Mesh-free Radial Basis Function Methods for Advection-Dominated Diffusion Problems

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

David Patrick Hunt
Department of Mathematics
University of Leicester

October 2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Mesh-free Radial Basis Function Methods for Advection-Dominated Diffusion Problems

David Patrick Hunt
University of Leicester

This thesis is concerned with the numerical solution of advection-dominated diffusion problems. There are essentially two key aspects to this work: the derivation of an *a priori* error estimate for a semi-Lagrangian mesh-free method using radial basis function interpolation to numerically approximate the first-order linear transport problem; and the design and testing of a semi-Lagrangian mesh-less method to numerically solve the full parabolic advection-diffusion problem, using radial basis function Hermite interpolation.

We begin by establishing the theory of radial basis function interpolation, including new results for the stability of interpolation via the class of radial basis functions known as polyharmonic splines, as well as error estimates for interpolation by the same class of function. These results provide us with the necessary tools to prove the *a priori* error estimate for the semi-Lagrangian advection scheme, given certain assumptions on the smoothness of the solution. We then validate both the scheme and the analysis with a series of numerical experiments.

By introducing the concept of Hermite interpolation, we develop and implement a new semi-Lagrangian method for the numerical approximation of advection-dominated diffusion problems, which is validated through two numerical experiments.

In memory of Professor Will Light

Acknowledgements

I have met some jolly nice people over the years, and I'd like single some of them out for thanks to show my appreciation of the impact they have had on my life.

Firstly, Jeremy Levesley and Paul Houston, for their joint supervision and guidance, and for keeping me focussed, albeit sometimes on twelve different things at once. Their expertise and experience is the foundation on which this thesis is built. I feel now is the time to reveal that I sometimes referred to them as Mum and Dad, but I refuse to clarify which one was which. I'd also like to thank the Department of Mathematics and its staff (both academic and support), the University of Leicester and the EPSRC for supporting my PhD via the Doctoral Training Grant scheme.

My colleague and, more importantly, my friend Rob Brownlee, for patiently answering hundreds of approximation-theory-based questions over the years, I wouldn't have been able to finish this PhD without his wisdom, insight, and capacity for beer.

My family, in particular my mother and father, for both their fiscal and unflappable emotional support, not just during my postgraduate days, but throughout my educational life (I suspect they don't really believe I've finally ceased to be a student).

The friends with whom I shared the house at Heather Rd — Lisa, Emma and Andrew — thank you for putting up with my various quirks and foibles, and for letting me make you countless cups of tea.

My wide circle of friends — including Amaro, Andy and Jo to name precisely three — thank you for the nights out and the seemingly constant laughter (also thanks for all paying your taxes in recent years, those grant cheques were much appreciated). Special thanks

to Ade & Sarah and Rob & Caz for their spare beds and delicious meals during the final months of writing-up, and their friendships throughout.

And to Stacey, for a friendship and love unbounded by distance and time-zones, who has taught me more about patience and understanding than I knew there was to learn. (And for sending me Oreo Cookies when I needed them most.)

Lastly, I would like to pay tribute to my original supervisor, Will Light, who died suddenly and unexpectedly in December 2002. Will was an inspiration even before I met him; his written introduction to the Approximation Theory course that I had opted to study, apparently knocked together off the top of his head whilst waiting for a flight in Christchurch International Airport, demonstrated both his supreme knowledge of the subject and the ease at which he could communicate the complicated ideas with clarity and humour. He still provides a great source of motivation and aspiration for me, not only intellectually, but in the way I saw him interact with other people, both professionally and personally.

The day after Will passed away, I wrote a short tribute which closed with the following paragraph.

> I know I'll remember him as a large inspiration to my continuing study of approximation theory, and people all over the world will remember him as uncovering as many interesting questions as he answered. I've not met many people who could genuinely say they've left a mark on the world, have contributed significantly to our understanding of complex ideas, and inspired many people to attain the same levels of understanding, but Will was definitely such a person. And I will be forever proud that I can say, I worked with Will Light.

David Hunt

Leicester, UK

October 2005

Contents

Acknowledgements iii

1 Introduction 1

1.1 The study of advection-diffusion problems 1

1.2 The structure of the thesis 4

1.3 Preliminaries 6

1.3.1 Numbers and multi-indexes 6

1.3.2 Differential operators and continuous functions 7

1.3.3 Inner products and norms 7

1.3.4 Function spaces, distributions, dual spaces and Fourier transforms . 9

2 Interpolation via Radial Basis Functions 12

2.1 Fundamental properties of an interpolant 12

2.2 Spaces of admissible functions 14

2.2.1 Strictly positive-definite functions 15

2.2.2 Conditionally strictly positive-definite functions 17

2.2.3 Characterisation of conditionally strictly positive-definite functions . 20

2.2.4 Radial basis functions 21

2.3 A variational approach to the interpolation problem 24

2.3.1 Reproducing kernel Hilbert spaces 24

2.3.2 Interpolants of minimal norm 25

2.4 Beppo Levi spaces 32
2.5 The Lagrange representation of the interpolant 35
2.6 Error estimates 36
2.6.1 Pointwise error estimates 37
2.6.2 L_p-error estimates for $BL^k(\Omega)$ 38
2.7 Local interpolation 50
2.7.1 Condition numbers and stability of polyharmonic spline local interpolation 52
2.7.2 Proof of stability of the thin-plate spline local interpolant 54
2.7.3 Proof of stability of the general polyharmonic spline local interpolant 62
2.7.4 Error estimates for local interpolation with polyharmonic splines . . 65
2.8 Hermite-Birkhoff interpolation 67
2.8.1 Non-symmetric and symmetric approaches 69
2.8.2 Application of Hermite-Birkhoff interpolation to solving linear partial differential equation problems 71

3 Linear Advection Problems **75**
3.1 Model problem and discretisation 75
3.1.1 Temporal discretisation 77
3.1.2 Spatial discretisation 78
3.2 Main results 80
3.3 Analysis of the semi-discrete scheme 81
3.4 Analysis of the radial basis function semi-Lagrangian numerical method . . 83

4 Linear Advection Numerical Experiments **87**
4.1 Experiment 1: Rotating cone problem 88
4.2 Experiment 2: Rotating cylinder problem 94
4.3 Experiment 3: Gaussian cone in a vortex 94

5 Advection-Diffusion Problems **98**

5.1 Temporal discretisation 100

5.2 Spatial discretisation 101

5.3 Analysis of the semi-discrete scheme 103

5.4 Remarks on the analysis of the full discretisation scheme 106

5.5 Numerical experiments for advection-diffusion problems 107

5.5.1 Experiment 1: Shear flow 107

5.5.2 Experiment 2: Rotating cylinder with small diffusion coefficient . . . 117

6 Conclusions and Future Work **120**

6.1 Introduction 120

6.2 Conclusions 120

6.3 Future work 122

6.3.1 Analysis of the linear advection scheme 122

6.3.2 The advection-diffusion scheme 123

A Nearest Neighbour Algorithms **126**

A.0.1 Fast exhaustive nearest neighbour algorithm 127

A.0.2 Comparison of fast full-search equivalent algorithm with brute force search 128

A.0.3 Alternative nearest neighbour algorithms 129

Bibliography **131**

Chapter 1

Introduction

This thesis is primarily concerned with the numerical solution of *advection* and *advection-diffusion* problems, which are broadly defined as models of the transport of a scalar quantity in a vector field. In the case of purely advective problems, the quantity is conserved along the *trajectories* (also known as the *characteristics*) of particles within the system; a physical example of this behaviour would be the transport of pollution in a river: the motion of the water carries the polluted water downstream. When diffusion occurs as well as advection, the quantity initially assigned to one particle "spreads" over a region as time moves forwards. With this in mind, it should be clear that the mathematical modelling and numerical approximation of purely advective or advection-dominated diffusion problems are of fundamental importance within a wide range of areas in applied mathematics; most notably, meteorology, oil reservoir simulation, aerodynamics and physiology, for example.

1.1 The study of advection-diffusion problems

Mathematically, we can state the advection-diffusion problem as a parabolic differential equation of the form

$$u_t + a \cdot \nabla u - \varepsilon \Delta u = f, \qquad (x,t) \in \Omega \times (0,T], \tag{1.1.1a}$$

$$u = g, \qquad (x,t) \in \partial\Omega \times (0,T], \tag{1.1.1b}$$

$$u(x,0) = u_0(x), \qquad x \in \Omega, \tag{1.1.1c}$$

where

- $a : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is the *velocity field*, which we assume is divergence-free (that is, $\nabla \cdot a = 0$);
- $f : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is the *forcing function*;
- $\varepsilon > 0$ is the *coefficient of diffusion*;
- $g : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ is a known function providing the value of the solution on $\partial\Omega$, the boundary of Ω, for all $t \in (0,T]$;
- the solution $u : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is, for example, the temperature of a fluid or the concentration of a pollutant, whose value is known for all $x \in \Omega \subset \mathbb{R}^d$ at time $t = 0$, represented by the *initial function* $u_0 : \mathbb{R}^d \to \mathbb{R}$.

When $\varepsilon = 0$ the problem reduces to one of a purely advective nature, sometimes referred to as *passive advection*, in which case (1.1.1a) becomes hyperbolic.

Typically, in many areas of practical importance, advection dominates diffusion, in the sense that $\|a\|/\varepsilon \gg 1$ for some appropriate norm of a. In this situation, it is well known that the application of many traditional numerical methods, developed for strictly parabolic or diffusion-dominated processes, often behave very poorly when applied to these types of problems. Indeed, among the common difficulties is a trade-off between excessive numerical diffusion and non-physical oscillations. However, these effects can be greatly reduced by employing numerical methods which directly exploit the hyperbolic character of the underlying partial differential equations. Indeed, one such class which we shall be concerned with — *semi-Lagrangian methods* — seek to combine the method of characteristics with a suitable algorithm for interpolating/projecting the solution onto a fixed (Eulerian) spatial mesh. Since their initial development in the early 1980s (see Robert (1981), for

example) they have been extensively studied; indeed, semi-Lagrangian finite difference and finite element methods have been studied by, e.g., Bercovier & Pironneau (1982), Pironneau (1981/82), Douglas & Russell (1982), Morton, Priestley & Süli (1988), Morton & Süli (1995), Houston & Süli (2001); see also the references cited therein. For the application of spectral methods within the context of semi-Lagrangian methods, we refer the reader to the article by Süli & Ware (1991).

The starting-point for the work contained in this thesis is a class of semi-Lagrangian methods based on employing a combination of the method of characteristics with *radial basis function* interpolation for the numerical approximation of the unsteady linear transport equation. This class of schemes was first introduced in the series of articles by Behrens & Iske (2002) and Behrens, Iske & Käser (2003); also see Iske (2003*b*) for the extension to advection-dominated diffusion problems.

The idea behind the method is quite straightforward: the computational domain is first discretised by a set of scattered points Ξ, referred to as a *point cloud*, since in contrast to finite difference-based methods, Ξ may be completely unstructured. Then, at each time-step, every point within Ξ is first traced backwards in time based on solving the system of ordinary differential equations which determine the characteristics of the underlying partial differential equation. The value of the numerical solution at this upstream point is then determined by *locally* interpolating the values of the computed solution within a neighbourhood of where the point has landed at the previous time level. Due to the hyperbolic nature of the problem, we may then assign a value to the particle at the *current time level* based on the computed value at the previous time level.

Using only a small number of the immediate neighbours of the upstream point over which to interpolate the numerical solution, ensures that the interpolation is computationally inexpensive since the resulting interpolation matrix is relatively small. Moreover, since this process is completely local, it can easily be parallelised, since each point within the point cloud Ξ can be treated independently.

The major advantage to using radial basis functions to solve the interpolation problem is that they can easily treat problems in high space dimensions, since, being based purely on the distance from each point in a given point set, radial basis function interpolation is essentially a dimensionally-independent process.

A natural progression from the study of advection problems is, of course, the solution of advection-diffusion problems, which are the focus of the penultimate chapter in this thesis; the strategy of deriving a numerical solution to this family of problems is similar to that described above, with the exception of it being not possible to straightforwardly localise the interpolation problem due to the global influence of diffusion. These concepts will be explained in greater depth as and when they occur.

1.2 The structure of the thesis

The thesis is structured as follows. After the preliminaries of notational conventions which close this chapter, Chapter 2 provides background theory on interpolation and function-reconstruction using the particular family of functions known as radial basis functions. In Sections 2.2 and 2.3 we see how linear combinations of shifts of these functions are utilised to construct a function which agrees with all values from given set of data, and thus forms an approximation to a function that corresponds to the given values.

The questions of which functions can be theoretically reconstructed in this manner, and how accurate the approximation is (termed *error estimation*), are addressed in Sections 2.4 and 2.6, respectively, where we concentrate on one particular class of radial basis function: *polyharmonic splines.* We then introduce the concept of local interpolation, describing how it differs from global interpolation methods, proving the stability of the interpolation operator for polyharmonic splines, which is undertaken in Section 2.7.1 using their property of invariance under scalings, followed by error estimates for local interpolation in Section 2.7.4.

Finally in this chapter, we broaden the discussion to consider interpolation problems wherein the data is not provided simply as function values but as the action of given functionals on a function. This is so-called *Hermite-Birkhoff interpolation*, for which in

Section 2.8 we describe some established results from the literature concerning its accuracy, as well as applications that will prove pertinent to the advection-diffusion problem.

In Chapter 3 we introduce the proposed radial basis function semi-Lagrangian method for the numerical approximation of the linear transport equation. Our main theoretical results are the *a priori* error bounds presented in Section 3.2. The proof of the bound for the semi-discrete (continuous in space) scheme is carried out in Section 3.3; in Section 3.4 the convergence proof for the full radial basis function semi-Lagrangian scheme is undertaken.

The analysis closely follows that presented in the article by Falcone & Ferretti (1998), where a class of mesh-free partition of unity methods were studied. The key difference in our work is that we consider a more general class of mesh-free methods based on radial basis function interpolation. In this case, the main technical difficulty lies in controlling the interpolation error at a given time level, evaluated at an interpolation point, with the corresponding quantity evaluated at the upstream point on the previous time level. To this end, we employ the Lagrangian representation of the underlying radial basis function interpolation operator which we describe in Section 2.5, together with approximation results based on employing a Taylor-series-argument that we prove in Section 2.7.4..

Although radial basis functions have been used to good effect elsewhere in numerically-solving advection-diffusion-type partial differential equations — see the work of Zerroukat, Djidjeli & Charafi (2000) for explicit and implicit mesh-less methods using radial basis functions; Zerroukat, Power & Chen (1998) for radial basis function collocation methods for solving the heat equation; Boztosun, Charafi, Zerroukat & Djidjeli (2002) and Ingber, Chen & Tanski (2004) for a description of a Crank-Nicholson (θ-weighted) radial basis function method; and Lorentz, Narcowich & Ward (2003) for related work on Eulerian-based radial basis function interpolation methods for linear hyperbolic problems — to the best of our knowledge this is the first proof of convergence of this class of numerical methods.

The numerical performance of the method is demonstrated in Chapter 4, for problems with (Sections 4.1 and 4.2) and without (Section 4.3) known analytical solutions; the former problems also serve to confirm the analysis of the previous chapter.

