, is unaffected by modifications in the basis function by polynomials from π_{2k-1}. Hence, the proof is immediate upon applying **3.3.8** with $|\beta| = 0$. ∎

Theory	Surface Splines	Gaussians	Inverse Multiquadrics	Multiquadrics
w	$\lvert\cdot\rvert^{2s}$, $s < \frac{n}{2}$	$e^{\lvert\cdot\rvert^2}$	$C\lvert\cdot\rvert^{-2k+(n-s)/2}/K_{(n-s)/2}(\lvert\cdot\rvert)$[1]	
k	$k + s - \frac{n}{2} > 0$	$k = 0$	$s > 0$, $k = 0$	$s < 0$, $s \neq 0, -2, -4, \ldots$ $k > -\frac{s}{2}$
μ	$\mu \leq s$	arb.	arb.	arb.
ϕ	$C\lvert\cdot\rvert^{2k+2s-n}\ln\lvert\cdot\rvert$, $2k + 2s - n$ is an even integer, $C\lvert\cdot\rvert^{2k+2s-n}$, otherwise.	$Ce^{-\lvert\cdot\rvert^2/4}$	$\Gamma(\frac{s}{2})(1 + \lvert\cdot\rvert^2)^{-\frac{s}{2}}$	

Table 5.1: Popular theories for multivariate interpolation.

Of course, taking $|\beta| = 0$ is a simplification — it merely allows us to highlight, in a simple manner, the style of error estimate. Since the basis functions are smooth up to order j, where j is the largest integer less than $2k + 2\mu - n$, by virtue of **4.2.5**, the strength of this type of error estimate depends highly on the value of μ. In general, though, the classical theory of polynomial approximation tells us that $|f(x) - u(x)|$ will be $\mathcal{O}(h^\lambda)$, where $\lambda = \min\{j/2, k\}$. We shall compare this with more rudimentary error estimates in the case of surface splines, which we discuss in the next section.

[1] K_ν is a modified Bessel function of the second kind, c.f. Abramowitz and Stegun [1], p.374.

5.1 Surface splines

Our first choice of weight function is $|\cdot|^{2s}$, $s < n/2$, with k subsequently chosen so that $k + s - n/2 > 0$. In this case, the space $X^{k,w}$ is of Beppo Levi type, as described in the fundamental papers of Duchon [10], and Deny and Lions [8], where they would be called $D^{-k}\widetilde{H}^s$ and $BL^k(Y^w)$ respectively. The basis functions which then arise from this space are all of the form $\phi + q$, where $q \in \pi_{2k-1}$, and, for suitable constants $d(k, s, n)$

$$\phi(y) = \begin{cases} d(k,s,n)|y|^{2k+2s-n}\ln|y|, & 2k+2s-n \text{ is an even integer,} \\ d(k,s,n)|y|^{2k+2s-n}, & \text{otherwise.} \end{cases} \tag{5.1}$$

Since we know the form of the basis functions, we will extend our error analysis, but first, some preliminaries.

5.1.1 Lemma *Let $\{p_1, \ldots, p_\lambda\}$ be the cardinal basis functions for π_κ, based on the points $a_1, \ldots, a_\lambda$ in $\mathbb{R}^n$. Let β be any multi-index, and let $q : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be any bivariate function, which, as a function of the first variable, lies in π_κ. Then, for all x in $\mathbb{R}^n$,*

$$\sum_{r,t=1}^{\lambda} (D^\beta p_r)(x)(D^\beta p_t)(x)q(a_t, a_r) - 2\sum_{r=1}^{\lambda}(D^\beta p_r)(x)(D^\beta q)(x, a_r) = -(D^\beta q)(x, x).$$

Proof. We begin by re-writing the expression on the left hand side as

$$\sum_{r=0}^{\lambda} \alpha_r \Big\{ \sum_{t=1}^{\lambda}(D^\beta p_t)(x)q(a_t, a_r) - (D^\beta q)(x, a_r)\Big\} - (D^\beta q)(x, x),$$

where $\alpha_0 = -1$, and $\alpha_r = (D^\beta p_r)(x)$, $r = 1, \ldots, \lambda$. Further manipulation then yields

$$\sum_{r=0}^{\lambda} \alpha_r \Big\{ \Big[D^\beta \Big\{ \sum_{t=1}^{\lambda} p_t(\cdot) q(a_t, a_r) \Big\} \Big](x) - (D^\beta q)(x, a_r) \Big\} - (D^\beta q)(x, x).$$

Since, for all z in $\mathbb{R}^n$, $q(\cdot, z)$ belongs to π_κ, we see that

$$D^\beta \left\{ \sum_{t=1}^{\lambda} p_t(\cdot) q(a_t, z) \right\}(x) = (D^\beta q)(x, z),$$

and so the expression reduces to $-(D^\beta q)(x, x)$. ∎

5.1.2 Lemma *Let E be the real-valued function defined on $\mathbb{R}^n \setminus \{0\}$ by $E(y) = \ln |y|^2$, $y = (y_1, \ldots, y_n)$. Then,*

$$\frac{\partial^\kappa E}{\partial y_j^\kappa} = \sum_{r=\sigma}^{\kappa} c_r y_j^{2r-\kappa} |y|^{-2r}, \qquad \kappa \geq 1,$$

where σ is the largest integer less than $(\kappa + 1)/2$, and the coefficients $\{c_r\}$ depend only on κ.

Proof. As in **4.2.6**, we use a two step inductive process. Suppose the result holds for $\kappa = d$, and consider $\kappa = d + 2$.