Chapter 5 introduces a method for numerically solving advection-diffusion problems using radial basis functions to solve Hermite-Birkhoff interpolation problems at each time-step. The effectiveness of this scheme is demonstrated in numerical experiments presented in Section 5.5.

Finally, in Chapter 6 we summarise the work presented in this thesis, draw some conclusions and suggest possible avenues for future work.

1.3 Preliminaries

In this section we conveniently gather some notational conventions and definitions that will be used throughout the rest of this thesis. Where beneficial, we shall remind the reader of these definitions and conventions as and when they subsequently occur.

From the outset, to avoid confusion we state here that when we write, for example,

$$f(x) = g(x) \qquad (x \in R^d),$$

we mean that the stated property is true *for all* $x \in \mathbb{R}^d$.

1.3.1 Numbers and multi-indexes

Let $\mathbb{N}$ denote the set of natural numbers, $\mathbb{Z}$ the set of integers, and $\mathbb{R}$ the set of real numbers. A subscript $+$ symbol on the latter two sets indicates the subset of non-negative numbers; e.g. $\mathbb{Z}_+$ indicates the set of non-negative integers, and is equivalent to $\mathbb{N} \cup \{0\}$.

$\mathbb{Z}_+^d$ denotes all multi-indexes with non-negative entries, and if $\alpha = (\alpha_1, \ldots, \alpha_d)^T \in \mathbb{Z}_+^d$, the *order of* α is defined to be $|\alpha| = \alpha_1 + \cdots + \alpha_d$, and $\alpha! = \alpha_1! \cdots \alpha_d!$ (which extends the convention that if $\alpha = (0, \ldots, 0)$, then $\alpha! = 1$). For $x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ and $\alpha = (\alpha_1, \ldots, \alpha_d)^T \in \mathbb{Z}_+^d$, $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, and x^α is called a *monomial.*

1.3.2 Differential operators and continuous functions

The differential operator D^α is

$$D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d},$$

where $\partial/\partial x_i$ denotes the partial derivative with respect to x_i, for $i = 1, \ldots, d$. D^α is equal to the identity operator for $\alpha = 0$.

For $m \in \mathbb{Z}_+ \cup \{\infty\}$, $C^m(\Omega)$ denotes the set of all continuous real-valued functions f defined on Ω such that $D^\alpha f$ is continuous on Ω for all $|\alpha| \leq m$. When $m = 0$ we simply write $C(\Omega)$ rather than $C^0(\Omega)$. $C^\infty(\Omega)$ denotes functions f for which $D^\alpha f$ is continuous over Ω for all $m = 0, 1, 2, \ldots$.

1.3.3 Inner products and norms

Semi-inner products are always denoted by $(\cdot, \cdot)$ whereas inner products will use angle brackets $\langle \cdot, \cdot \rangle$, with an appropriate subscript specifying over which space the (semi-)inner product is defined. Semi-norms that arise from semi-inner products are denoted by $|\cdot|$ with the corresponding subscript — $|\cdot|_F$ for the semi-norm over the function space F, for example — whereas full norms are denoted $\|\cdot\|$ with a subscript. In particular, the norms for the well-known Lebesgue spaces of measurable functions of finite norm, which for an arbitrary domain $\Omega \subset \mathbb{R}^d$ are denoted $L_p(\Omega)$, $1 \leq p \leq \infty$, are defined as follows:

$$\|f\|_{L_p(\Omega)} := \left(\int_\Omega |f(x)|^p \, \mathrm{d}x\right)^{1/p} \tag{1.3.1}$$

when $1 \leq p < \infty$, and

$$\|f\|_{L_\infty(\Omega)} := \operatorname{ess\,sup}_{x \in \Omega} |f(x)| \tag{1.3.2}$$

for the case $p = \infty$.

It has become standard in the literature (or at least, the literature concerned with radial basis function methods) to denote the Euclidean norm $\|\cdot\|_2$ by $|\cdot|$; when the argument is a real number the same notation denotes the modulus of the number, and as it is always clear from the context which metric is being deployed we shall defer to this tradition, although we may re-introduce the former notation where extra clarity can be gained from its use. In general, for any vector $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$,

$$\|a\|_p := \left(\sum_{i=1}^{d} |a_i|^p \right)^{1/p}$$

for $1 \leq p < \infty$, and

$$\|a\|_\infty := \max_{1 \leq i \leq d} |a_i|.$$

Given an $n \times n$ real matrix A, the matrix p-norm is defined for $1 \leq p \leq \infty$ by

$$\|A\|_p := \max_{\|x\|_p = 1} \|Ax\|_p,$$

where $x \in \mathbb{R}^d$.

For a real-valued function f defined on a finite set of n points $\{x_1, \ldots, x_n\} = \Xi \subset \mathbb{R}^d$, the *discrete infinity norm* is

$$\|f\|_{\infty,\Xi} := \max_{x \in \Xi} |f(x)|.$$

Let $0 = t^0 < t^1 < \ldots < t^{n-1} < t^n = T$ be a subdivision of the interval $[0, T]$. For any function space F,

$$\|f\|_{\ell^\infty([0,T];F)} := \max_{0 \leq i \leq n} \|f(t^i)\|_F, \tag{1.3.3}$$

whenever f can be interpreted as a function of t.

For $f \in C^1(\mathbb{R}^d) \times C((0,T])$,

$$\|f\|_{C((0,T];C^1(\mathbb{R}^d))} := \sup_{t\in(0,T]} \left[\max_{|\alpha|\le 1} \sup_{x\in\mathbb{R}^d} |D^\alpha f(x,t)| \right]. \tag{1.3.4}$$

1.3.4 Function spaces, distributions, dual spaces and Fourier transforms

We denote by Π_m^d the space of real-valued d-variate polynomials with real coefficients of total degree less than or equal to m. Hence $\Pi_0^d = \mathbb{R}$, Π_1^d is the space of all linear polynomials in d variables, and so on. By convention the zero polynomial (that is, the constant polynomial that is zero everywhere) has degree $-\infty$, so $\Pi_{-k}^d = \{0\}$ for all $k > 0$ and $d \ge 1$.

The classical Sobolev space $W_p^k(\Omega)$ is defined for $k, p \in \mathbb{Z}_+$ as the space

$$W_p^k(\Omega) := \{v : D^\alpha v \in L_p(\Omega), 0 \le |\alpha| \le k\}, \tag{1.3.5}$$

with norms

$$\|v\|_{W_p^k(\Omega)} := \left(\sum_{|\alpha|\le k} \|D^\alpha v\|_{L_p(\Omega)}^p \right)^{1/p}$$

when $1 \le p < \infty$, and

$$\|v\|_{W_\infty^k(\Omega)} := \max_{|\alpha|\le k} \|D^\alpha v\|_{L_\infty(\Omega)}$$

for $p = \infty$.

It is common in the literature to write $H^k(\Omega)$ for $W_2^k(\Omega)$, but for simplicity in this thesis we shall avoid introducing this extra notation.

The Schwartz space of rapidly decreasing functions consists of functions in $C^\infty(\mathbb{R}^d)$ for which

$$\sup_{|\alpha|\le N} \sup_{x\in\mathbb{R}^d} (1+|x|^2)^N |(D^\alpha f)(x)| < \infty \qquad (N \in \mathbb{Z}_+). \tag{1.3.6}$$

$\mathscr{S}$ denotes the vector space of these functions with a topology given by the countable collection of norms (1.3.6), and $\mathscr{S}'$ denotes the space of linear functionals on $\mathscr{S}$ that are continuous with respect to this topology, which is termed the space of *(tempered) distributions.* For a distribution $\Lambda \in \mathscr{S}'$ and a so-called *test function* $\phi \in \mathscr{S}$, $[\Lambda, \phi]$ denotes the action of the distribution on the test function. For $f \in C^\infty(\mathbb{R}^d)$, $f\Lambda$ is a distribution with an action defined by

$$[f\Lambda, \phi] = [\Lambda, f\phi] \qquad (\phi \in \mathscr{S}).$$

We extend the notion of differentiation to encompass distributional derivatives by defining the distribution $D^\alpha \Lambda$ via

$$[D^\alpha \Lambda, \phi] = [\Lambda, (-1)^{|\alpha|} D^\alpha \phi] \qquad (\phi \in \mathscr{S}).$$

For any space of functions F, we denote the *dual space* of distributions over F by F^*.

The *Fourier transform* of a function $f \in L_1(\mathbb{R}^d)$ is denoted $\widehat{f}$ and given by

$$\widehat{f}(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(y) e^{-ix^T y} \, dy,$$

which is continuous and vanishes at infinity. Moreover, the Fourier transform defines a continuous, linear, one-to-one mapping of $\mathscr{S}$ onto $\mathscr{S}$ whose continuous inverse $f^\vee$ is given by

$$f^\vee(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(y) e^{ix^T y} \, dy,$$

and we have $\widehat{\widehat{f}} = f(-\cdot)$. We may extend the Fourier transform to a distributional setting by defining, for $\Lambda \in \mathscr{S}'$,

$$[\widehat{\Lambda}, \phi] = [\Lambda, \widehat{\phi}] \qquad (\phi \in \mathscr{S}),$$

and the distributional Fourier transform is also a continuous, linear, one-to-one mapping, this time of $\mathscr{S}'$ onto $\mathscr{S}'$, with a continuous inverse $\Lambda^\vee$ defined in the obvious way.

It is a useful property of the Fourier domain that $\widehat{D^\alpha f} = (i\cdot)^\alpha \widehat{f}$—in other words, the Fourier transform of the derivative of a function is merely a polynomial multiplication of the Fourier transform of the function. (This property is key in generalising the notion of differentiation to a fractional-order setting.)

Parseval's formula

$$\int_{\mathbb{R}^d} \widehat{f}(x)\overline{\widehat{g}}(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\overline{g}(x)\,\mathrm{d}x \qquad (f, g \in L_2(\mathbb{R}^d))$$

reveals, via the special case $f = g$, that the Fourier transform is an isometry on $L_2(\mathbb{R}^d)$:

$$\|\widehat{f}\|_{L_2(\mathbb{R}^d)} = \|f\|_{L_2(\mathbb{R}^d)} \qquad (f \in L_2(\mathbb{R}^d)), \tag{1.3.7}$$

a result known as the Plancherel theorem. More specifically, the theorem proves the existence of a linear isometry $\Psi : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ such that $\Psi f = \widehat{f}$ for all $f \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$, with inverse $\Psi^{-1} f = f^\vee$ for all $f \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$, which extends the notion of the Fourier transform from $L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ to $L_2(\mathbb{R}^d)$. The extension Ψ is still referred to as the Fourier transform and we continue to write $\widehat{f}$ instead of Ψf for all $f \in L_2(\mathbb{R}^d)$.

For extensive further details on the above spaces and operations, the books of Adams (1978) and Rudin (1991) are excellent resources.

Chapter 2

Interpolation via Radial Basis Functions

In a fairly abstract setting, the interpolation problem involves finding a function $\mathcal{S}$ which, for a set of pairwise distinct *interpolation points* $\mathcal{A} = \{a_1, \ldots, a_n\} \subset \mathbb{R}^d$, $d \geq 1$, satisfies

$$\mathcal{S}(a) = f_a \qquad (a \in \mathcal{A}),$$

where the f_a are n *interpolation values* which correspond to each of the n interpolation points.

Usually, and certainly in the settings we will discuss, the f_a are real numbers; if we define the *data function* $f : \mathcal{A} \to \mathbb{R}$ the problem becomes one of finding an interpolant $\mathcal{S}_f : \mathbb{R}^d \to \mathbb{R}$ such that $\mathcal{S}_f(a) = f(a)$ for all $a \in \mathcal{A}$. Note that aside from ensuring that they are pairwise distinct, we have placed very little restriction on the distribution of the interpolation points in $\mathbb{R}^d$—they may be scattered wildly throughout space.

2.1 Fundamental properties of an interpolant

There are some properties of an interpolant $\mathcal{S}_f$ which would be sensible to demand. Firstly, we would hope that if the data function f was scaled by some $\alpha \in \mathbb{R}$, an interpolant to this

new set of points would be simply $\alpha\mathcal{S}_f$.

Additionally, suppose that, as well as the data f and their interpolant $\mathcal{S}_f$, we have another set of data g whose interpolant is $\mathcal{S}_g$; it would be quite natural to expect the function $\mathcal{S}_{f+g} = \mathcal{S}_f + \mathcal{S}_g$ to interpolate the data $f + g$.

Consequently, we have implied that our potential interpolants must belong to a linear space of functions, which we denote by $\mathcal{F}$ — this space is sometimes referred to as the *space of admissible functions.*

As one might expect, we cannot make such demands on our interpolant without paying a price: suppose that $v \in \mathcal{F}$ vanishes at all the interpolation points $a \in \mathcal{A}$. Let $\mathcal{S}_f \in \mathcal{F}$ be an interpolant to the values $f(a)$, $a \in \mathcal{A}$. Then

$$\mathcal{V}_{f,\alpha}(a) := \mathcal{S}_f(a) + \alpha v(a) = \mathcal{S}_f(a) + \alpha.0 = \mathcal{S}_f(a) = f(a)$$

for each $a \in \mathcal{A}$ and all $\alpha \in \mathbb{R}$. Hence $\mathcal{V}_{f,\alpha}$ interpolates the data as well as $\mathcal{S}_f$, so the interpolant is not uniquely defined — a property which is vital in practice, for numerical stability. In situations where this problem may arise, we can make an assumption about our interpolation points which neatly circumvents the issue.

Definition 2.1.1 (Unisolvent sets of points). *Let K be a space of functions defined on $\Omega \subset \mathbb{R}^d$. A subset $\mathcal{A}$ of Ω is called* unisolvent with respect to K *if the only function in K that vanishes on $\mathcal{A}$ is the zero function.*

Thereby, as long as the interpolation points form a unisolvent set with respect to the space of admissible functions, if the interpolant exists then it is unique. We shall see later that this requirement is not as restrictive on the choice of interpolation points as one might imagine. Unfortunately, before we even reach that point, things again take a turn for the worse: the following theorem is attributed to Mairhuber (1956).

Theorem 2.1.2. *Let K be an n-dimensional space of continuous, real-valued functions on $\mathbb{R}^d$. Assume that all sets of points $\mathcal{A} \subset \mathbb{R}^d$ with $|\mathcal{A}| = n$ are unisolvent with respect to K. Then either $n = 1$ or $d = 1$.*

Hence we are, to say the least, rather restricted in our choices for dimension or number of interpolation points if we want to demand that all sets of interpolation points are unisolvent with respect to every function in our space of interpolants.

2.2 Spaces of admissible functions

In order to overcome the problem revealed to us by the theorem of Mairhuber, one thing we can do is allow the space of admissible functions to depend on the interpolation points themselves — thus by changing the interpolation points, we alter the space of functions. One way to do this is to define a function $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$; then, for a set of interpolation points $\mathcal{A} \subset \mathbb{R}^d$, the space of admissible functions can be defined as

$$\mathcal{F}_{\mathcal{A},\psi} = \operatorname{span}\{\psi(\,\cdot\,, a) \,:\, a \in \mathcal{A}\}\,. \tag{2.2.1}$$

In such a setting the single function ψ generates a space of functions; we refer to ψ as the *basis function* (or sometimes as the *basic function*, since it is really the n functions $\psi(\,\cdot\,, a)$ which form a basis for $\mathcal{F}_{\mathcal{A},\psi}$).