$$\begin{aligned}
\frac{\partial^{d+2} E}{\partial y_j^{d+2}} &= \frac{\partial^2}{\partial y_j^2} \left\{ \frac{\partial^d E}{\partial y_j^d} \right\} \\
&= \frac{\partial}{\partial y_j} \left\{ \sum_{r=\sigma}^{d} c_r (2r - d) y_j^{2r-d-1} |y|^{-2r} - \sum_{r=\sigma}^{d} 2r c_r y_j^{2r-d+1} |y|^{-2r-2} \right\} \\
&= \sum_{r=\tau}^{d} c_r (2r - d)(2r - d - 2) y_j^{2r-d-2} |y|^{-2r} + \sum_{r=\sigma}^{d} 2c_r (2r - d) y_j^{2r-d} |y|^{-2r-2} \\
&\quad - \sum_{r=\sigma}^{d} 2r c_r (2r - d + 1) y_j^{2r-d} |y|^{-2r-2} + \sum_{r=\sigma}^{d} 4r(r+1) c_r y_j^{2r-d+2} |y|^{-2r-4},
\end{aligned}$$

where, once more, τ is chosen to be $\sigma + 1$ when d is odd, and $\sigma + 2$ when d is even. Rearranging these terms, we have

$$\sum_{r=\tau}^{d} c_r (2r - d)(2r - d - 2) y_j^{2r-d-2} |y|^{-2r} + \sum_{r=\sigma+1}^{d+1} 2c_{r-1} (2r - d - 2) y_j^{2r-d-2} |y|^{-2r}$$

$$-\sum_{r=\sigma+1}^{d+1} 2(r-1)c_{r-1}(2r-d-1)y_j^{2r-d-2}|y|^{-2r} + \sum_{r=\sigma+2}^{d+2} 4(r-2)(r-1)c_{r-2}y_j^{2r-d-2}|y|^{-2r},$$

which, when collected together yield the desired form

$$\sum_{r=\sigma+1}^{d+2} b_r y_j^{2r-(d+2)}|y|^{-2r}.$$

Observing that our initial cases for odd and even k are

$$\frac{\partial E}{\partial y_j} = y_j|y|^{-2} \qquad \text{and} \qquad \frac{\partial^2 E}{\partial y_j^2} = |y|^{-2} + 2y_j^2|y|^{-4},$$

we complete the proof. ■

5.1.3 Corollary *Let γ be any non-zero multi-index. Then, for all $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$,*

$$(D^\gamma E)(y) = \sum_{j_1=\sigma_1}^{\gamma_1} c_{j_1} \cdots \sum_{j_n=\sigma_n}^{\gamma_n} c_{j_n}|y|^{-2(j_1+\cdots+j_n)} \prod_{r=1}^{n} y_r^{2j_r-\gamma_r},$$

where σ_r is the largest integer less than $(\gamma_r+1)/2$, and the coefficients $\{c_{j_r}\}$ depend only on γ.

We are now in a position to prove the following error estimate for surface splines.

5.1.4 Theorem *Choose positive integers k and n, a real number $s < n/2$, and a multi-index β satisfying $|\beta| < k+s-n/2$. Let Ω be an open subset of $\mathbb{R}^n$, having the cone property. Let $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$ define a π_{k-1}-unisolvent subset of Ω, with separation distance h defined by $h = \sup_{t\in\Omega} \inf_{a\in\mathcal{A}} |t-a|$.*

Given any f from $X^{k,w}$, let u denote the minimal norm interpolant to f on $\mathcal{A}$. Then there exist constants $h_0 > 0$ and $C > 0$, independent of f and h such that,

$$|(D^\beta f)(x) - (D^\beta u)(x)| \le Ch^{k+s-\frac{n}{2}-|\beta|}|f|_{k,w},$$

whenever $h < h_0$.

Proof. We begin by following the proof of **3.3.8**, word for word, until we obtain the estimate

$$\begin{aligned} |(D^\beta f)(x) - (D^\beta u)(x)|^2 \;\; \leq \;\; |f|^2_{k,w} \Bigg| (-1)^{|\beta|}(D^{2\beta}\phi)(0) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta \phi)(x-a_r) \\ + \sum_{r,t=1}^{\ell} \phi(a_t - a_r)(D^\beta p_r)(x)(D^\beta p_t)(x)\Bigg|, \quad (5.2) \end{aligned}$$

and the refinement

$$\begin{aligned} |(D^\beta f)(x) - (D^\beta u)(x)|^2 \;\; \leq \;\; |f|^2_{k,w}\Bigg\{ |(D^{2\beta}\phi)(0)| + 2Kh^{-|\beta|}\max_{0\leq|y|\leq Ch}\{|(D^\beta\phi)(y)|\} \\ +K^2h^{-2|\beta|}\max_{0\leq|y|\leq Ch}\{|\phi(y)|\}\Bigg\}. \quad (5.3) \end{aligned}$$

The two basic choices for ϕ in (5.1) require individual treatment. Considering the case when $2k+2s-n$ is not an even integer, it can be seen from **4.2.6**, that for the specified values of β, $(D^{2\beta}\phi)(0) = 0$. In addition,

$$\max_{0\leq|y|\leq Ch}\{|\phi(y)|\} \sim \mathcal{O}(h^{2k+2s-n}),$$

and

$$\max_{0\leq|y|\leq Ch}\{|(D^\beta\phi)(y)|\} \sim \mathcal{O}(h^{2k+2s-n-|\beta|}).$$

The indicated rate of convergence then follows from (5.3).

However, using such a rudimentary analysis in the case when $2k+2s-n$ is an even integer, might, on casual inspection, introduce logarithmic terms into the error, and so we approach the problem with more care using a technique found in Powell [25]. Define two functions $\Phi(x)$ and $\Psi(x)$ by

$$\Phi(x) = \sum_{r,t=1}^{\ell} (D^\beta p_r)(x)(D^\beta p_t)(x)\phi(a_t - a_r) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta \phi)(x - a_r),$$

and

$$\Psi(x) = \sum_{r,t=1}^{\ell} (D^\beta p_r)(x)(D^\beta p_t)(x)\psi_\alpha(a_t - a_r) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta \psi_\alpha)(x - a_r),$$

where $\psi_\alpha(y) = d|y|^{2k+2s-n}\ln|\alpha y|$, $\alpha \in \mathbb{R}_+$. Setting $E(y) = |y|^{2\lambda}$, with $\lambda = k + s - n/2$, we may write

$$\Psi(x) = \Phi(x) + \ln|\alpha|\left\{ \sum_{r,t=1}^{\ell} (D^\beta p_r)(x)(D^\beta p_t)(x)E(a_t - a_r) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta E)(x - a_r)\right\}. \tag{5.4}$$

Now, to apply **5.1.1** to the bracketed term, we wish to know when $E(y-z)$ is a polynomial of degree at most, $k-1$, in either y or z, since in this case $E(y-z) = E(z-y)$. Now, λ is integral and so, we can find binomial coefficients $\{c_{\lambda,r}\}$, $0 \le r \le \lambda$, such that

$$\begin{aligned} |y - z|^{2\lambda} &= \sum_{r=0}^{\lambda} c_{\lambda,r}|y|^{2(\lambda-r)}\{|z|^2 - 2zy\}^r \\ &= \sum_{r=0}^{\lambda} c_{\lambda,r}|y|^{2(\lambda-r)}\sum_{t=0}^{r} c_{r,t}|z|^{2(r-t)}(-2)^t(zy)^t. \end{aligned}$$

In this form, we can split E into two polynomials — each one a function of y and z, yet each of degree no more than $k-1$ in one of the variables. Firstly, $E(\cdot - z)$ is a polynomial of degree $2\lambda - 2r + t$ for all z and $E(y - \cdot)$ is a polyomial of degree $2r - t$ for all y. Separating the terms for which $2\lambda - 2r + t < k$, we have a polynomial in z, denoted E_1, which lies in π_{k-1}. For the remaining terms, we know that $2\lambda - 2r + t \ge k$, or alternatively, $2r - t \le k + 2s - n$. However, $s < n/2$ by assumption, and so $2r - t < k$. Therefore, the remaining terms can be collected into a polynomial in y, denoted E_2, which also lies in π_{k-1}. We may now apply **5.1.1** to (5.4) after splitting E into E_1 and E_2. Bringing the results together again yields the equation.

$$\Psi(x) = \Phi(x) - (D^{\beta}E)(0)\ln|\alpha|.$$

From **4.2.6**, $(D^{\beta}E)(0) = 0$ for the specified values of β, and so, we can re-write the error estimate in (5.2) as

$$\begin{aligned}|(D^{\beta}f)(x) - (D^{\beta}u)(x)|^2 &\leq \|f\|_{k,w}^2\{|(D^{2\beta}\phi)(0)| + 2Kh^{-|\beta|}\max_{0\leq|y|\leq Ch}\{|(D^{\beta}\psi_{\alpha})(y)|\} \\ &\quad +K^2h^{-2|\beta|}\max_{0\leq|y|\leq Ch}\{|\psi_{\alpha}(y)|\}\}.\end{aligned}$$

Once again, **4.2.6** and **5.1.3** show us that $(D^{2\beta}\phi)(0) = 0$. Moreover, elementary calculus tells us that

$$\begin{aligned}&\max_{0\leq|y|\leq Ch}\{|\psi_{\alpha}(y)|\} = \\ &\qquad \max\left\{(Ch)^{2k+2s-n}\ln|C\alpha h|, \frac{1}{2k+2s-n}\left(\frac{1}{\alpha}\exp\left[\frac{-1}{2k+2s-n}\right]\right)^{2k+2s-n}\right\}.\end{aligned}$$

In either case, setting $\alpha = 1/h$ yields

$$\max_{0\leq|y|\leq Ch}\{|\psi_{\alpha}(y)|\} \sim \mathcal{O}(h^{2k+2s-n}).$$

We now use Leibniz's formula, **4.2.6** and **5.1.2** to write, for any multi-index γ,

$$\begin{aligned}(D^{\gamma}\psi_{\alpha})(y) &= \sum_{\tau\leq\gamma} c_{\tau\gamma}(D^{\gamma-\tau}\{|\cdot|^{2\lambda}\})(y)(D^{\tau}\{\ln|\alpha\cdot|^2\})(y) \\ &= (D^{\gamma}\{|\cdot|^{2\lambda}\})(y)\ln|\alpha y| + \sum_{0<\tau\leq\gamma} c_{\tau\gamma}(D^{\gamma-\tau}\{|\cdot|^{2\lambda}\})(y)(D^{\tau}\{\ln|\alpha\cdot|^2\})(y)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{|\rho|=\lambda} c_{\rho\gamma} y^{2\rho-\gamma} \ln|\alpha y| \\
&+ \sum_{0<\tau\leq\gamma} c_{\tau\gamma}|\alpha|^{-|\tau|} \sum_{|\rho|=\lambda} c_{\rho\gamma\tau} y^{2\rho-\gamma+\tau} \sum_{j_1=\sigma_1}^{\tau_1} c_{j_1} \cdots \sum_{j_n=\sigma_n}^{\tau_n} c_{j_n} |y|^{-2(j_1+\cdots+j_n)} \prod_{r=1}^{n} y_r^{2j_r-\tau_r} \qquad (5.5)\\
&= y^{-\gamma}\Big\{ \sum_{|\rho|=\lambda} c_{\rho\gamma} y^{2\rho} \ln|\alpha y| \\
&+ \sum_{0<\tau\leq\gamma} c_{\tau\gamma}|\alpha|^{-|\tau|} \sum_{|\rho|=\lambda} c_{\rho\gamma\tau} y^{2\rho-\gamma+\tau} \sum_{j_1=\sigma_1}^{\tau_1} c_{j_1} \cdots \sum_{j_n=\sigma_n}^{\tau_n} c_{j_n} |y|^{-2(j_1+\cdots+j_n)} \prod_{r=1}^{n} y_r^{2j_r} \Big\}. \qquad (5.6)
\end{aligned}
$$

Here, the coefficients $c_{\rho\gamma}$, $c_{\tau\gamma}$ and $c_{\rho\tau\gamma}$ are simple products of binomial coefficients from Leibniz's formula, and the factors introduced by differentiation. For the prescribed values of β, these coefficients are positive.

It is clear from (5.5) that, for these values of β, $D^\beta\psi_\alpha$ has an extremum at the origin. Any further extrema then depend on the sign of (5.6). If the combined effect of the coefficients $\{c_{j_r}\}$ makes (5.6) positive, then $D^\beta\psi_\alpha$ may have a second set of extrema when $\ln|\alpha y|$ is negative. This occurs when $0 < |\alpha y| < 1$. If the coefficients have the opposite effect, making (5.6) negative, then we may find extrema when $|\alpha y| > 1$. In either case, there exists a constant A such that extrema of $D^\beta\psi_\alpha$ occur at the origin, and at A/α. Of course, these extrema may lie outside of the ball of radius Ch in which we are interested, but we still include the possibility. Hence, there exist constants b_r such that

$$
\begin{aligned}
\max_{0\leq|y|\leq Ch}\{|(D^\beta\psi_\alpha)(y)|\} = \quad & \max\Big\{ b_1 h^{2\lambda-|\beta|} \ln|b_1\alpha h| + b_2 \sum_{0<\tau\leq\beta} c_{\tau\beta} h^{2\lambda-|\beta|+|\tau|} |\alpha h|^{-|\tau|}, \\
& b_3\alpha^{|\beta|-2\lambda} + b_4 \sum_{0<\tau\leq\beta} c_{\tau\beta}\alpha^{|\beta|-2\lambda-|\tau|}\Big\}.
\end{aligned}
$$

Selecting $\alpha = 1/h$, we again obtain the desired rate, which completes the proof. ∎