Suppose we want to interpolate the function f at the interpolation points $a \in \mathcal{A}$ with a function from $\mathcal{F}_{\mathcal{A},\psi}$. The interpolant would be of the form

$$\mathcal{S}_f(x) = \sum_{a \in \mathcal{A}} \alpha_a \psi(x, a), \tag{2.2.2}$$

for all $x \in \mathbb{R}^d$, and the coefficients α_a would be determined by the requirements

$$\sum_{a \in \mathcal{A}} \alpha_a \psi(b, a) = f(b), \tag{2.2.3}$$

for all $b \in \mathcal{A}$; in matrix form this is written

$$A\alpha = f, \tag{2.2.4}$$

with the obvious definitions of the so-called *interpolation matrix* $A \in \mathbb{R}^{n\times n}$ and the vectors $\alpha, f \in \mathbb{R}^n$. From this representation we can see that in order to guarantee a unique solution to the interpolation problem, A must be a *non-singular matrix*. One way of doing this is to ensure A is *positive-definite*, which in turn means that ψ must be a *strictly positive-definite function*.

2.2.1 Strictly positive-definite functions

Definition 2.2.1 (Positive-definite matrix). *An $n \times n$ matrix M is called* positive-definite *if, for any $\gamma \in \mathbb{R}^n$,*

$$\gamma^T M \gamma \geq 0,$$

and, in addition, $\gamma M \gamma^T = 0$ if and only if $\gamma = 0$.

Definition 2.2.2 (Strictly positive-definite function). *A function $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is* strictly positive-definite on $\mathbb{R}^d$ *if, for any finite set of points $x_1, \ldots, x_n \in \mathbb{R}^d$, the $n \times n$ matrix A with entries given by*

$$A_{ij} = \psi(x_i, x_j) \qquad (1 \leq i, j \leq n),$$

is positive-definite.

Using a positive-definite function as our basic function, we immediately have a unique interpolant of the form (2.2.2), as proved in the following well-known theorem.

Theorem 2.2.3. *Let $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a strictly positive-definite function on $\mathbb{R}^d$. Let $\mathcal{A} \subset \mathbb{R}^d$ be an arbitrary set of n interpolation points, and let $f : \mathcal{A} \to \mathbb{R}$ be the corresponding data function. Then there exists a unique function $S_f \in \mathcal{F}_{\psi,\mathcal{A}}$ such that*

$$S_f(b) = f(b) \qquad (b \in \mathcal{A}).$$

Proof. From the matrix form (2.2.4) of the interpolation problem, we wish to show that the matrix A given by $A_{ij} = \psi(a_i, a_j)$, $1 \leq i, j \leq n$, is non-singular; we do this by showing that it has only zero in its kernel.

Let $\gamma \in \mathbb{R}^n$ be in $\ker A$, so that $A\gamma = 0$. Then $\gamma^T A\gamma = 0$. Since ψ is strictly positive-definite, A is positive-definite, and thus by definition, $\gamma = 0$. □

Here we see that the strict positive-definiteness of ψ takes care of the uniqueness of the interpolant, so we do not need to impose a unisolvency condition on $\mathcal{A}$. We now meet our first example of a strictly positive-definite function.

Example 2.2.4. *The function $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by $\psi(x, y) = e^{-\gamma|x-y|^2}$, $\gamma > 0$, is strictly positive-definite (Light 1992, p. 118). ψ is referred to as* the Gaussian.

In fact, the Gaussian is a member of a whole family of strictly positive-definite functions, revealed to us by the following lemma.

Lemma 2.2.5. *If $f \in L_1(\mathbb{R}^d)$ is continuous, non-negative and non-vanishing, then*

$$\phi(x) = \int_{\mathbb{R}^d} f(\omega) e^{-ix^T\omega} \, \mathrm{d}\omega \qquad (x \in \mathbb{R}^d),$$

is strictly positive-definite.

For a proof of this lemma, see (Wendland 2005, Corollary 6.9). In particular, this holds for $f \in L_1(\mathbb{R})$ and $x \in \mathbb{R}$ with $x = B(y, z)$ for some map $B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Thus

$$\psi(y, z) := (\phi \circ B)(y, z) = \int_{\mathbb{R}} f(\omega) e^{-iB(y,z)\omega} \, \mathrm{d}\omega,$$

is strictly positive-definite. Setting $f(\omega) = \frac{1}{2}(\pi\gamma)^{(-1/2)} e^{-\frac{\omega^2}{4\gamma}}$, $w \in \mathbb{R}$, and $B(y, z) = |y - z|$, $y, z \in \mathbb{R}^d$, verifies that the Gaussian is strictly positive-definite.

However, there are certain types of functions that have also proved effective at solving interpolation problems, but are positive-definite only on a *subspace* of $\mathbb{R}^d$.

2.2.2 Conditionally strictly positive-definite functions

We modify the definition of strictly positive-definite functions slightly to encompass those functions which are strictly positive-definite on an important subset of $\mathbb{R}^d$, and call these functions *conditionally strictly positive-definite of order m.*

Definition 2.2.6 (Conditionally (strictly) positive-definite function of order m). *A function $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is* conditionally positive-definite of order m on $\mathbb{R}^d$ *if, for any finite set of points $x_1, \ldots, x_n \in \mathbb{R}^d$, the $n \times n$ matrix A given by $A_{ij} = \psi(x_i, x_j)$, $1 \leq i, j \leq n$, satisfies*

$$\gamma^T A \gamma > 0,$$

for all $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n \setminus 0$ satisfying

$$\sum_{j=1}^{n} \gamma_j p(x_j) = 0 \qquad (p \in \Pi_{m-1}^d).$$

If, in addition, $\gamma^T A \gamma = 0$ if and only if $\gamma = 0$ then we say ψ is a conditionally strictly positive-definite function of order m on $\mathbb{R}^d$.

Thus, a strictly positive-definite function is a conditionally strictly positive-definite function of order 0.

We still wish to use these functions to solve the interpolation equations (2.2.3), which give us n degrees of freedom. By using conditionally strictly positive-definite functions we have increased the degrees of freedom of the problem by $\ell = \dim(\Pi_{m-1}^d)$. In order that we recover a square system, we require ℓ extra conditions, and so we adjust our interpolant by a polynomial of an appropriate degree, which thus has the form

$$\mathcal{S}_f(x) = \sum_{a \in \mathcal{A}} \alpha_a \psi(x, a) + \sum_{i=1}^{\ell} \beta_i p_i(x) \qquad (x \in \mathbb{R}^d), \tag{2.2.5}$$

where $\beta_1, \ldots, \beta_\ell \in \mathbb{R}$, $p_1, \ldots, p_\ell$ form a basis for Π_{m-1}^d, and we demand what are often

referred to as the *natural side conditions*:

$$\sum_{a\in\mathcal{A}} \alpha_a p_i(a) = 0 \qquad (i = 1, \ldots, \ell).$$

This choice of extra conditions, as well as the use of a polynomial modifier in the interpolant — may seem a little contrived, but we will see in Section 2.3 that, in fact, these conditions arise naturally from the general set-up.

Of course we must still satisfy the interpolation conditions

$$\sum_{a\in\mathcal{A}} \alpha_a \psi(b,a) + \sum_{i=1}^{\ell} \beta_i p_i(b) = f(b) \qquad (b \in \mathcal{A}),$$

and writing this in matrix form it is easy to see how we have now recovered a square system:

$$\begin{pmatrix} \psi(a_1,a_1) & \psi(a_1,a_2) & \cdots & \psi(a_1,a_n) & p_1(a_1) & \cdots & p_\ell(a_1) \\ \psi(a_2,a_1) & \psi(a_2,a_2) & \cdots & \psi(a_2,a_n) & p_1(a_2) & \cdots & p_\ell(a_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \psi(a_n,a_1) & \psi(a_n,a_2) & \cdots & \psi(a_n,a_n) & p_1(a_n) & \cdots & p_\ell(a_n) \\ p_1(a_1) & p_1(a_2) & \cdots & p_1(a_n) & 0 & \cdots & 0 \\ p_2(a_1) & p_2(a_2) & \cdots & p_2(a_n) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_\ell(a_1) & p_\ell(a_2) & \cdots & p_\ell(a_n) & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_\ell \end{pmatrix} = \begin{pmatrix} f(a_1) \\ f(a_2) \\ \vdots \\ f(a_n) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

or,

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}.$$

If the $(n+\ell)\times(n+\ell)$ matrix above is non-singular then we have a unique solution to the interpolation problem. Due to our polynomial modifier, we can see that as in Section 2.1, the interpolant may not be unique unless we demand the interpolation points are unisolvent

with respect to our space of admissible functions. In fact, because of the conditionally strict positive-definiteness of ψ, we need only demand unisolvency over the polynomial space.

Theorem 2.2.7. *Let ψ be conditionally strictly positive-definite of order m on $\mathbb{R}^d$, and let $\mathcal{A} \subset \mathbb{R}^d$ be a set of n interpolation points, unisolvent with respect to Π^d_{m-1}, with a corresponding data function $f : \mathcal{A} \to \mathbb{R}$.*

Then there exists a unique solution $(\alpha, \beta)^T \in \mathbb{R}^{n+\ell}$ to the interpolation equations

$$\sum_{a \in \mathcal{A}} \alpha_a \psi(b, a) + \sum_{i=1}^{\ell} \beta_i p_i(b) = f(b) \qquad (b \in \mathcal{A}), \tag{2.2.6}$$

$$\sum_{a \in \mathcal{A}} \alpha_a p_i(a) = 0 \qquad (i = 1, \ldots, \ell), \tag{2.2.7}$$

and hence a unique interpolant of the form (2.2.5).

Proof. Suppose that $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^\ell$ are such that $(\alpha, \beta)^T \in \mathbb{R}^{n+\ell}$ is in the kernel of $\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix}$. Then $A\alpha + P\beta = 0$ and $P^T\alpha = 0$. The second of these equations can be rewritten $\alpha^T P = 0$, and utilising this fact when pre-multiplying the first equation by α^T gives

$$\alpha^T A \alpha = 0.$$

Since $\alpha^T P = 0$, that is, $\sum_{a \in \mathcal{A}} \alpha_a p_i(a) = 0$ for $1 \leq i \leq \ell$, we can use the conditionally strictly positive-definiteness of ψ to conclude that $\alpha = 0$.

Hence we have $P\beta = 0$. Each of the n rows of this may be written in full as

$$q(a) := \beta_1 p_1(a) + \beta_2 p_2(a) + \ldots + \beta_\ell p_\ell(a) = 0,$$

for some $q \in \Pi^d_{m-1}$, for all $a \in \mathcal{A}$. So $q(a) = 0$ for all $a \in \mathcal{A}$, and by the unisolvency of $\mathcal{A}$, this must mean that $q = 0$, thus $\beta = 0$.

Therefore $\ker \begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix}$ is trivial and so the matrix is non-singular. □

Remark 2.2.8. *As one would hope, if we take $m = 0$ then clearly the above theorem reduces to be equivalent to Theorem 2.2.3.*

2.2.3 Characterisation of conditionally strictly positive-definite functions

In the mid-Eighties, Micchelli (1986) proved a conjecture of Franke which said that there exists a unique solution to the interpolation problem in $\mathbb{R}^2$ when interpolating via a function from the family known as *multiquadrics*, given by $\psi(x,y) = (1+|x-y|)^{\frac{1}{2}}$. The same paper also contains the first half of a characterisation of a class of conditionally strictly positive-definite functions $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ in terms of a relatively easily-checkable property. The characterisation was completed some years later by Guo, Hu & Sun (1993). We first recall the definition of *completely monotone functions.*

Definition 2.2.9 (Completely monotone function). *A function f is said to be* completely monotone on $(0,\infty)$ *if $f \in C^\infty(0,\infty)$ and $(-1)^k f^{(k)}$ is non-negative for all $k = 0,1,2,\ldots$.*

Theorem 2.2.10. *Let $\phi \in C^\infty(0,\infty) \cap C(\mathbb{R}_+)$. The function $\psi := \phi \circ |\cdot|^2$ is conditionally strictly positive-definite of order m if and only if $(-1)^m \phi^{(m)}$ is completely monotone but not identically constant on $(0,\infty)$.*

Note the subtlety here: m is the smallest number such that $(-1)^{m+k}\phi^{(m+k)} \geq 0$ for all $k = 0,1,2,\ldots$.

Corollary 2.2.11. *Let $\phi \in C^\infty(0,\infty) \cap C(\mathbb{R}_+)$ be completely monotone but not identically constant on $(0,\infty)$. Then $\psi := \phi \circ |\cdot|^2$ is strictly positive-definite on $\mathbb{R}^d$.*

Motivated by these characterisations, it is convenient to make the following definition.

Definition 2.2.12 ((Conditionally) strictly positive-definite univariate function). *A univariate function $\phi : [0,\infty) \to \mathbb{R}$ is (conditionally) strictly positive-definite (of order m) on $\mathbb{R}^d$ if the corresponding multivariate function $\psi := \phi \circ |\cdot|$ is (conditionally) strictly positive-definite (of order m) on $\mathbb{R}^d$.*

Thus, given a function $\phi \in C^\infty(0,\infty) \cap C(\mathbb{R}_+)$, in order to prove that $\psi := \phi \circ |\cdot|$ is conditionally strictly positive-definite of order m over $\mathbb{R}^d$, it suffices to show that $(-1)^{m+k}\phi^{(m+k)}(\sqrt{\cdot})$ is completely monotone but not constant in $(0,\infty)$ for $k = 0,1,2,\ldots$.

2.2.4 Radial basis functions

As interest and work in interpolation progressed, and now motivated by Michelli's characterisation, one particular type of basis function emerged as a popular and powerful tool: those functions which are *radial* — the value of the function depending only on the Euclidean distance of the argument.

Definition 2.2.13 (Radial function). *A function $\psi : \mathbb{R}^d \to \mathbb{R}$ is said to be* radial *if there exists a function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that $\psi = \phi \circ |\cdot|$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$.*

Interpolation schemes which utilise such a function refer to ψ as the *radial basis function.* It is usual, and convenient, to also refer to the univariate function ϕ as the radial basis function, despite it actually being translates of the radial function ψ which form the basis for the space of admissible functions.

With the characterisations given by Theorem 2.2.10 and Corollary 2.2.11, it is relatively easy to verify that the most commonly-used radial basis functions are strictly positive-definite or conditionally strictly positive-definite (and to determine their order in the latter case).