Duchon, in [10], also showed that, in these cases, the spaces $X^{k,w}$ are contained in $W^{k+s,2}_{\text{loc}}(\mathbb{R}^n)$, the fractional order Sobolev space. The continuity of elements in $X^{k,w}$ can then be seen from the Sobolev embedding theorems (c.f. Adams [2]), although, to obtain a continuous embedding is highly technical, due to the structure of $W^{k+s,2}_{\text{loc}}(\mathbb{R}^n)$. However, when s is zero, the weight w is the constant function 1, and Parseval's relation then reduces $|\cdot|_{k,w}$ to the Sobolev semi-norm, $|\cdot|_{k,2,\mathbb{R}^n}$. In this refined setting, local L^p error estimates are possible, as detailed fully in [11]. However, we take the opportunity here to continue our error analysis in this direction so that we may discuss several interesting questions that arise. The next lemma relates the $X^{k,w}$ semi-norm to the local Sobolev semi-norm $|\cdot|_{k,2,\Omega}$, for certain domains Ω, using interpolation as an extension operator.

5.1.5 Lemma (Duchon [11]) *Let Ω be an open subset of $\mathbb{R}^n$ with a Lipschitz boundary, and let f belong to $W^{k,2}(\Omega)$. Then there exists a unique element f^Ω in $X^{k,w}$ such that $f^\Omega|_\Omega = f$, and amongst all elements of $X^{k,w}$ satisfying this condition, $|f^\Omega|_{k,w}$ is minimal. Furthermore, there exists a constant $K = K(\Omega)$ such that, for all f in $W^{k,2}(\Omega)$,*

$$|f^\Omega|_{k,w} \leq K|f|_{k,2,\Omega}.$$

Proof. We begin by observing that, since $|\cdot|_{k,2,\Omega}$ is a semi-norm on $W^{k,2}(\Omega)$ with kernel π_{k-1}, $(W^{k,2}(\Omega)/\pi_{k-1}, |\cdot|_{k,2,\Omega})$ is a normed linear space. Furthermore, from Chapter Three of Ciarlet's work [6], we know that, on $W^{k,2}(\Omega)/\pi_{k-1}$, $|\cdot|_{k,2,\Omega}$ is equivalent to the usual quotient norm $\inf_{p\in\pi_{k-1}} \|\cdot+p\|_{k,2,\Omega}$ with respect to which $W^{k,2}(\Omega)/\pi_{k-1}$ is complete. Hence $(W^{k,2}(\Omega)/\pi_{k-1}, |\cdot|_{k,2,\Omega})$ is a Banach space.

Take any f in $W^{k,2}(\Omega)$. The conditions on Ω then allow us to construct many extensions of f which lie in $W^{k,2}(\mathbb{R}^n)$ (c.f. Stein [36]). Since $W^{k,2}(\mathbb{R}^n) \subset X^{k,w}$ in this case, we may also construct a unique function f^Ω which agrees with any of these extensions on Ω, and whose $X^{k,w}$-norm is minimal amongst those of the extensions.

We assert that $\mathcal{F} : f \mapsto |f^\Omega|_{k,w}$ is a norm which makes $W^{k,2}(\Omega)/\pi_{k-1}$ a Banach space. The preservation of elements in the kernel of $|\cdot|_{k,w}$ by the interpolation process, and the nature of the semi-norm confirm that $\mathcal{F}$ is a norm on $W^{k,2}(\Omega)/\pi_{k-1}$. Taking a Cauchy sequence $\{f_j\}$ in $(W^{k,2}(\Omega)/\pi_{k-1}, \mathcal{F})$, we can find, given any $\epsilon > 0$, a threshold N such that

$$\mathcal{F}(f_r - f_t) = |f_r^\Omega - f_t^\Omega|_{k,w} < \epsilon,$$

whenever $r, t \geq N$. Therefore, $\{f_j^\Omega\}$ is a sequence in $X^{k,w}$ which is Cauchy with respect to $|\cdot|_{k,w}$. Our earlier work in **4.1.16** shows us that we can find many elements in $X^{k,w}$ to which this sequence converges, all differing by polynomials from π_{k-1}. Using the aforementioned containment of $X^{k,w}$ in $W^{k,2}_{\text{loc}}(\mathbb{R}^n)$, we can choose any of these limits, knowing that it is unique in the factor space $X^{k,w}/\pi_{k-1}$, and hence unique in $W^{k,2}(\Omega)/\pi_{k-1}$. We will denote the limit in the factor space by g so that we may write,

$$\mathcal{F}(f_j - g) = |f_j^\Omega - (g|_\Omega)^\Omega|_{k,w} = |\{f_j - g|_\Omega\}^\Omega|_{k,w}.$$

The minimal norm property of the interpolation process on Ω then suggests that we can bound the right hand side by the $X^{k,w}$-norm of any element which agrees with $\{f_j - g|_\Omega\}^\Omega$ on Ω. Making the choice $f_j^\Omega - g$, we see that

$$\mathcal{F}(f_j - g) \leq |f_j^\Omega - g|_{k,w},$$

and the convergence as $j \to \infty$ of the latter implies the former. Therefore, $W^{k,2}(\Omega)/\pi_{k-1}$, endowed with the norm $\mathcal{F}$, is a Banach space.

At this point, we have equipped $W^{k,2}(\Omega)/\pi_{k-1}$ with two Banach space topologies, namely $|\cdot|_{k,2,\Omega}$ and $\mathcal{F}$. To complete the proof we observe that, for all f in $W^{k,2}(\Omega)/\pi_{k-1}$,

$$|f|_{k,2,\Omega} = |f^\Omega|_{k,2,\Omega} \leq |f^\Omega|_{k,2,\mathbb{R}^n} = |f^\Omega|_{k,w} = \mathcal{F}(f).$$

The result then follows by a well-known corollary of the open mapping theorem (c.f. Friedman [12], **4.6.3**). ■

5.1.6 Lemma *Let Ω be a measurable subset of $\mathbb{R}^n$. If σ denotes the linear change of variables $x \mapsto t + h(x-a)$, where $h > 0$, and $a, t \in \mathbb{R}^n$, then, for all u in $W^{k,p}(\Omega)$,*

$$|u|_{k,p,\sigma(\Omega)} = h^{\frac{n}{p}-k}|u \circ \sigma|_{k,p,\Omega}.$$

Proof. We have, using the change of variable formula for integration (c.f. Apostol [3]),

$$\begin{aligned} |u|^p_{k,p,\sigma(\Omega)} &= \sum_{|\alpha|=k} c_\alpha \int_{\sigma(\Omega)} |(D^\alpha u)(y)|^p \, dy \\ &= \sum_{|\alpha|=k} c_\alpha \int_\Omega h^n |(D^\alpha u \circ \sigma)(y)|^p \, dy. \end{aligned}$$

Now, if $|\alpha| = k$, then

$$(D^\alpha u)(y) = [D^\alpha(u \circ \sigma \circ \sigma^{-1})](y) = h^{-k}[D^\alpha(u \circ \sigma)](\sigma^{-1}(x)).$$

Hence, for such values of α,

$$(D^\alpha u \circ \sigma)(y) = (D^\alpha u)(\sigma(y)) = h^{-k}[D^\alpha(u \circ \sigma)](y).$$

Finally,

$$\begin{aligned} |u|^p_{k,p,\sigma(\Omega)} &= \sum_{|\alpha|=k} c_\alpha h^n \int_\Omega h^{-kp} |[D^\alpha(u \circ \sigma)](y)|^p \, dy \\ &= h^{n-kp}|u \circ \sigma|^p_{k,p,\Omega}. \qquad ■ \end{aligned}$$

5.1.7 Lemma *Let B be any ball of radius h in $\mathbb{R}^n$, and let f be taken from $W^{k,2}(B)$. Then there exists a unique element f^B in $X^{k,w}$ such that $f^B|_B = f$, and amongst all such* *elements of* $X^{k,w}$, $|f^B|_{k,w}$ *is minimal. Moreover, there exists a constant* C, *independent of* B, *such that, for all* f *in* $W^{k,2}(B)$,

$$|f^B|_{k,w} \leq C|f|_{k,2,B}.$$

Proof. This result is identical to **5.1.5**, except for the fact that C can be taken independent of B. To see this, let B be the ball defined for some a in $\mathbb{R}^n$ by $\{y \in \mathbb{R}^n : |y-a| \leq h\}$, and define σ by $\sigma(y) = h^{-1}(y-a)$. Let B_0 denote the unit ball centred at the origin. Then $\sigma(B) = B_0$. Take any f from $W^{k,2}(B)$. Then $f \circ \sigma^{-1} \in W^{k,2}(B_0)$. Set $F = f \circ \sigma^{-1}$. It is an elementary property of the semi-norm, that F^B, as defined by **5.1.5**, is equal to $f^B \circ \sigma^{-1}$. Also by **5.1.5**, $|f^B \circ \sigma^{-1}|_{k,w} \leq K(B_0)|f^B \circ \sigma^{-1}|_{k,2,B_0}$. Therefore, using **5.1.6**, we obtain $h^{\frac{n}{2}}|f^B|_{k,w} \leq K(B_0)h^{\frac{n}{2}}|f^B|_{k,2,B}$, and taking $C = K(B_0)$ concludes the proof. ∎

5.1.8 Theorem *Let* Ω *be an open subset of* $\mathbb{R}^n$ *with a Lipschitz boundary, let* p *be chosen so that* $2 \leq p \leq \infty$, *and let* β *be any multi-index satisfying* $|\beta| < k - n/2$. *For* $h > 0$, *let* $\mathcal{A}_h$ *be a finite,* π_{k-1}*-unisolvent subset of* Ω *with* $\sup_{t\in\Omega} \inf_{a\in\mathcal{A}_h} |t-a| \leq h$. *For each* f *in* $W^{k,2}(\Omega)$, *let* u_h *denote the minimal norm interpolant to* f *on* $\mathcal{A}_h$, *from* $X^{k,w}$. *Then there exists a constant* $h_0 > 0$ *and a constant* $C > 0$, *both independent of* h, *such that, for all* f *in* $W^{k,2}(\Omega)$,

$$\|D^\beta(f-u_h)\|_{p,\Omega} \leq Ch^{k-|\beta|-\frac{n}{2}+\frac{n}{p}}|f|_{k,2,\Omega},$$

whenever $h < h_0$.

Proof. We proceed as we did in **3.3.8** until $a_1, \ldots, a_\ell$ have been chosen from $B = B(t, MRh)$, and the corresponding Hilbert space theory applied. Next, we define f^Ω in accordance with **5.1.5** and set $g = f^\Omega$. We then note that interpolants to f and g are identical, being made up from the same data, so that, defining $(g-u_h)^B$ in accordance with **5.1.7**, we see that $(g - u_h)^B|_B = (g - u_h)|_B = (f - u_h)|_B$. Then,

$$\begin{aligned} |(D^\beta f)(x) - (D^\beta u_h)(x)|^2 &= |(D^\beta g^B)(x) - (D^\beta u_h^B)(x)|^2 \\ &\leq |\Phi(x)||(g - u_h)^B|_{k,w}, \end{aligned} \tag{5.7}$$

where

$$\Phi(x) = \sum_{r,s=1}^{\ell} (D^\beta p_r)(x)(D^\beta p_s)(x)\phi(a_s - a_r) - 2\sum_{r=1}^{\ell}(D^\beta p_r)(x)(D^\beta \phi)(x - a_r).$$

Now, using **5.1.7** gives

$$|(D^\beta f)(x) - (D^\beta u_h)(x)|^2 \leq C_1|\Phi(x)||g - u_h|^2_{k,2,B},$$

and so,

$$\begin{aligned} \|D^\beta(f - u_h)\|_{p,B} &\leq \sqrt{C_1}|g - u_h|_{k,2,B}\left\{\int_B |\Phi(x)|^{\frac{p}{2}}\, dx\right\}^{\frac{1}{p}} \\ &\leq \sqrt{C_1}|g - u_h|_{k,2,B}\left\{\max_{x\in B}|\Phi(x)|\right\}^{\frac{1}{2}} \text{ meas } (B)^{\frac{1}{p}}. \end{aligned}$$

Using earlier techniques from **5.1.4**, we know that

$$\max_{x\in B}\{|\Phi(x)|\} \sim \mathcal{O}(h^{2k-2|\beta|-n}).$$

Therefore, since the volume of the ball B is $\mathcal{O}(h^n)$, there exists a constant K such that

$$\|D^\beta(f - u_h)\|_{p,B} \leq Kh^{k-|\beta|-\frac{n}{2}+\frac{n}{p}}|g - u_h|_{k,2,B}.$$