Example 2.2.14. *We have already seen in Example 2.2.4 that the Gaussian $\phi(r) = e^{-\gamma r^2}$ is strictly positive-definite, with $r = |x - y|$. To prove this in light of Theorem 2.2.10, observe that $\phi^{(k)}(\sqrt{r}) = (-1)^k \gamma^k e^{-\gamma r}$ for $r \in \mathbb{R}$, $k = 0, 1, 2, \ldots$. Then $(-1)^k \phi^{(k)}(\sqrt{r}) = \gamma^k \phi(\sqrt{r}) > 0$ (and not constant) for $\gamma, r > 0$.*

Example 2.2.15. *Similarly, we can prove the positive-definiteness of the inverse multiquadrics $\phi(r) = (c^2 + r^2)^{-\gamma}$, $c, \gamma > 0$, by observing that*

$$(-1)^k \phi^{(k)}(\sqrt{r}) = (-1)^k (-1)^k \gamma(\gamma + 1) \ldots (\gamma + k - 1)(c^2 + r)^{-(\gamma + k)} \geq 0,$$

for $k = 0, 1, 2, \ldots$.

Example 2.2.16. *Consider the multiquadric* $\phi(r) = -(c^2 + r^2)^{\frac{1}{2}}$, $c > 0$. ϕ *is clearly not strictly positive-definite since* $\phi(\sqrt{r}) < 0$ *for all* $r > 0$. *However, since* $(-1)^m \phi^{(m)}(\sqrt{r}) = (-1)^{m+1} \frac{1}{2}(\frac{1}{2} - 1) \cdots (\frac{1}{2} - m + 1)(c^2 + r)^{\frac{1}{2}-m} > 0$ *for* $m = 1, 2, \ldots$, *we can see that* ϕ *is conditionally strictly positive-definite of order* 1. *In fact, we can generalise this to the whole family of multiquadrics* $\phi(r) = (-1)^{\lceil \gamma \rceil}(c^2 + r^2)^{\gamma}$, $\gamma > 0$, $\gamma \notin \mathbb{N}$, *which are conditionally strictly positive-definite of order* $\lceil \gamma \rceil$.

Example 2.2.17. *The functions* $\phi(r) = (-1)^{\lceil k/2 \rceil} r^k$, *where* k *is an odd natural number, are conditionally strictly positive-definite of order* $\lceil k/2 \rceil$.

We wish to find the smallest m *such that* $(-1)^m \phi^{(m)}(\sqrt{\cdot})$ *is completely monotone on* $(0, \infty)$. *Let* $f_k(r) := \phi(\sqrt{r}) = (-1)^{\lceil k/2 \rceil} r^{\frac{k}{2}}$. *Then*

$$f_k^{(m)}(r) = (-1)^{\lceil k/2 \rceil} \frac{k}{2} \left(\frac{k}{2} - 1 \right) \cdots \left(\frac{k}{2} - m + 1 \right) r^{\frac{k}{2} - m},$$

and thus, observing that $\lceil \frac{k}{2} \rceil = \frac{k+1}{2}$, $(-1)^m f_k^{(m)}$ *is completely monotone for* $m \geq \lceil \frac{k}{2} \rceil$ *but not for* $m = \lceil \frac{k}{2} \rceil - 1$.

Example 2.2.18. *The functions* $\phi(r) = (-1)^{k+1} r^{2k} \ln(r)$ *are conditionally strictly positive-definite of order* $k + 1$ *on* $\mathbb{R}^d$.

As noted by Wendland (2005), since $2\phi(r) = (-1)^{k+1} r^{2k} \ln(r^2)$, *it suffices to show that* $f_k(r) := (-1)^{k+1} r^k \ln(r)$ *is conditionally strictly positive-definite of order* $k + 1$.

This follows from the observation that

$$f_k^{(m)}(r) = (-1)^{k+1} k(k-1)(k-2) \cdots (k-m+1) r^{k-m} \ln(r) + p_m(r) \qquad (1 \leq m \leq k),$$

where p_m *is a polynomial of degree* $k - m$. *In particular, when* $m = k$, $(-1)^k f_k^{(k)} = (-1)^{2k+1} k! \ln(r) + c$ *is not completely monotone on* $(0, \infty)$, *whereas since for* $m = k + 1$, $f_k^{(k+1)} = (-1)^{k+1} k! r^{-1}$, *we have* $(-1)^{k+1} f_k^{(k+1)} = k! r^{-1}$ *which is completely monotone on* $(0, \infty)$ *and so* ϕ *is conditionally strictly positive-definite of order* $k + 1$.

Radial basis function	$\phi(r),\ r > 0$	m
Gaussian	$e^{-\lambda r^2},\ \lambda > 0$	0
Inverse multiquadrics	$(c^2 + r^2)^{-\gamma},\ c, \gamma > 0$	0
Multiquadrics	$(-1)^{\lceil \gamma \rceil}(c^2 + r^2)^{\gamma},\ c, \gamma > 0$	$\lceil \gamma \rceil$
Polyharmonic splines	$(-1)^{\lceil k/2 \rceil} r^k,\ k > 0,\ k \notin 2\mathbb{N}_0$ $(-1)^{k+1} r^{2k} \ln(r),\ k > 0$	$\lceil k/2 \rceil$ $k + 1$

Table 2.1: Commonly-used radial basis functions and their order m of conditionally strictly positive-definiteness.

Remark 2.2.19. *The functions of Examples 2.2.17 and 2.2.18 are known collectively as the* polyharmonic splines. *This is because they are the multi-dimensional generalisations of the well-known one-dimensional cubic spline,* $\phi(r) = r^3$.

We have collected together the results of the above examples for easy reference to the common names and corresponding order of conditionally strict positive-definiteness of each of the radial basis functions we have just met — see Table 2.1.

So it seems that things are progressing nicely: we apparently have a good collection of functions with which to carry out our interpolation, and we know that they produce unique interpolants to our data. But we have said little about why the form of the interpolant (2.2.5) comes about — for instance, why the addition of a polynomial term would be the right thing to do. And since the space of admissible functions is essentially determined by linear combinations of the basis function, we are not really sure exactly what kind of functions we can interpolate — they may be of no practical value whatsoever. We need to be able to say more about the properties of the functions we are interpolating — in particular, their smoothness.

To address these issues, it seems appropriate to tackle the problem from an alternative, almost opposite direction: start with a space of functions that we would like to be able to interpolate, and see if we can deduce anything about the form or properties of the interpolant. This strategy is called a *variational approach*, wherein assumed properties of the space of functions itself are used to show that the form of the interpolant (2.2.5) arises quite naturally.

2.3 A variational approach to the interpolation problem

As discussed early in Section 2.1, it is prudent to assume that our space of functions is linear, and we would like to be able to talk about the value of the functions in a pointwise sense over $\mathbb{R}^d$, so let $\mathcal{F}$ be a linear space of continuous functions $\mathbb{R}^d \to \mathbb{R}$. Assume $\mathcal{F}$ is equipped with a semi-inner product $(\cdot, \cdot)_{\mathcal{F}}$, and hence a semi-norm $|\cdot|_{\mathcal{F}} := (\cdot, \cdot)_{\mathcal{F}}^{\frac{1}{2}}$, which determines an ℓ-dimensional kernel K.

As before, let $\mathcal{A} \subset \mathbb{R}^d$ be a finite set of n interpolation points, with given real-valued interpolation data $f_a := f(a)$ for $a \in \mathcal{A}$, for some function $f \in \mathcal{F}$ which we are attempting to interpolate. To go some way to ensuring uniqueness, we must assume $\mathcal{A}$ is unisolvent with respect to K, for if not, we may adjust any interpolant by a function $v \in K$ which vanishes on $\mathcal{A}$.

With our set of interpolation points in place, we can adjust the semi-inner product $(\cdot, \cdot)_{\mathcal{F}}$ in a certain way to define a true inner product on $\mathcal{F}$. Pick a subset $\mathcal{A}' \subset \mathcal{A}$ of size ℓ which is unisolvent with respect to K. Such a set exists, for if it did not, then $\mathcal{A}$ itself would not be unisolvent with respect to K. (It is time to come clean and admit that we have tacitly assumed throughout that $n > \ell$.) Now define

$$\langle f, g \rangle_{\mathcal{F}} := (f, g)_{\mathcal{F}} + \sum_{a \in \mathcal{A}'} f(a) g(a), \tag{2.3.1}$$

for $f, g \in \mathcal{F}$. This is a genuine inner product on $\mathcal{F}$ since $(\cdot, \cdot)_{\mathcal{F}}$ is a semi-inner product. Denote the induced norm by $\|\cdot\|_{\mathcal{F}}$, and assume $\mathcal{F}$ is complete with respect to this norm. Hence we have a Hilbert space.

2.3.1 Reproducing kernel Hilbert spaces

As mentioned above, for the purposes of interpolation, it is important to be able to determine pointwise values of functions in $\mathcal{F}$, so we have assumed the functions are continuous—in other words, that the point evaluation functional δ_x such that $\delta_x(f) = f(x)$ for all $x \in \mathbb{R}^d$ is a bounded linear functional in the dual of $\mathcal{F}$. This means that for each $x \in \mathbb{R}^d$, there

exists a constant C_x such that $|f(x)| \leq C_x \|f\|_{\mathcal{F}}$, for all $f \in \mathcal{F}$. We may now apply the Riesz representation theorem for Hilbert spaces (Adams 1978), which states that δ_x is a bounded linear functional if and only if $\mathcal{F}$ is a *reproducing kernel Hilbert space.*

Definition 2.3.1 (Reproducing kernel Hilbert space). *Let $\mathcal{F}$ be a Hilbert space of real-valued functions on $\mathbb{R}^d$, equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{F}}$. If, for each $x \in \mathbb{R}^d$, there exists a unique element $R_x \in \mathcal{F}$ such that $f(x) = \langle f, R_x \rangle_{\mathcal{F}}$ for all $f \in \mathcal{F}$, then $\mathcal{F}$ is called a* reproducing kernel Hilbert space. *The function R_x is called the* reproducing kernel for x in $\mathcal{F}$.

The power of the reproducing kernels will be revealed in a moment; firstly, because there may be more than one function in $\mathcal{F}$ that satisfies the interpolation conditions (that is, functions that agree with f at the interpolation points), we need a consistent way of defining the function which we will call "the interpolant to f".

2.3.2 Interpolants of minimal norm

Consider the set $M = \{v \in \mathcal{F} : v(a) = f(a), a \in \mathcal{A}\}$. This is the set of all elements of $\mathcal{F}$ which interpolate $f \in \mathcal{F}$. Observe that M is a closed, convex, non-empty subset of $\mathcal{F}$ and as such possesses a unique element of minimal norm (Cheney 1966, page 22).

Definition 2.3.2 (Minimal norm interpolant). *The* minimal norm interpolant to f on $\mathcal{A}$ from $\mathcal{F}$ *is the unique element of M of minimal norm, which shall be denoted $\mathcal{I}_{\mathcal{A}} f$.*

Henceforth, whenever we speak of *the* interpolant to a function, we will mean the interpolant of minimal norm, and it is in terms of the minimal norm interpolant that we formulate the *variational problem*: find $\mathcal{I}_{\mathcal{A}} f$ such that

$$\mathcal{I}_{\mathcal{A}} f(a) = f(a) \text{ for all } a \in \mathcal{A} \text{ (the interpolation conditions)}, \tag{2.3.2a}$$

$$|\mathcal{I}_{\mathcal{A}} f|_{\mathcal{F}} \leq |g|_{\mathcal{F}} \text{ for all } g \in M \text{ (the minimal norm condition)}. \tag{2.3.2b}$$

(The second of these conditions is a reduction of the more "natural" condition $\|\mathcal{I}_{\mathcal{A}} f\|_{\mathcal{F}} \leq \|g\|_{\mathcal{F}}$ for all $g \in M$, by the definition of $\|\cdot\|_{\mathcal{F}}$.)

One might ask for further justification for the choice of the element of minimal norm to be *the* interpolant, and we shall address this shortly, in Remark 2.3.5. For now, we learn that, rather surprisingly, we can calculate $\mathcal{I}_{\mathcal{A}}f$ explicitly, via the reproducing kernels for $\mathcal{A}$.

Theorem 2.3.3. *Let $\mathcal{F}$ be a reproducing kernel Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ defined by* (2.3.1). *Let $f \in \mathcal{F}$ and let $\mathcal{A}$ be a set of n pairwise distinct interpolation points, unisolvent with respect to K, the kernel of $\mathcal{F}$ with respect to the semi-inner product $(\cdot, \cdot)_{\mathcal{F}}$. Let $\mathcal{I}_{\mathcal{A}}f$ be the minimal norm interpolant to f on $\mathcal{A}$ from $\mathcal{F}$. Then*

$$\mathcal{I}_{\mathcal{A}}f = \sum_{c \in \mathcal{A}} \alpha_c R_c,$$

where the coefficients α_c are determined by the equations

$$\sum_{b \in \mathcal{A}} \alpha_b \langle R_b, R_c \rangle_{\mathcal{F}} = \langle f, R_c \rangle_{\mathcal{F}},$$

for all $c \in \mathcal{A}$, where R_c denotes the reproducing kernel for c in $\mathcal{F}$.

Remark 2.3.4. *Recall that $\langle f, R_c \rangle_{\mathcal{F}} = f(c)$, which means that, as we would expect, the coefficients α_c are determined by the values of f at the interpolation points.*

The proof of Theorem 2.3.3 that follows is a mild rewriting of the result found in Cheney & Light (2000, Chapter 30, Theorem 3).

Proof of Theorem 2.3.3. Define a subset V of $\mathcal{F}$ consisting of all functions which vanish on the interpolation points—that is $V := \bigcap_{a \in \mathcal{A}} \{v \in \mathcal{F} : v(a) = 0\}$. The minimal norm condition is equivalent to minimising $|\mathcal{I}_{\mathcal{A}}f|_{\mathcal{F}}$ subject to $(f - \mathcal{I}_{\mathcal{A}}f) \in V$, and setting $v = f - \mathcal{I}_{\mathcal{A}}f$, we can restate this as minimising $|f - v|_{\mathcal{F}}$ subject to $v \in V$.

This is a standard problem of best approximation and its solution is characterised by the conditions $v \in V$ and $f - v \perp V$ (Cheney & Light 2000); hence $\mathcal{I}_{\mathcal{A}}f \perp V$.