Setting $\Omega^* = \bigcup_{t\in T_{Rh}} B(t, MRh)$, let χ_t denote the characteristic function of the ball $B(t, MRh)$, and let λ denote $k - |\beta| - n/2 + n/p$, for convenience. Using the fact that, if

$a \in \mathbb{R}^m$, then for all $p \geq 2$, $\|a\|_p \leq \|a\|_2$, we have

$$\begin{aligned}
\|D^\beta(f-u_h)\|_{p,\Omega} &\leq \|D^\beta(f-u_h)\|_{p,\Omega^*} \\
&\leq \left\{ \sum_{t \in T_{Rh}} \|D^\beta(f-u_h)\|_{p,B}^p \right\}^{\frac{1}{p}} \\
&\leq Kh^\lambda \left\{ \sum_{t \in T_{Rh}} |g-u_h|_{k,2,B}^p \right\}^{\frac{1}{p}} \\
&\leq Kh^\lambda \left\{ \sum_{t \in T_{Rh}} |g-u_h|_{k,2,B}^2 \right\}^{\frac{1}{2}} \\
&\leq Kh^\lambda \left\{ \sum_{t \in T_{Rh}} \int_{\mathbb{R}^n} \chi_t(y) \left[\sum_{|\alpha|=k} c_\alpha |[D^\alpha(g-u_h)](y)|^2 \right] dy \right\}^{\frac{1}{2}} \\
&\leq Kh^\lambda \left\{ |g-u_h|_{k,2,\mathbb{R}^n}^2 \sum_{t \in T_{Rh}} \chi_t \right\}^{\frac{1}{2}},
\end{aligned}$$

and using **3.3.7**, we obtain,

$$\begin{aligned}
\|D^\beta(f-u_h)\|_{p,\Omega} &\leq K\sqrt{M_1}h^\lambda |g-u_h|_{k,2,\mathbb{R}^n} \\
&= K\sqrt{M_1}h^\lambda |g-u_h|_{k,w} \\
&\leq KC_2\sqrt{M_1}h^\lambda |g|_{k,w} \\
&= Ch^{k-|\beta|-\frac{n}{2}+\frac{n}{p}} |f|_{k,2,\Omega}. \qquad \blacksquare
\end{aligned}$$

Our first observation is that we build up the L^p error estimate from the pointwise error. This is in direct contrast with Duchon's work, where he capitalises on the ability of the interpolation process to preserve polynomials in the kernel, π_{k-1}, and the containment of $X^{k,w}$ in $W_{\text{loc}}^{k,2}(\mathbb{R}^n)$. Therefore, our approach, which is more restrictive in demanding that $X^{k,w}$ be a space of smooth functions, only recovers the full strength of Duchon's result in the L^∞ case, or when $|\beta| = 0$, as might be expected. In the other norms, higher derivatives exist in $L^p(\Omega)$, but are beyond the range of our theory, which requires that $|\beta| < k - n/2$.

This problem might lead us to examine different choices of the linear functional γ in the basic theory. Such a choice might, for suitable multi-indices β, be

$$\gamma(f) = \int_{\mathbb{R}^n} \overline{\psi}(y)(D^\beta f)(y)\, dy, \qquad \psi \in C_0^\infty(\mathbb{R}^n),$$

since, by taking the supremum over all ψ whose support lies in some domain Ω, we might recover the L^2 norms for higher derivatives. However, this requires fresh analysis of the boundedness of the linear functional, and the resulting representer, before error estimates can be derived.

We therefore content ourselves with building local L^p error estimates from pointwise errors, but ask the question, when can this be done? The crucial result here, is **5.1.5**, since it allows us, via **5.1.7**, to completely 'localize' the error estimate to the balls which cover Ω, as seen in (5.7). However, in the general cases, the result becomes increasingly technical and non-trivial, as a definition for a 'local' version of the $X^{k,w}$ space is sought.

5.2 Gaussians, multiquadrics, and inverse multiquadrics

We group the remaining examples of popular applications under one heading because the weight functions involved each exhibit a certain property that sets them apart in terms of the error analysis.

From Abramowitz and Stegun [1], p.374, the modified Bessel functions of the second kind K_ν, $\nu \in \mathbb{R}$, are continuous and positive in the complement of the origin, when they depend on the radial distance. Moreover, for all ν in $\mathbb{R}$,

$$K_\nu(|x|) = \begin{cases} \mathcal{O}(e^{-|x|}), & |x| \to \infty \\ \mathcal{O}(|x|^{-\nu}), & |x| \to 0 \end{cases}.$$

This behaviour verifies that, in the cases of multiquadrics and inverse multiquadrics highlighted in Table 5.1, the weight functions satisfy **4.2.1**.

In addition, we can now see the relationship between the three theories, namely the exponential decay of the reciprocal of the weight function at infinity. As Table 5.1 indicates, this allows us to choose μ to be as large as we like, and hence, by **4.2.5**, the basis functions are infinitely differentiable.

As the following theorem shows, this is enough to extend the error analysis of Chapter 3 in a new direction. However, the nature of the convergence is tricky, and so we will attempt to explain the statement of the result just prior to the proof.

5.2.1 Theorem *Let w_μ be any weight function, for which the corresponding choice of integers k and n ensure that the basis function for minimal norm interpolation in $X^{k,w}$ is infinitely differentiable. Let Ω be any open, connected subset of $\mathbb{R}^n$ having the cone property, and let $\mathcal{A} = \{a_r \in \mathbb{R}^n : r = 1, \ldots, m\}$ define a π_k-unisolvent subset of Ω with separation distance h defined by $h = \sup_{t\in\Omega} \inf_{a\in\mathcal{A}} |t-a|$. Then there exist positive constants ϵ_0, M, and R, and, for each $h < \epsilon_0/R$, a set of centres T_{Rh}, such that $\Omega \subset \bigcup_{t\in T_{Rh}} B(t, MRh)$.*

Choose any f in $X^{k,w}$, and let u denote the minimal norm interpolant based on the data $f(a_1), \ldots, f(a_m)$. Fix x in $\mathbb{R}^n$ so that, for some τ in T_{Rh}, $x \in B(\tau, MRh)$, and let $\lambda_{\max}$ be the largest integer for which $\dim \pi_{\lambda_{\max}} \leq m$. If, for some $k \leq \lambda \leq \lambda_{\max}$, N points from $\mathcal{A}$, $N = \dim \pi_\lambda$, can be found in $B(\tau, MRh)$ which form a π_λ-unisolvent set, then there exists a constant C, independent of h, such that, whenever $0 < h < \epsilon_0/R$,

$$|f(x) - u(x)| \leq Ch^{\lambda/2}\sqrt{\langle f, f\rangle}.$$

This theorem highlights one aspect of the convergence process which did not enter into our previous estimate in **3.3.8**, namely the idea that, as more interpolation nodes are introduced into a local neighbourhood of x, the rate of convergence increases accordingly.

Furthermore, in the theorem above, we have, for convenience, neglected to highlight the dependence of N and λ on h. This dependence is the crux of the matter. If interpolation nodes are added to $\mathcal{A}$, then the covering of Ω will change. Indeed, it is possible to shrink the size of the ball in which x is located, without adding any nodes to that ball. Thus, the rate of convergence should always be viewed with care, taking these considerations into mind.

Proof. The proof, initially, follows that of **3.3.8** closely, but we will include the details for the sake of clarity. We begin by taking a π_k-unisolvent set of points $\{v_1, \ldots, v_\ell\}$ from $\mathbb{R}^n$. By **3.3.4**, there exists $\delta > 0$ such that every choice of ℓ-tuple from $B(v_1, \delta) \times \cdots \times B(v_\ell, \delta)$ is π_k-unisolvent. Dilation by a factor δ^{-1} creates a new set of points $\{x_1, \ldots, x_\ell\}$ such that the set $B(x_1, 1) \times \cdots \times B(x_\ell, 1)$ also generates unisolvent ℓ-tuples from $(\mathbb{R}^n)^\ell$. Choose $R > 0$ such that $B(x_r, 1) \subset B(0, R)$, $r = 1, \ldots, \ell$.

Now, applying **3.3.7** to Ω yields two constants, ϵ_0 and M, with which the following properties are associated. Firstly, to each $0 < h < \epsilon_0/R$ there corresponds a set of centres T_{Rh} such that, for all t in T_{Rh}, $B(t, Rh) \subset \Omega$, and, secondly, $\Omega \subset \bigcup_{t \in T_{Rh}} B(t, MRh)$. This completes the first part of the Theorem.

Now suppose x lies in Ω. Then x lies in $B(\tau, MRh)$, for some τ in T_{Rh}. Define $\sigma : B(\tau, MRh) \longrightarrow B(0, MR)$ by $\sigma(y) = h^{-1}(y - t)$, where $y \in B(\tau, MRh)$. Each ball $B(x_r, 1)$ must contain at least one image under σ of a point in $\mathcal{A}$. Hence, we can select $a_1, \ldots, a_\ell$ in $B(t, Rh)$ such that $\sigma(a_r) \in B(x_r, 1)$, $r = 1, \ldots, \ell$. As in **3.3.8**, the points $a_1, \ldots, a_\ell$ will be used in the heart of the Hilbert space theory, with $p_1, \ldots, p_\ell$ defined in accordance with assumption (iii).

Now, by hypothesis, we can find N points from $\mathcal{A}$ which lie in $B(\tau, MRh)$ and are unisolvent with respect to π_λ. Clearly, the above argument shows that $\lambda \geq k$. Moreover, for convenience, and without loss of generality, we can assume that these points can be ordered $a_1, \ldots, a_N$. Therefore, in a similar manner to **3.3.8**, let $L_a : C(B(t, MRh)) \longrightarrow \pi_\lambda$

be the Lagrange interpolation operator associated with $a = \{a_1, \ldots, a_N\}$. As before, the construction allows us to find a constant K, using **3.3.4** and **3.3.5**, which is independent of $a_1, \ldots, a_N$ and which bounds the norm of L_a. We will denote the cardinal basis functions for π_λ by $\tilde{p}_1, \ldots, \tilde{p}_N$. Then, from **3.3.2**, $L_a f = \sum_{r=1}^N f(a_r)\tilde{p}_r$, and the boundedness of the norm of L_a means that $\sup_{x \in B(\tau, MRh)} \sum_{r=1}^N |\tilde{p}_r(x)| \le K$. Choose coefficients $\beta_1, \ldots, \beta_N$ such that

$$\sum_{r=1}^N \beta_r p(a_r) = -p(x),$$

for all p in π_λ. Applying this set of equations to each $\tilde{p}_r$, $r = 1, \ldots, N$ shows that $\beta_r = -\tilde{p}_r(x)$, $r = 1, \ldots, N$. Since $\pi_k \subset \pi_\lambda$, we can then construct an error estimate using **3.2.6**, and the associated theory presented there. Therefore, since ϕ is real-valued,

$$|f(x) - u(x)| \le \sqrt{\langle f, f\rangle} \left| \phi(0) + \sum_{r,s=1}^N \beta_r \beta_s \phi(a_s - a_r) + 2\sum_{r=1}^N \beta_r \phi(x - a_r) \right|^{\frac{1}{2}}.$$

The differentiability of ϕ then allows us to expand each term using a Taylor's series to obtain,

$$\begin{aligned} |f(x) - u(x)| \le\ & \sqrt{\langle f, f\rangle} \\ & \left| \phi(0) + 2\sum_{r=1}^N \beta_r \left\{ \sum_{|\alpha|<\lambda} \frac{(D^\alpha \phi)(0)}{\alpha!}(x - a_r)^\alpha + \psi_\lambda(x - a_r) \right\} \right. \\ & \left. + \sum_{r,s=1}^N \beta_r \beta_s \left\{ \sum_{|\alpha|<\lambda} \frac{(D^\alpha \phi)(0)}{\alpha!}(a_s - a_r)^\alpha + \psi_\lambda(a_s - a_r) \right\} \right|^{\frac{1}{2}}, \end{aligned}$$

where $|\psi_\lambda(y)| \le C|y|^\lambda$, for some constant C, dependent on Ω. The conditions on $\beta_1, \ldots, \beta_N$ now allow us to replace the terms which are polynomials in π_λ so that,

$$|f(x) - u(x)| \le \sqrt{\langle f, f\rangle} \Big\{ \phi(0) - 2\phi(0) + \phi(0)$$

$$
\begin{aligned}
& +2C\sum_{r=1}^{N}|\widetilde{p}_r(x)||x-a_r|^{\lambda}+C\sum_{r,s=1}^{N}|\widetilde{p}_r(x)||\widetilde{p}_s(x)||a_s-a_r|^{\lambda}\Big\}^{\frac{1}{2}} \\
\leq\ & \{2CK+K^2\}h^{\frac{\lambda}{2}}\sqrt{\langle f,f\rangle},
\end{aligned}
$$

which completes the proof. ■

This concludes our analysis of some of the more popular theories in radial basis function interpolation. As we mentioned in the introduction, these examples provided the motivation for this research, yet as Chapter 4 shows, radial basis functions correspond to the specific case of the theory in which the weight function exhibits radial symmetry. The constructive nature of our approach, centred on **4.2.1**, hopefully shows that many more examples exist.