Since V is a subset of $\mathcal{F}$ we know there exist unique reproducing kernels R_a such that, for all $v \in V$, $\langle v, R_a \rangle_{\mathcal{F}} = v(a) = 0$ for all $a \in \mathcal{A}$. Thus we can rewrite V in terms of these reproducing kernels:

$$V = \bigcap_{a \in \mathcal{A}} \{v \in \mathcal{F} : \langle v, R_a \rangle_{\mathcal{F}} = 0\} \equiv \bigcap_{a \in \mathcal{A}} R_a^{\perp}.$$

This is the set of elements of $\mathcal{F}$ which are orthogonal to R_a for all $a \in \mathcal{A}$, and hence $V = \Big(\text{span}\{R_a : a \in \mathcal{A}\}\Big)^{\perp}$. Thus $\mathcal{I}_{\mathcal{A}} f \in \text{span}\{R_a : a \in \mathcal{A}\}$ and the coefficients in the span are determined by the interpolation equations. □

We saw in the above proof that $\mathcal{I}_{\mathcal{A}} f \perp V$, so for any $v \in V$ we have

$$\begin{aligned} \|v - \mathcal{I}_{\mathcal{A}} f\|_{\mathcal{F}}^2 &= \langle v - \mathcal{I}_{\mathcal{A}} f, v - \mathcal{I}_{\mathcal{A}} f \rangle_{\mathcal{F}} \\ &= \langle v, v - \mathcal{I}_{\mathcal{A}} f \rangle_{\mathcal{F}} - \langle \mathcal{I}_{\mathcal{A}} f, v - \mathcal{I}_{\mathcal{A}} f \rangle_{\mathcal{F}} \\ &= \langle v, v \rangle_{\mathcal{F}} + \langle \mathcal{I}_{\mathcal{A}} f, \mathcal{I}_{\mathcal{A}} f \rangle_{\mathcal{F}} \\ &= \|v\|_{\mathcal{F}}^2 + \|\mathcal{I}_{\mathcal{A}} f\|_{\mathcal{F}}^2, \end{aligned}$$

giving us a *Pythagorean law* relating elements of V to the interpolant. Now put $v = \mathcal{I}_{\mathcal{A}} f - f$, and apply the interpolation conditions to deduce

$$|f - \mathcal{I}_{\mathcal{A}} f|_{\mathcal{F}}^2 = |f|_{\mathcal{F}}^2 - |\mathcal{I}_{\mathcal{A}} f|_{\mathcal{F}}^2, \tag{2.3.3}$$

for all $f \in \mathcal{F}$, which is called the *Pythagorean property of the interpolant.*

Remark 2.3.5. *We return to the question of why the element of minimal norm should be the one chosen to be the interpolant to f. To justify this choice, suppose that, in addition to the values of f at the interpolation points, we are told that $\|f\|_{\mathcal{F}} \leq r$, for some $r \in \mathbb{R}$. Then we can summarise all the information we have about f by:*

$$f \in C := \{c \in M : \|c\|_{\mathcal{F}} \leq r\} = \{c \in \mathcal{F} : \|c\|_{\mathcal{F}} \leq r, c(a) = f(a), a \in \mathcal{A}\}.$$

Notice that, due to its minimal norm property, $\mathcal{I}_{\mathcal{A}}f \in C$. Moreover, for $c \in C$ with $\|c\|_{\mathcal{F}} = r$, by definition there exists $v = (c - \mathcal{I}_{\mathcal{A}}f) \in V$, so by the Pythagorean law for elements of V,

$$\|c\|_{\mathcal{F}}^2 = \|\mathcal{I}_{\mathcal{A}}f - c\|_{\mathcal{F}}^2 + \|\mathcal{I}_{\mathcal{A}}f\|_{\mathcal{F}}^2.$$

Thus $\|\mathcal{I}_{\mathcal{A}}f - c\|_{\mathcal{F}}^2 = r^2 - \|\mathcal{I}_{\mathcal{A}}f\|_{\mathcal{F}}^2$, so $\mathcal{I}_{\mathcal{A}}f$ is equidistant from all points on the circumference of C — in other words, the minimal norm interpolant lies at the centre of C; since we know nothing of the location of f within C, the centre of C is a reasonable choice to make for the approximation to f.

By Theorem 2.3.3, we will be able to determine the form of the interpolant to data prescribed on $\mathcal{A}$ provided we know the reproducing kernels R_c for each $c \in \mathcal{A}$. Light & Wayne (1998) proved the form of the reproducing kernels in terms of a function $\phi \in C(\mathbb{R}^d)$ on the condition that the kernel K has a *Lagrange basis.*

Definition 2.3.6 (Lagrange basis). *Let F be a space of continuous functions in $\mathbb{R}^d$ of dimension n. Let $\mathcal{A}$ be a finite set of n pairwise distinct points in $\mathbb{R}^d$. The* Lagrange basis for F *is the set of functions $\lambda_a \in F$, $a \in \mathcal{A}$ such that*

$$\lambda_a(b) = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases}$$

for all $b \in \mathcal{A}$.

For convenience, we gather the so-far-assumed properties of the space $\mathcal{F}$.

Assumption 2.3.7. *The space $\mathcal{F} = \mathcal{F}(\mathbb{R}^d)$ of functions $\mathbb{R}^d \to \mathbb{R}$ has the following properties:*

1. *$\mathcal{F} \subset C(\mathbb{R}^d)$;*
2. *A semi-inner product $(\cdot, \cdot)_{\mathcal{F}}$ is defined on $\mathcal{F}$ with corresponding semi-norm $|\cdot|_{\mathcal{F}}$ having kernel K;*

3. *Let $\mathcal{A}'$ be a finite set of pairwise-distinct points unisolvent with respect to K and let $\{p_a\}_{a\in\mathcal{A}'}$ be a Lagrange basis for K. Then there exists a function $\phi \in C(\mathbb{R})$ such that, for each $x \in \mathbb{R}^d$,*

$$r_x(y) = \phi(|x-y|) - \sum_{a\in\mathcal{A}'} p_a(x)\phi(|y-a|) \qquad (y \in \mathbb{R}^d),$$

defines a function $r_x \in \mathcal{F}$ with $\langle f, r_x\rangle_{\mathcal{F}} = f(x)$ for all $f \in \mathcal{F}(\mathbb{R}^d)$ with the property that $f(a) = 0$ for all $a \in \mathcal{A}'$. Here, for $u, v \in \mathcal{F}$,

$$\langle u, v\rangle_{\mathcal{F}} = (u, v)_{\mathcal{F}} + \sum_{a\in\mathcal{A}'} u(a)v(a).$$

Remark 2.3.8. *The use of the symbol ϕ, which was used previously to represent a basic function, in this new context is not meant to confuse the reader; on the contrary, it is intended to seed the notion that the function will eventually turn out to be a radial basis function that we have already met.*

We now quote the aforementioned result of Light & Wayne (1998).

Lemma 2.3.9. *Let $\mathcal{F}$ satisfy Assumption 2.3.7. Then the representer of point evaluation — that is, the element $R_x \in \mathcal{F}$ such that $\langle f, R_x\rangle_{\mathcal{F}} = f(x)$ for all $f \in \mathcal{F}$ — has the form*

$$\begin{aligned} R_x(y) = \phi(|y-x|) &- \sum_{a\in\mathcal{A}'} p_a(x)\phi(|y-a|) \\ &- \sum_{b\in\mathcal{A}'} \Big(\phi(|b-x|) - \sum_{a\in\mathcal{A}'} p_a(x)\phi(|b-a|)\Big)p_b(y) + \sum_{b\in\mathcal{A}'} p_b(x)p_b(y). \end{aligned}$$

So if, in addition to the above assumptions, $\mathcal{F}$ is complete with respect to the norm $\|\cdot\|_{\mathcal{F}} = \langle\cdot,\cdot\rangle_{\mathcal{F}}^{\frac{1}{2}}$, then by Theorem 2.3.3 we know we have a unique interpolant of minimal norm, and using Lemma 2.3.9 gives us the form of the representers.

In particular, for the interpolation points $c \in \mathcal{A}$, we have, for all $y \in \mathbb{R}^d$,

$$\begin{aligned} R_c(y) = \phi(|y-c|) &- \sum_{a\in\mathcal{A}'} p_a(c)\phi(|y-a|) \\ &- \sum_{b\in\mathcal{A}'} \big(\phi(|b-c|) - \sum_{a\in\mathcal{A}'} p_a(c)\phi(|b-a|)\big)p_b(y) + \sum_{b\in\mathcal{A}'} p_b(c)p_b(y). \end{aligned} \tag{2.3.4}$$

When $c \in \mathcal{A}'$, it is clear from (2.3.4) that $R_c(y) = p_c(y)$. Moreover, we can see that for each $c \in \mathcal{A} \setminus \mathcal{A}'$ there exist constants γ_a^c, $a \in \mathcal{A}$, and ν_b^c, for $b \in \mathcal{A}'$, such that we can rewrite (2.3.4) as

$$R_c(y) = \sum_{a \in \mathcal{A}} \gamma_a^c \phi(|y - a|) + \sum_{b \in \mathcal{A}'} \nu_b^c p_b(y)$$

with constants γ_a^c explicitly given by

$$\gamma_a^c = \begin{cases} 1, & a = c, \\ -p_a(c), & a \in \mathcal{A}', \\ 0, & \text{otherwise}, \end{cases}$$

and constants ν_b^c whose explicit description is not important here. Via the two lemmas that follow we shall see that we have been led quite naturally to an interpolant with coefficients selected in an identical manner to that previously seen in the "direct" construction which resulted in Theorem 2.2.7.

Lemma 2.3.10. *Let $c \in \mathcal{A} \setminus \mathcal{A}'$. For each $b \in \mathcal{A}'$, $\sum_{a \in \mathcal{A}} \gamma_a^c p_b(a) = 0$.*

Proof. Using the definition of γ_a^c and properties of the Lagrange basis functions,

$$\sum_{a \in \mathcal{A}} \gamma_a^c p_b(a) = p_b(c) - \sum_{a \in \mathcal{A}'} p_a(c) p_b(a) = p_b(c) - p_b(c) = 0.$$

□

Hence, Theorem 2.3.3 tells us that can write the interpolant of f as

$$\begin{aligned} \mathcal{I}_{\mathcal{A}} f(y) &= \sum_{c \in \mathcal{A}} \alpha_c R_c(y) \\ &= \sum_{c \in \mathcal{A} \setminus \mathcal{A}'} \alpha_c R_c(y) + \sum_{b \in \mathcal{A}'} \alpha_b R_b(y) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \left(\sum_{a\in\mathcal{A}} \gamma_a^c \phi(|y-a|) + \sum_{b\in\mathcal{A}'} \nu_b^c p_b(y) \right) + \sum_{b\in\mathcal{A}'} \alpha_b R_b(y) \\
&= \sum_{a\in\mathcal{A}} \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \gamma_a^c \phi(|y-a|) + \sum_{b\in\mathcal{A}'} \left(\sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \nu_b^c + \alpha_b \right) p_b(y) \\
&= \sum_{a\in\mathcal{A}} \gamma_a \phi(|y-a|) + \sum_{b\in\mathcal{A}'} \mu_b p_b(y),
\end{aligned}
$$

where

$$
\gamma_a = \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \gamma_a^c, \qquad \mu_b = \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \nu_b^c + \alpha_b.
$$

Lemma 2.3.11. *Let $b \in \mathcal{A}'$. Then $\sum_{a\in\mathcal{A}} \gamma_a p_b(a) = 0$.*

Proof.

$$
\sum_{a\in\mathcal{A}} \gamma_a p_b(a) = \sum_{a\in\mathcal{A}} \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \gamma_a^c p_b(a) = \sum_{c\in\mathcal{A}\setminus\mathcal{A}'} \alpha_c \sum_{a\in\mathcal{A}} \gamma_a^c p_b(a) = 0,
$$

by Lemma 2.3.10. □

Thus the interpolation problem boils down to finding constants γ_b for $b \in \mathcal{A}$ and μ_b for $b \in \mathcal{A}'$ that satisfy

$$
\begin{aligned}
\sum_{b\in\mathcal{A}} \gamma_b \phi(|a-b|) + \sum_{b\in\mathcal{A}'} \mu_b p_b(a) &= f(a) \qquad (a \in \mathcal{A}), \\
\sum_{b\in\mathcal{A}} \gamma_b p_a(b) &= 0 \qquad (a \in \mathcal{A}').
\end{aligned}
$$

Hence we have deduced that the unique solution to the variational problem (2.3.2) is of precisely the same form as the solution displayed in Theorem 2.2.7. Furthermore, the selection of the coefficients is in the same manner in both cases, providing justification for the apparently artificial selection of the natural conditions that we saw in Section 2.2.2.

Now that we know more about the form of the interpolant, we can investigate certain well-known spaces which happen to satisfy the assumptions we have made on $\mathcal{F}$, and use

these to provide more insight into how interpolants comprising a sum of translates of a radial basis function arise quite naturally from the variational approach.

2.4 Beppo Levi spaces

As we have said in Assumption 2.3.7, the space of functions we are interpolating must be at least globally continuous; beyond that assumption, we would like to know precisely how smooth the functions are. To this end, we define suitable spaces of continuous functions, the *Beppo Levi spaces*, popularly-used in the radial basis function literature, for they are the native space for the polyharmonic splines.

Definition 2.4.1 (Fractional-order Beppo Levi space on $\mathbb{R}^d$). *Let $d, k \in \mathbb{Z}_+$ and $0 \leq \mu < 1$ such that $k + \mu > d/2$. The* fractional-order local Beppo Levi space on $\mathbb{R}^d$ *is defined as*

$$BL^{k+\mu}(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) \,:\, |f|_{BL^{k+\mu}(\mathbb{R}^d)} < \infty \right\},$$

where the weighted semi-inner product given by

$$(f, g)_{BL^{k+\mu}(\mathbb{R}^d)} = \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^d} (\widehat{D^\alpha f})(x) \overline{(\widehat{D^\alpha g})(x)} |x|^{2\mu} \, \mathrm{d}x,$$

for $f, g \in BL^{k+\mu}(\mathbb{R}^d)$, induces the semi-norm $|\cdot|_{BL^{k+\mu}(\mathbb{R}^d)} = (\cdot, \cdot)^{\frac{1}{2}}_{BL^{k+\mu}(\mathbb{R}^d)}$.

The constants c_α are chosen so that the semi-norm is rotationally invariant —

$$\sum_{|\alpha|=k} c_\alpha x^{2\alpha} = |x|^{2k} \qquad (x \in \mathbb{R}^d),$$

and are given explicitly by $c_\alpha = \binom{k}{\alpha!}$.

When $\mu = 0$, by the Plancherel theorem (1.3.7) *we can write the semi-norm alternatively as $|f|_{BL^k(\mathbb{R}^d)} = \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^d} |D^\alpha f|^2$.*

Note that k and μ are assumed to be optimal, by which we mean $0 \leq \mu < 1$. The following theorem shows that $BL^{k+\mu}(\mathbb{R}^d)$ is an example of the function space $\mathcal{F}$ that we have been building-up over the previous sections.

Theorem 2.4.2. $\left(BL^{k+\mu}(\mathbb{R}^d), |\cdot|_{BL^{k+\mu}(\mathbb{R}^d)}\right)$ *satisfies the properties of* $\mathcal{F}$ *in Assumption 2.3.7 with* $\phi \equiv \phi_{d,k,\mu} \in C(\mathbb{R})$ *the polyharmonic spline given by*

$$\phi_{d,k,\mu}(r) = \begin{cases} (-1)^{k+\mu-d/2+1} r^{2k+2\mu-d} \ln(r), & 2k+2\mu-d \text{ even}, \\ (-1)^{\lceil k+\mu-d/2 \rceil} r^{2k+2\mu-d}, & \text{otherwise.} \end{cases} \tag{2.4.1}$$

For $\mu = 0$ *the semi-norm has kernel* Π^d_{k-1}*; when* $\mu > 0$ *the kernel is* Π^d_k.

Given the dependence on μ of the semi-norm kernel, it is convenient to define

$$\mathcal{K}_{k,\mu} = \begin{cases} \Pi^d_{k-1}, & \mu = 0, \\ \Pi^d_k, & \mu > 0. \end{cases} \tag{2.4.2}$$

For the integer-order case (that is, when $\mu = 0$), this theorem is proved over several results in Wendland's (2005, Chapter 10) book; for the general result, see the Ph.D. thesis of Wayne (1996).

Remark 2.4.3. *As suggested by Examples 2.2.17 and 2.2.18, the constants* $(-1)^{k+\mu-d/2+1}$ *and* $(-1)^{\lceil k+\mu-d/2 \rceil}$ *are required to ensure* ϕ *is conditionally strictly positive-definite and thus provides a unique solution to the interpolation problem, in accordance with Theorems 2.2.7 and 2.2.10.*

Remark 2.4.4. *Another way to approach this set-up is to choose a set* $\mathcal{A}' \subset \mathcal{A}$ *of size* $\ell = \dim(\mathcal{K}_{k,\mu})$ *which is unisolvent with respect to* $\mathcal{K}_{k,\mu}$*, and modify the semi-norm* $|\cdot|_{BL^{k+\mu}(\mathbb{R}^d)}$ *as described on page 24, to form a genuine norm* $\|\cdot\|_{BL^{k+\mu}(\mathbb{R}^d)}$*. Then, with* $\phi \equiv \phi_{d,k,\mu}$*, the completion of* $\mathcal{F}$ *with respect to this norm is* $BL^{k+\mu}(\mathbb{R}^d)$.

The notion of Beppo Levi spaces on $\mathbb{R}^d$ may be generalised to domains $\Omega \subseteq \mathbb{R}^d$, in terms of the tempered distributions defined in Section 1.3, but it is not simply a case of restricting

the integration to Ω instead of $\mathbb{R}^d$, due to the global nature of the Fourier transform. As shown by Levesley & Light (1999), one must first determine what they call the "direct form" of the global semi-norm—that is, an expression in terms of the original function rather than its Fourier transform. Once this form has been found, it can then be restricted to the local domain, and thus we arrive at the following definition.

Definition 2.4.5 (Fractional-order Beppo Levi space on Ω). *Let $d, k \in \mathbb{Z}_+$ and $0 \leq \mu < 1$. Let $\Omega \subset \mathbb{R}^d$. The* fractional-order local Beppo Levi space on Ω *is defined as*

$$BL^{k+\mu}(\Omega) = \left\{ f \in \mathscr{S}' : |f|_{BL^{k+\mu}(\Omega)} < \infty \right\},$$

where the local weighted semi-inner product is given by

$$(f,g)_{BL^{k+\mu}(\Omega)} = \sum_{|\alpha|=k} c_\alpha \int_\Omega \int_\Omega \frac{\Big((D^\alpha f)(x) - (D^\alpha f)(y)\Big)\Big((D^\alpha g)(x) - (D^\alpha g)(y)\Big)}{|x-y|^{2\mu+d}} \,\mathrm{d}x\,\mathrm{d}y,$$

for $f, g \in BL^{k+\mu}(\Omega)$, which induces the semi-norm $|\cdot|_{BL^{k+\mu}(\Omega)} = (\cdot,\cdot)^{\frac{1}{2}}_{BL^{k+\mu}(\Omega)}$. The constants c_α are as in Definition 2.4.1.

Remark 2.4.6. *The weight $|\cdot|^{-2\mu-d}$ may seem a little mysterious at first glance, but we should not be too worried: it is merely a consequence of the theory of Levesley & Light (1999, Theorem 3.7) that we are taking the Fourier transform of the "global weight" $|\cdot|^{2\mu}$.*

If the local spaces $BL^{k+\mu}(\Omega)$ are unfamiliar to the reader, the following result should provide some more comfortable ground.

Lemma 2.4.7. *Let $\Omega \subseteq \mathbb{R}^d$. Let $0 \leq \mu < 1$, and k and d be integers chosen so that $k \geq 0$, $d \geq 1$ and $k + \mu - d/2 > j \geq 0$. Then $BL^{k+\mu}(\Omega) \subset C^{[j]}(\Omega)$ for $\Omega \subseteq \mathbb{R}^d$, where $[j]$ denotes the largest integer that does not exceed j.*

Proof. Beppo Levi spaces are an example of the spaces studied in the Ph.D. thesis of Wayne (1996), see Theorem 4.1.18 and Section 5.1 therein; alternatively see Light & Wayne's (1999, Theorem 2.12) paper. □

Remark 2.4.8. *In light of Lemma 2.4.7, henceforth we assume $k + \mu - d/2$ is at least positive, so that point evaluation makes sense.*

2.5 The Lagrange representation of the interpolant

There is an alternative representation of the radial basis function interpolant which utilises a Lagrange basis for the whole of $\mathcal{F}$ over $\mathcal{A}$ (as opposed to a Lagrange basis for the kernel space) — in other words, those functions $\lambda_a \in \mathcal{F}$ satisfying $\lambda_a(b) = \delta_{ab}$ for $a, b \in \mathcal{A}$, where δ_{ab} is the Kronecker delta function. These functions, once known, immediately give us the interpolant

$$\mathcal{I}_{\mathcal{A}} f(x) = \sum_{a \in \mathcal{A}} \lambda_a(x) f(a), \tag{2.5.1}$$

for all $x \in \mathbb{R}^d$, which clearly satisfies the interpolation conditions $\mathcal{I}_{\mathcal{A}} f(b) = f(b)$ for all $b \in \mathcal{A}$.

Finding the values of the Lagrange basis functions at an arbitrary point $x \in \mathbb{R}^d$ requires the solution of the linear system

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda(x) \\ \tau(x) \end{pmatrix} = \begin{pmatrix} \varphi(x) \\ \pi(x) \end{pmatrix},$$

where A and P are the usual interpolation matrices as described in Section 2.2.2, and, in the context of the kernel $\mathcal{K}_{k,\mu}$,

$$\begin{aligned} \lambda(x) &= (\lambda_a(x))_{a \in \mathcal{A}} \in \mathbb{R}^n, \\ \tau(x) &= (\tau_\alpha(x))_{|\alpha| < \tilde{k}} \in \mathbb{R}^\ell, \\ \varphi(x) &= (\phi(|x - a|))_{a \in \mathcal{A}} \in \mathbb{R}^n, \\ \pi(x) &= (x^\alpha)_{|\alpha| < \tilde{k}} \in \mathbb{R}^\ell, \end{aligned}$$

with $\alpha \in \mathbb{Z}_+^d$ and

$$\tilde{k} = \begin{cases} k, & \mu = 0 \\ k+1 & \mu > 0. \end{cases}$$

As ever, assuming that $\mathcal{A}$ is unisolvent with respect to $\mathcal{K}_{k,\mu}$, for any $p \in \mathcal{K}_{k,\mu}$ we have polynomial reproduction, in that

$$\mathcal{I}_{\mathcal{A}} p = \sum_{a \in \mathcal{A}} \lambda_a p(a) = p.$$

Because pointwise evaluation entails solving a linear system, computing values of an interpolant based on its Lagrange basis form is an expensive procedure practically speaking, but this representation can prove useful from a theoretical perspective.

2.6 Error estimates

Arguably the most important question one can ask of an interpolant is: how "close" does it come to reproducing the function it interpolates? Ideally the so-called *error estimate*, however we define this quantity, would be given in terms of something over which we have a good deal of control — for example, some measure of the distance between the interpolation points. Intuitively, we would expect the interpolant to be more accurate the greater the number of interpolation points we use, but if we double the number of points, does the error only halve? It is one of the great achievements of approximation theory that we can provide error estimates that answer this question.

Suppose that our finite set of interpolation points $\mathcal{A}$ is contained in some open bounded subset Ω of $\mathbb{R}^d$, and let $\overline{\Omega}$ denote the closure of Ω (in other words, $\overline{\Omega} = \Omega \cup \partial\Omega$).

Definition 2.6.1 (Fill distance). *The* fill distance *of $\mathcal{A}$ in Ω is defined to be the quantity*

$$h = \sup_{x \in \overline{\Omega}} \min_{a \in \mathcal{A}} |x - a|.$$

One can think of h as being the radius of the largest open ball, centred at a point in Ω, such that no points in $\mathcal{A}$ lie inside the ball.

Although the radial basis function interpolant that we construct is globally supported, lying in the space $\mathcal{F}$, the interpolation points are the only points where we know the precise values of the data function, so we cannot expect to be able to determine much about the accuracy of our interpolant far away from these points. Yet of course we would like to be able to say something about the error away from the interpolation points (by construction, the error at each interpolation point is zero), so we try to deduce error estimates over Ω rather than the whole of $\mathbb{R}^d$ (for example, when interpolating with polyharmonic splines, we may assume the interpoland f is in a local Beppo Levi space rather than a global one).

2.6.1 Pointwise error estimates

A natural quantity to look at is $|f(x) - \mathcal{I}_{\mathcal{A}} f(x)|$ — that is, the difference in modulus between the function and its interpolant evaluated at a particular point $x \in \Omega$. Let us assume that $\mathcal{F}$ is complete with respect to $\|\cdot\|_{\mathcal{F}}$. By construction $f(a) = \mathcal{I}_{\mathcal{A}} f(a)$ for all $a \in \mathcal{A}$, thus $f - \mathcal{I}_{\mathcal{A}} f$ is a member of the set

$$V' := \bigcap_{a \in \mathcal{A}'} \{v \in \mathcal{F} : v(a) = 0\},$$

where $\mathcal{A}'$ is a unisolvent subset of $\mathcal{A}$ of size $\ell = \dim(\mathcal{K}_{k,\mu})$, and V' itself is a Hilbert space, so it seems prudent to examine, for each $x \in \Omega$, the value of $|v(x)|$ where $v \in V'$, to ascertain whether this tells us anything about the behaviour of the interpolant. More specifically, we define what is known as the *power function* and examine its properties.

Definition 2.6.2 (Power function). *Let $x \in \Omega$. The* power function *P is given by*

$$P(x) := \sup_{v \in V'} \{|v(x)| : |v|_{\mathcal{F}} = 1\}. \tag{2.6.1}$$

Thus for any $v \in V'$ we have $P(x) \geq |v(x)|/|v|_{\mathcal{F}}$ for all $x \in \Omega$; in other words, $|v(x)| \leq P(x)|v|_{\mathcal{F}}$, and in particular,

$$|f(x) - \mathcal{I}_{\mathcal{A}} f(x)| \leq P(x)|f - \mathcal{I}_{\mathcal{A}} f|_{\mathcal{F}}. \tag{2.6.2}$$

Recall the Pythagorean property of the interpolant (2.3.3); substituting this into (2.6.2) gives an error estimate of the form

$$|f(x) - \mathcal{I}_{\mathcal{A}} f(x)| \leq P(x)|f|_{\mathcal{F}}, \qquad (x \in \Omega). \tag{2.6.3}$$

For this to be of any practical use, we obviously need to be able to say a little more about the power function P, ideally describing it in terms of (powers of) h as mentioned at the beginning of this section. With Assumption 2.3.7 in place, Light & Wayne (1998) provide the precise form of the power function.

Theorem 2.6.3. *Let $\mathcal{I}_{\mathcal{A}} f$ be the interpolant to f from $\mathcal{F}$ over a finite set of interpolation points $\mathcal{A} \subset \mathbb{R}^d$, which contains a subset $\mathcal{A}'$ that is unisolvent with respect to the kernel K. Then*

$$P(x) = \phi(0) - 2 \sum_{a \in \mathcal{A}'} p_a(x)\phi(|x - a|) + \sum_{a,b \in \mathcal{A}'} p_a(x)p_b(x)\phi(|a - b|).$$

Substituting this into (2.6.3), we have an expression for the behaviour of the error, at least in a pointwise sense, so we turn our attention to establishing stronger results, which will require more stringent demands on the domain Ω that contains our interpolation points.

2.6.2 L_p-error estimates for $BL^k(\Omega)$

In this section we prove error estimates in the L_p-norms (1.3.1) and (1.3.2) in the case of the polyharmonic spline radial basis functions with $\mu = 0$. Recall from Theorem 2.4.2 that these basis functions give rise to the integer-order Beppo Levi spaces $BL^k(\mathbb{R}^d)$ (with kernel

Π^d_{k-1}), which happily satisfy our assumptions on the general space $\mathcal{F}$, so naturally we shall be interpolating functions from the local space $BL^k(\Omega)$.

Some preliminary results arm us with properties of the sets of points with which we deal, the first of which, found in Light & Wayne (1998) for example, tells us that we may perturb a unisolvent set and still retain its unisolvency property.

Lemma 2.6.4. *Let $B(x,r) := \{y \in \mathbb{R}^d : |x-y| \leq r\}$ be the closed ball of radius r, centred at x. Let $b = (b_1, \ldots, b_\ell)$ denote an ℓ-tuple of points in $\mathbb{R}^d$ which is unisolvent with respect to Π^d_{k-1}. Then there exists $\delta > 0$ such that if $c = (c_1, \ldots, c_\ell) \in B(b_1, \delta) \times \cdots \times B(b_\ell, \delta)$, then $c_1, \ldots, c_\ell$ is also a set of points unisolvent with respect to Π^d_{k-1}.*

This next result, due to Duchon (1978), shows that our set Ω may be covered with a finite number of closed balls, provided that Ω satisfies the well-known *cone condition.*

Definition 2.6.5. *Let $\Omega \subset \mathbb{R}^d$. Ω satisfies the* cone condition *(or possesses the* cone property*) if, for each $x \in \Omega$, there exists an $\xi_x \in \mathbb{R}^d$ of unit length such that, for fixed positive ρ and θ,*

$$\left\{x + \lambda\eta : \eta \in \mathbb{R}^d,\ |\eta| = \rho,\ \eta^T \xi_x \geq \cos\theta,\ 0 \leq \lambda \leq 1\right\} \subset \Omega.$$

(Note that ρ and θ are fixed but may depend on the domain.)

A domain which satisfies the cone condition may not, in particular, contain any cusps.

Lemma 2.6.6 (Duchon (1978)). *Let Ω be an open subset of $\mathbb{R}^d$ satisfying the cone condition. Then there exist M, M_1 and $h_0 > 0$ such that to each $0 < \xi < h_0$, there corresponds a set $T_\xi \subset \Omega$ with*

(D1) $B(t,\xi) \subset \Omega$ for all $t \in T_\xi$;

(D2) $\Omega \subset \bigcup_{t \in T_\xi} B(t, M\xi)$;

(D3) $\sum_{t \in T_\xi} \chi_{B(t,M\xi)} \leq M_1$.

Here χ_W is the so-called characteristic function *which, for any set W, has value one on W and zero elsewhere.*

Assumption 2.6.7. *The following are useful properties of a set $\Omega \subset \mathbb{R}^d$:*

(P1) Ω is open, bounded and connected;

(P2) Ω has the cone property;

(P3) Ω has a Lipschitz boundary.

Assuming 2.6.7, Duchon (1978) showed that there exist extensions of functions in $BL^k(\Omega)$ to the native space of the interpolant—that is, to $BL^k(\mathbb{R}^d)$.