In pursuing this approach to multivariate interpolation, we hope we have made some small steps towards balancing the theory. Without doubt, one of the disadvantages of the Hilbert space theory is that you must have prior knowledge of the quadratic functional $\langle\cdot,\cdot\rangle$. The resulting analysis may then yield an unwieldy form of basis function, not suited to computational techniques and error estimation. Therefore, from a certain point of view, the second approach, whose foundations lie with the choice of conditionally positive definite function, is a more attractive prospect. However, the intention of this work is to complement, not compete with existing theories. We hope that the work clearly shows the evolutionary steps one might need to perform in order to pose the kind of interpolation problems under consideration, and in doing so, we believe a subtle, yet fundamental difference is maintained with the alternative approach.

Bibliography

[1] Abramowitz, M. and I.A. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington (1964).

[2] Adams, R.A. *Sobolev Spaces*, Academic Press, Boston, Ma. (1978).

[3] Apostol, T. *Mathematical Analysis* 2nd ed., Addison-Wesley, Reading, Ma. (1974).

[4] Atteia, M. *Fonctions spline et noyaux reproduisants d'Aronszajn-Bergman*, R.I.R.O. 4, **No. 3** (1970), pp31-43.

[5] Ball, K., N. Sivakumar, and J.D. Ward, *On the sensitivity of radial basis interpolation to minimal separation distance*, Constructive Approximation, 8 (1992), pp401–426.

[6] Ciarlet, P.G. *The Finite Element Method for Elliptic Problems*, North Holland (1978).

[7] Davis, P.J. *Interpolation and Approximation*, Dover Publications, New York (1975).

[8] Deny, J. and J.L. Lions, *Les espaces du type de Beppo Levi*, Ann. Inst. Fourier **5** (1954), pp305–370.

[9] Ditzian, Z. *Multivariate Bernstein and Markov inequalities*, Journal of Approximation Theory **70**, Academic Press (1992), pp273-283.

[10] Duchon, J. *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller (eds.), Springer-Verlag, Oberwolfach (1976), pp85–100.

[11] Duchon, J. *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, R.A.I.R.O. Analyse Numerique **12**, **No. 4** (1978), pp325-334.

[12] Friedman, A. *Foundations of Modern Analysis*, Dover (1970).

[13] Gel'fand, I.M. and N.Ya. Vilenkin, *Generalized Functions* **IV**, Academic Press, New York (1964).

[14] Golomb, M. and H.F. Weinberger, *Optimal approximation and error bounds*, in *On Numerical Approximation*, ed. R.E. Langer, University of Wisconsin Press (Madison) (1959), pp117–190.

[15] Harder, R.L. and R.N. Desmarais, *Interpolation using surface splines*, J. Aircraft **9** (1972), pp189-191.

[16] Hardy, R.L. *Multiquadric equations of topography and other irregular surfaces*, J. Geophys. Res. **76** (1971), pp1905–1915.

[17] Holladay, J.C. *A smoothest curve approximation*, Math. of Comp. **11** (1957), pp233-243.

[18] Hörmander, L. *Linear Partial Differential Operators*, Die Grundlehren der Mathematischen Wissenschaften **116**, Springer-Verlag, Berlin (1969).

[19] Hörmander, L. *On the division of distributions by polynomials*, Arkiv For Matematik **3**, **No. 53** (1958), pp555–568.

[20] Madych, W.R., and S.A. Nelson, *Multivariate interpolation, a variational theory*, preprint (1983).

[21] Madych, W.R., and S.A. Nelson, *Multivariate interpolation and conditionally positive definite functions*, Approximation Theory and its Applications **4**, **No. 4** (1988), pp77-89.

[22] Madych, W.R., and S.A. Nelson, *Multivariate interpolation and conditionally positive definite functions II*, Mathematics of Computation **54**, **No. 189**, AMS (1990), pp211–230.

[23] Micchelli, C.A. *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constructive Approximation **2**, Springer-Verlag (1986), pp11-22.

[24] Narcowich, F.J. and J.D. Ward, *Norms of inverses and condition numbers for matrices associated with scattered data*, Journal of Approximation Theory **64**, Academic Press (1991), pp69–94.

[25] Powell, M.J.D. *The uniform convergence of thin-plate spline interpolation in two dimensions*, Numerische Mathematik **68**, **No. 1**, Springer-Verlag (1994), pp107-128.

[26] Powell, M.J.D. *Approximation Theory and Methods*, Cambridge University Press (1981).

[27] Rudin, W. *Functional Analysis*, International Series in Pure and Applied Mathematics, 2nd ed. McGraw-Hill, New York (1973).

[28] Rudin, W. *Real and Complex Analysis*, McGraw-Hill Series in Higher Mathematics, McGraw-Hill, New York (1966).

[29] Schaback, R. *Comparison of radial basis function interpolants*, in *Multivariate Approximation — from CAGD to Wavelets*, K. Jetter and F. Utreras (eds.), World Scientific (1993), pp293–305.

[30] Schaback, R. *Error estimates and condition numbers for radial basis function interpolation*, Advances in Computational Mathematics **3**, J.C. Baltzer AG (1995), pp251–264.

[31] Schoenberg, I.J. *Metric spaces and completely monotone functions*, Ann. of Math **39** (1938), pp811-841.

[32] Schumaker, L.L. *Spline functions: basic theory*, Krieger Publishing (1993).

[33] Schwartz, L. Theorie des Distributions, Tome I, Actualites Scientifiques et Industrielles **1091**, Hermann, Paris (1950).

[34] Schwartz, L. Theorie des Distributions, Tome II, Actualites Scientifiques et Industrielles **1122**, Hermann, Paris (1951).

[35] Shapiro, H.S. *Topics in Approximation Theory*, Lecture Notes in Mathematics **187**, Springer-Verlag (1971).

[36] Stein, E. *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press (1970).

[37] Stein, E. and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, N.J. (1971).

[38] Sun, X. *Norm estimates for inverses of Euclidean distance matrices*, Journal of Approximation Theory **70**, Academic Press (1992), pp339–347.

[39] Wu, Z. and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA Journal of Numerical Analysis **13** (1993), pp13–27.