Lemma 2.6.8. *Let Ω satisfy Assumption 2.6.7, and let $f \in BL^k(\Omega)$. Then there exists a unique element $f^\Omega \in BL^k(\mathbb{R}^d)$ such that $f^\Omega|_\Omega = f$, and amongst all elements of $BL^k(\mathbb{R}^d)$ with this property, $|f^\Omega|_{BL^k(\mathbb{R}^d)}$ is minimal.*

Moreover, there exists a constant C_Ω, dependent on Ω, such that, for all $f \in BL^k(\Omega)$, $|f^\Omega|_{BL^k(\mathbb{R}^d)} \leq C_\Omega |f|_{BL^k(\Omega)}$.

Light & Wayne (1998) were able to adapt this result to show that $C_\Omega \equiv C$, that is, the constant is *independent* of the domain, when Ω is a ball in $\mathbb{R}^d$.

Lemma 2.6.9. *Let B be any ball in $\mathbb{R}^d$, and let $f \in BL^k(B)$. Then there exists a unique element $f^B \in BL^k(\mathbb{R}^d)$ such that $f^B|_B = f$, and amongst all elements of $BL^k(\mathbb{R}^d)$ with this property, $|f^B|_{BL^k(\mathbb{R}^d)}$ is minimal.*

Moreover, there exists a constant C, independent of B, such that, for all $f \in BL^k(B)$, $|f^B|_{BL^k(\mathbb{R}^d)} \leq C|f|_{BL^k(B)}$.

Alternative basis function for even dimensions d

In the case of an even dimension d, we may use a slightly different basis function, ψ_σ, in place of the usual polyharmonic spline, and still arrive at the same power function given in Theorem 2.6.3, and furthermore, we know the maximum value of the modulus of this basis

function, inside a ball depending on the fill distance h, in terms of h; we prove this series of results via work adapted, for clarity, from Light & Wayne (1998).

Lemma 2.6.10. *Let $\{a_0, \ldots, a_m\} \subset \mathbb{R}^d$ be any set of points unisolvent with respect to $\mathcal{K}_{k,\mu}$. Let $\alpha_0, \ldots, \alpha_m \in \mathbb{R}$ be chosen so that, for all $p \in \Pi^d_{k-1}$, $\sum_{r=0}^m \alpha_r p(a_r) = 0$, and assume $k > d/2$. Then*

$$\sum_{r,s=0}^{m} \alpha_r \alpha_s |a_r - a_s|^{2k-d} = 0.$$

Note that in the case of polyharmonic splines, $\phi(0) = 0$, so we are free to drop this term from the power function.

Lemma 2.6.11. *Let $\mathcal{A}'$ be a subset of $\mathbb{R}^d$ unisolvent with respect to Π^d_{k-1}. Let $\phi = (-1)^{(k-d/2+1)} r^{2k-d} \ln(r)$, where d is even and $k > d/2$, and let*

$$P(x) = -2 \sum_{a \in \mathcal{A}'} p_a(x) \phi(|x-a|) + \sum_{a,b \in \mathcal{A}'} p_a(x) p_b(x) \phi(|a-b|). \tag{2.6.4}$$

Define $\psi_\sigma(r) = (-1)^{(k-d/2+1)} r^{2k-d} \ln(\sigma r)$, where $\sigma > 0$, and set

$$P_\sigma(x) = -2 \sum_{a \in \mathcal{A}'} p_a(x) \psi_\sigma(|x-a|) + \sum_{a,b \in \mathcal{A}'} p_a(x) p_b(x) \psi_\sigma(|a-b|). \tag{2.6.5}$$

Then $P_\sigma(x) = P(x)$ for all $\sigma > 0$, for $x \in \mathbb{R}^d$.

Proof. Let $n_{d,k} = (-1)^{k-d/2+1}$. Since $\psi_\sigma = \phi(x) + n_{d,k} r^{2k-d} \ln(\sigma)$, we have

$$P_\sigma(x) = P(x) - n_{d,k} \left\{ 2 \sum_{a \in \mathcal{A}'} p_a(x) |x-a|^{2k-d} - \sum_{a,b \in \mathcal{A}'} p_a(x) p_b(x) |a-b|^{2k-d} \right\} \ln(\sigma). \tag{2.6.6}$$

We claim that,

$$2 \sum_{a \in \mathcal{A}'} p_a(x) |x-a|^{2k-d} - \sum_{a,b \in \mathcal{A}'} p_a(x) p_b(x) |a-b|^{2k-d} = 0. \tag{2.6.7}$$

To see this, first observe that for any $x \in \mathbb{R}^d$ we may write $p(x) = \sum_{a \in \mathcal{A}'} \alpha_a p_a(x)$, for some coefficients α_a, $a \in \mathcal{A}'$—in particular, for $x = b \in \mathcal{A}'$. Thus $p(b) = \sum_{a \in \mathcal{A}'} \alpha_a p_a(b) = \alpha_b$ since $p_a(b) = \delta_{ab}$, for all $b \in \mathcal{A}'$.

Hence $p(x) = \sum_{a \in \mathcal{A}'} p(a) p_a(x)$ for all $x \in \mathbb{R}^d$, which we can rewrite as

$$\sum_{a \in \mathcal{A}' \cup \{x\}} \alpha_a p(a) = 0,$$

where

$$\alpha_a = \begin{cases} 1, & a = x, \\ -p_a(x), & a \in \mathcal{A}'. \end{cases}$$

With this explicit definition of the coefficients α_a for $a \in \mathcal{A}' \cup \{x\}$, we may rewrite (2.6.7) as follows:

$$\begin{aligned} 2 \sum_{a \in \mathcal{A}'} p_a(x) |x - a|^{2k-d} - \sum_{a,b \in \mathcal{A}'} p_a(x) p_b(x) |a - b|^{2k-d} & \\ = -\left(2 \sum_{b \in \mathcal{A}'} \alpha_b |x - b|^{2k-d} + \sum_{a,b \in \mathcal{A}'} \alpha_a \alpha_b |a - b|^{2k-d} \right) & \\ = -\left(2 \sum_{a \in \{x\}} \sum_{b \in \mathcal{A}'} \alpha_a \alpha_b |a - b|^{2k-d} + \sum_{a,b \in \mathcal{A}'} \alpha_a \alpha_b |a - b|^{2k-d} \right) & \\ = -\left(\sum_{a,b \in \mathcal{A}' \cup \{x\}} \alpha_a \alpha_b |a - b|^{2k-d} \right). & \end{aligned}$$

Appealing to Lemma 2.6.10 in light of the coefficients α_a possessing the required property, we have that $\sum_{a,b \in \mathcal{A}' \cup \{x\}} \alpha_a \alpha_b |a - b|^{2k-d} = 0$, and hence $P_\sigma(x) = P(x)$. □

Lemma 2.6.12. *Let $\phi_{d,k,\sigma} : \mathbb{R} \to \mathbb{R}$ be defined as follows:*

$$\phi_{d,k,\sigma}(r) = \begin{cases} (-1)^{(k-d/2+1)} r^{2k-d} \ln(\sigma r), & d \text{ even}, \\ (-1)^{\lceil k-d/2 \rceil} r^{2k-d}, & d \text{ odd}. \end{cases}$$

Then

$$\max_{0\leq r\leq Ch} |\phi_{d,k,1/h}(r)| = \mathcal{O}(h^{2k-d})$$

for all $h > 0$, where C is a constant independent of h.

Proof. For the case of d odd, the result is obvious, so let d be even and set $n_{d,k} = (-1)^{k-d/2+1}$. Then clearly

$$\max_{0\leq r\leq Ch} |\phi_{d,k,\sigma}(r)| = \max\left\{n_{d,k}(Ch)^{2k-d}|\ln(\sigma Ch)|, |\phi_{d,k,\sigma}(s)|\right\},$$

where $0 \leq s \leq Ch$ is such that $\left(\frac{\mathrm{d}}{dr}\phi_{d,k,\sigma}\right)(s) = 0$. Some elementary calculus shows that $s = \frac{1}{\sigma}e^{-1/(2k-d)}$ and hence $\phi_{d,k,\sigma}(s) = K(1/\sigma)^{2k-d}$ for a suitable constant K. Putting $\sigma = 1/h$ completes the proof. □

The final result we require before proving the main result of this section is a rephrasing of Theorem 2.6.3, which we state in the integer-order Beppo Levi space case with which we are concerned.

Corollary 2.6.13. *Let $\mathcal{A}' \subset \Omega \subset \mathbb{R}^d$ be unisolvent with respect to Π^d_{k-1}, and let $g \in BL^k(\mathbb{R}^d)$ satisfy $g(a) = 0$ for all $a \in \mathcal{A}'$. Then for all $x \in \Omega$,*

$$|g(x)|^2 \leq P(x)|g|^2_{BL^k(\Omega)},$$

where the power function P is as in Theorem 2.6.3.

We are now in a position to state the main result of this section, which gives the L_p-norm error for functions in $BL^k(\Omega)$ in terms of the fill distance of the underlying interpolation points.

Theorem 2.6.14. *Let Ω be a subset of $\mathbb{R}^d$ satisfying Assumption 2.6.7, let $1 \leq p \leq \infty$ and $k > d/2$. For each $h > 0$, let $\mathcal{A}_h$ be a finite, subset of Ω, unisolvent with respect to Π^d_{k-1}, and with fill distance less than or equal to h.*

Let $f \in BL^k(\Omega)$ and $\mathcal{I}_{\mathcal{A}_h} f$ be the minimal norm interpolant to f over $\mathcal{A}_h$ from $BL^k(\mathbb{R}^d)$. Then there exists a constant $h_0 > 0$ and a constant C independent of h such that

$$\|f - \mathcal{I}_{\mathcal{A}_h} f\|_{L_p(\Omega)} \leq \begin{cases} Ch^k |f|_{BL^k(\Omega)}, & 1 \leq p \leq 2, \\ Ch^{k+d/p-d/2} |f|_{BL^k(\Omega)}, & 2 \leq p \leq \infty, \end{cases} \tag{2.6.8}$$

for all $f \in BL^k(\Omega)$ and $h \leq h_0$.

Remark 2.6.15. *The following proof of this result largely follows Light & Wayne's (1998) strategy, but we present it here with some modifications and clarifications in light of the theory as we have built it up thus far, as well as to provide insight into the general ideas that these error estimates use in their proofs.*

Proof of Theorem 2.6.14. Take $v_1, \ldots, v_\ell$ to be a set of points in $\mathbb{R}^d$ that are unisolvent with respect to Π^d_{k-1}. Then by 2.6.4, there exists a $\delta > 0$ such that if

$$(x_1, \ldots, x_\ell) \in B(v_1, \delta) \times \cdots \times B(v_\ell, \delta), \tag{2.6.9}$$

then $\{x_1, \ldots, x_\ell\}$ is a set of points unisolvent with respect to Π^d_{k-1}. Since unisolvency is unaffected by scaling, it follows that if

$$(y_1, \ldots, y_\ell) \in B\left(\frac{v_1}{\delta}, 1\right) \times \cdots \times B\left(\frac{v_\ell}{\delta}, 1\right), \tag{2.6.10}$$

then $\{y_1, \ldots, y_\ell\}$ is also a set of points unisolvent with respect to Π^d_{k-1}. Set $u_i = v_i/\delta$, $i = 1, \ldots, \ell$.

Choose $R > 0$ such that $B(u_i, 1) \subset B(0, R)$ for $i = 1, \ldots, \ell$ (note that $R > 1$). Select h_0 in accordance with Lemma 2.6.6, and fix $h > 0$ so that $Rh < h_0$. (So in the language of Lemma 2.6.6, $\xi = Rh$.) Thus we have effectively fixed the Π^d_{k-1}-unisolvent set $\mathcal{A}_h$. By Lemma 2.6.6 there exists a set $T_{Rh} \subset \Omega$ such that $B(t, Rh) \subset \Omega$ for all $t \in T_{Rh}$ and $\Omega \subset \bigcup_{t \in T_{Rh}} B(t, MRh)$ for some constant M.

Hence, fixing $x \in \Omega$, $x \in B(t, MRh)$ for some $t \in T_{Rh}$. Define

$$\sigma : B(t, MRh) \to B(0, MR)$$

by

$$\sigma(y) = h^{-1}(y - t) \qquad (y \in \mathbb{R}^d).$$

We claim that there exist ℓ points $a_1, \ldots, a_\ell \in \mathcal{A}_h$ such that $\{a_1, \ldots, a_\ell\}$ is unisolvent with respect to Π^d_{k-1}, and

$$\sigma(a_i) \in B(u_i, 1) \qquad (i = 1, \ldots, \ell).$$

To see this, note firstly that for each $i = 1, \ldots, l$, $\sigma^{-1}(B(u_i, 1))$ is a ball of radius h, and

$$\begin{aligned} |\sigma^{-1}(u_i) - t| &= |(hu_i + t) - t| \\ &= h|u_i| \\ &< hR, \end{aligned}$$

since $u_i \in B(0, R)$. Thus

$$\sigma^{-1}(B(u_i, 1)) \subset B(t, Rh) \subset \Omega \tag{2.6.11}$$

by hypothesis. Since the fill distance of $\mathcal{A}_h$ in Ω is at most h, $\sigma^{-1}(B(u_i, 1))$ must contain at least one point of A_h, and so by (2.6.11), we can choose $a_1, \ldots, a_\ell \in B(t, Rh) \cap \mathcal{A}_h$ such that $\sigma(a_i) \in B(u_i, 1)$ for $i = 1, \ldots, \ell$. Thus we have a set of ℓ points which are unisolvent with respect to Π^d_{k-1} — denote this set by $\mathcal{A}'_h$.

Let $\mathcal{I}_{\mathcal{A}_h} f^\Omega$ be the minimal norm interpolant to f^Ω over $\mathcal{A}_h$ from $BL^k(\mathbb{R}^d)$. Henceforth we denote $B(t, MRh)$ by B_t. For each $t \in T_{Rh}$, define $g_t \in BL^k(\mathbb{R}^d)$ such that

(G1) $g_t|_{B_t} = (f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega)|_{B_t}$;

(G2) $|g_t|_{BL^k(\mathbf{R}^d)} \leq C|g_t|_{BL^k(B_t)} = C|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)}$.

Note that g_t exists, and the constant C is independent of B_t, by virtue of Lemma 2.6.9.

Since $g_t(a) = 0$ for all $a \in \mathcal{A}'_h$, we can apply the error estimate in Corollary 2.6.13 to find

$$|g_t(x)|^2 \leq P(x)|g_t|^2_{BL^k(\mathbf{R}^d)} \leq P(x)C^2|g_t|^2_{BL^k(B_t)} = C^2P(x)|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|^2_{BL^k(B_t)}.$$

Hence

$$\|g_t\|_{L_p(B_t)} \leq C|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)} \left(\int_{B_t} |P(x)|^{\frac{p}{2}} \, \mathrm{d}x \right)^{\frac{1}{p}}. \tag{2.6.12}$$

We now turn our attention to dealing with the power function P. Recall from Lemma 2.6.3 that

$$P(x) = \phi(0) - 2 \sum_{a \in \mathcal{A}'} p_a(x)\phi(|x-a|) + \sum_{a,b \in \mathcal{A}'} p_a(x)p_b(x)\phi(|a-b|),$$

where here $\phi \equiv \phi_{d,k,1/h}$ is as defined in Lemma 2.6.12. Since $x \in B_t$, $|t-x| \leq MRh$, and since each $a \in B(t, Rh)$, $|t-a| \leq Rh$. Hence $|x-a|$ and $|a-b|$ are each at most $(M+1)Rh$ for all $a, b \in \mathcal{A}'_h$. Set

$$\lambda = \max\{|\phi(y)| \, : \, y \in B(0, (M+1)Rh)\},$$

$$\Gamma = \max\left\{ \sum_{a \in \mathcal{A}'_h} |p_a(y)| \, : \, y \in B(t, MRh) \right\}.$$

Then by Lemma 2.6.12, $\lambda = \mathcal{O}(h^{2k-d})$ and so

$$\begin{aligned}\int_{B_t} |P(x)|^{\frac{p}{2}}\,\mathrm{d}x &\le \int_{B_t} \Big(|\phi(0)| + 2\sum_{a\in\mathcal{A}_h'} |p_a(x)||\phi(|x-a|)| + \sum_{a,b\in\mathcal{A}_h'} |p_a(x)||p_b(x)||\phi(|a-b|)|\Big)^{\frac{p}{2}}\,\mathrm{d}x \\ &\le \lambda^{\frac{p}{2}} \int_{B_t} (1+2\Gamma+\Gamma^2)^{\frac{p}{2}}\,\mathrm{d}x \\ &= C(h^{k-d/2})^p(1+\Gamma)^p(MRh)^d \\ &= C(h^{k-d/2})^p h^d.\end{aligned}$$

Here and henceforth, C denotes an arbitrary constant independent of h. Hence from (2.6.12) we have

$$\|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega\|_{L_p(B_t)}^p = \|g_t\|_{L_p(B_t)}^p \le C\Big(h^{k+d/p-d/2} C_{B_t} |f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)}\Big)^p.$$

Now set $\Omega^* = \bigcup_{t\in T_{Rh}} B(t, MRh)$; by Lemma 2.6.6, $\Omega \subset \Omega^*$, and so

$$\begin{aligned}\|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega\|_{L_p(\Omega)}^p &\le \|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega\|_{L_p(\Omega^*)}^p \\ &= \int_{\Omega^*} |f^\Omega(x) - (\mathcal{I}_{\mathcal{A}_h} f^\Omega)(x)|^p\,\mathrm{d}x \\ &\le \sum_{t\in T_{Rh}} \int_{B_t} |f^\Omega(x) - (\mathcal{I}_{\mathcal{A}_h} f^\Omega)(x)|^p\,\mathrm{d}x \\ &= \sum_{t\in T_{Rh}} \|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega\|_{L_p(B_t)}^p \\ &\le C \sum_{t\in T_{Rh}} \Big(h^{k+d/p-d/2} |f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)}\Big)^p. \qquad (2.6.13)\end{aligned}$$

We now consider the two cases $1 \le p \le 2$ and $2 \le p \le \infty$ separately, taking the latter case first.

For any vector $v \in \mathbb{R}^m$, $\|v\|_p \le \|v\|_2$ for $p \ge 2$, and observe that

$$\sum_{t\in T_{Rh}} \Big(h^{k+d/p-d/2} |f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)}\Big)^p =: \Big\| \Big(h^{k+d/p-d/2} |f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)}\Big)_{t\in T_{Rh}} \Big\|_p^p.$$

Since $\|f - \mathcal{I}_{\mathcal{A}_h} f\|^p_{L_p(\Omega)} = \|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega\|^p_{L_p(\Omega)}$, from (2.6.13) we find

$$\begin{aligned}
\|f - \mathcal{I}_{\mathcal{A}_h} f\|^p_{L_p(\Omega)} &\leq C \left(\sum_{t \in T_{Rh}} \left(h^{k+d/p-d/2} |f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)} \right)^2 \right)^{\frac{p}{2}} \\
&= C h^{kp+d-dp/2} \left(\sum_{t \in T_{Rh}} \left(|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(B_t)} \right)^2 \right)^{\frac{p}{2}} \\
&= C h^{kp+d-dp/2} \left(\sum_{t \in T_{Rh}} \sum_{|\alpha|=k} c_\alpha \int_{B_t} |D^\alpha (f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega)|^2 \right)^{\frac{p}{2}}.
\end{aligned}$$

By introducing the characteristic function χ_{B_t}, which has the value one on B_t and is zero everywhere else, we can rewrite this as

$$\begin{aligned}
\|f - \mathcal{I}_{\mathcal{A}_h} f\|^p_{L_p(\Omega)} &\leq C h^{kp+d-dp/2} \left(\sum_{t \in T_{Rh}} \sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^d} \chi_{B_t} |D^\alpha (f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega)|^2 \right)^{\frac{p}{2}} \\
&= C h^{kp+d-dp/2} \left(\sum_{|\alpha|=k} c_\alpha \int_{\mathbb{R}^d} |D^\alpha (f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega)|^2 \sum_{t \in T_{Rh}} \chi_{B_t} \right)^{\frac{p}{2}} \\
&\leq C h^{kp+d-dp/2} \left(|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|_{BL^k(\mathbb{R}^d)} \right)^{\frac{p}{2}},
\end{aligned}$$

by once again appealing to Lemma 2.6.6, using (D3) on this occasion. Hence we have

$$\begin{aligned}
\|f - \mathcal{I}_{\mathcal{A}_h} f\|^p_{L_p(\Omega)} &\leq C h^{kp+d-dp/2} \left(|f^\Omega - \mathcal{I}_{\mathcal{A}_h} f^\Omega|^2_{BL^k(\mathbb{R}^d)} \right)^{\frac{p}{2}} \\
&\leq C h^{kp+d-dp/2} \left(|f^\Omega|^2_{BL^k(\mathbb{R}^d)} \right)^{\frac{p}{2}} \\
&\leq C h^{kp+d-dp/2} |f|^p_{BL^k(\Omega)},
\end{aligned}$$

by recalling the Pythagorean property of the interpolant (2.3.3), then using Lemma 2.6.8. Taking the p-th root, we obtain

$$\|f - \mathcal{I}_{\mathcal{A}_h} f\|_{L_p(\Omega)} \leq C h^{k+d/p-d/2} |f|_{BL^k(\Omega)} \qquad (2 \leq p \leq \infty), \tag{2.6.14}$$

as required.

Now suppose $1 \leq p \leq 2$. From (2.6.13) we know

$$\|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}\|_{L_p(\Omega)} \leq \left(C \sum_{t \in T_{Rh}} \left(h^{k+d/p-d/2} |f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^p \right)^{1/p}$$
$$= Ch^{k+d/p-d/2} \left(\sum_{t \in T_{Rh}} \left(|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^p \right)^{1/p}.$$

Let q be the number such that $p/2 + 1/q = 1$; applying Hölder's inequality for sums (which will hold for $1 \leq p \leq 2$),

$$\|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}\|_{L_p(\Omega)} \leq Ch^{k+d/p-d/2} \left\{ \left(\sum_{t \in T_{Rh}} \left(|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^2 \right)^{p/2} \left(\sum_{t \in T_{Rh}} 1^q \right)^{1/q} \right\}^{1/p}$$
$$= Ch^{k+d/p-d/2} \left(\sum_{t \in T_{Rh}} \left(|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^2 \right)^{1/2} \left(\sum_{t \in T_{Rh}} 1^q \right)^{1/pq}$$
$$= Ch^{k+d/p-d/2} \left(\sum_{t \in T_{Rh}} \left(|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^2 \right)^{1/2} h^{-d/pq}$$
$$= Ch^k \left(\sum_{t \in T_{Rh}} \left(|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}|_{BL^k(B_t)} \right)^2 \right)^{1/2}.$$

As in the case $2 \leq p \leq \infty$, we may now introduce χ_{B_t} into the above, apply (D3) of Lemma 2.6.6 and Lemma 2.6.8 to conclude

$$\|f^{\Omega} - \mathcal{I}_{\mathcal{A}_h} f^{\Omega}\|_{L_p(\Omega)} \leq Ch^k |f|_{BL^k(\Omega)} \qquad (1 \leq p \leq 2).$$

□

2.7 Local interpolation

We have seen in Section 2.2.2, that, for a set of n interpolation points $\mathcal{A}$, finding the coefficients $(\alpha, \beta)^T \in \mathbb{R}^{n+\ell}$ of the interpolant

$$\mathcal{I}_{\mathcal{A}} f = \sum_{a \in \mathcal{A}} \alpha_a \phi(|\cdot - a|) + p$$

requires inverting an $(n+\ell) \times (n+\ell)$ matrix of the form

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix}.$$

Using direct solvers, this is an order $(n+\ell)^3$ operation. But, if one reasons that, for the evaluation of the interpolant at a point $x \in \Omega$, interpolation points far away from x should not have as much influence on the value of the continuous function we are interpolating (and hence on the value of interpolant itself) as those interpolation points in a neighbourhood of x, then we can use the idea of *local interpolation* to reduce the number of operations we need to carry out.

Let $B(x, r)$ be a ball of radius r centred at the point $x \in \Omega$. Choose a *local interpolation set* $\mathcal{N} = \mathcal{N}(x) \subset \mathcal{A}$ of size N such that $\mathcal{N} \subset B(x, r)$ is in the neighbourhood of x, and is unisolvent with respect to $\mathcal{K}_{k,\mu}$, the kernel defined in (2.4.2). Then for any suitable function f, the value of the local interpolant $\mathcal{I}_{\mathcal{N}} f$ at x is given by

$$\mathcal{I}_{\mathcal{N}} f(x) = \sum_{a \in \mathcal{N}} \alpha_a \phi(|x - a|) + \sum_{i=1}^{\ell} \beta_i p_i(x), \tag{2.7.1}$$

with the coefficient vector $(\alpha, \beta)^T \in \mathbb{R}^{N+\ell}$ given by the constraints

$$\mathcal{I}_{\mathcal{N}} f(x) = f(x) \qquad (x \in \mathcal{N}), \tag{2.7.2a}$$

$$\sum_{a \in \mathcal{N}} \alpha_a p(x) = 0 \qquad (x \in \mathcal{N},\ p \in \mathcal{K}_{k,\mu}), \tag{2.7.2b}$$

which we can write in matrix form as

$$\begin{pmatrix} A_{\mathcal{N}} & P_{\mathcal{N}} \\ P_{\mathcal{N}}^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix},$$

where the local interpolation matrices are given by

$$\begin{aligned} (A_{\mathcal{N}})_{ij} &= \phi(|a_i - a_j|), && (1 \leq i, j \leq N), \\ (P_{\mathcal{N}})_{ij} &= p_j(a_i), && (1 \leq i \leq N,\ 1 \leq j \leq \ell). \end{aligned}$$

As in the global case, the local interpolant has a Lagrange basis representation, to wit:

$$\mathcal{I}_{\mathcal{N}} f(x) = \sum_{a \in \mathcal{N}} \lambda_a(x) f(a), \tag{2.7.3}$$

where

$$\lambda_a(b) = \begin{cases} 1, & a = b, \\ 0, & \text{otherwise}, \end{cases}$$

for $a, b \in \mathcal{N}$.

For problems where the number of points at which the value of the interpolant is required — call this number M — is such that

$$M(N + \ell)^3 \ll (n + \ell)^3,$$

using a local interpolation scheme is clearly advantageous. Additionally, the M interpolation problems being independent from one another allows for easy parallelisation.

2.7.1 Condition numbers and stability of polyharmonic spline local interpolation

Numerical stability of an interpolation operator is of great importance; we need to be sure that, as we refine a set of interpolation points (in other words, let the fill distance h tend to zero), the method does not become unstable. In this section we provide new stability results for local interpolation with polyharmonic splines.

The stability of the interpolation process is intrinsically linked to the eigenvalues of the interpolation matrix — specifically, the smallest and largest eigenvalues, with which we define the *condition number* of the matrix.

Definition 2.7.1 (Condition number). *For a given symmetric matrix A, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its smallest and largest eigenvalues, respectively.*

The condition number of A, *denoted* $\mathrm{cond}(A)$, *is given by*

$$\mathrm{cond}(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$

The astute reader may realise that, whilst for a positive-definite interpolation matrix A, $\lambda_{\min}(A)$ coincides with $\|A^{-1}\|_2$, if the basis function is only conditionally strictly positive-definite, we must be more careful about what we really mean by $\lambda_{\min}(A)$, for A is only positive-definite on a subset of $\mathbb{R}^d$. But we are getting ahead of ourselves, and these ideas will be explored in more detail later, by invoking the Rayleigh Quotient.

In the case of local interpolation, we keep N, the size of the local interpolation set, fixed as we refine the global set of interpolation points $\mathcal{A}$ — the net result of this is that as $h \to 0$, the radius of the ball containing our local interpolation set around a particular point $x \in \Omega$ also tends to zero. Naturally this may be cause for concern, for we are carrying out our interpolation calculations on points that are getting closer and closer together. Fortunately, in the case of local interpolation via polyharmonic splines, it turns out that the Lagrange basis functions have a particularly useful property, as shown by Iske (2003*a*).

Lemma 2.7.2. *The Lagrange basis functions of polyharmonic spline interpolation are invariant under uniform scalings.*

Let $\mathcal{W} = \{w_1, \ldots, w_N\} \subset \mathbb{R}^d$ be a fixed set of points with fill distance $h_{\mathcal{W}}$, where N is a given fixed number, and suppose that there exists an $R > 0$ such that $\mathcal{W} \subset B_R$, where B_R denotes the ball of radius R. As usual, for the interpolant over $\mathcal{W}$ to be well defined, we must demand that $\mathcal{W}$ is unisolvent with respect to $\mathcal{K}_{k,\mu}$, but we shall say a little more about the unisolvency condition in a moment, with regard to the local interpolation points $\mathcal{N}$ and the global set $\mathcal{A}$. Let $\mathcal{W}^h$ be the uniform scaling of $\mathcal{W}$ by $h > 0$. Then Lemma 2.7.2 tells us

$$(\mathcal{I}_{\mathcal{W}^h} g)(x) = (\mathcal{I}_{\mathcal{W}}(g \circ \sigma^h))(\sigma^h(x)) \tag{2.7.4}$$

for an arbitrary suitable continuous function $g : \mathbb{R}^d \to \mathbb{R}$, where σ^h is the scaling such that $\sigma^h(\mathcal{W}^h) = \mathcal{W}$. Thus, in order to prove the stability of the interpolant over $\mathcal{W}^h$ as $h \to 0$, it suffices to show that the interpolation operator is bounded over the "reference set" $\mathcal{W}$. In other words, we need to find a bound on $\|\mathcal{I}_{\mathcal{W}}\hat{g}\|_{\infty,B_R} := \max_{y \in B_R} |(\mathcal{I}_{\mathcal{W}}\hat{g})(y)|$ in terms of $\|\hat{g}\|_{L_\infty(B_R)}$, where $\hat{g} := g \circ \sigma^h$.

Before we proceed we must be a little more careful about the relationship between any given set of local interpolation points $\mathcal{N}$ and the reference set $\mathcal{W}$.

Defining $P_{\mathcal{W}}$ analogously to $P_{\mathcal{N}}$—that is, $(P_{\mathcal{W}})_{ij} = p_j(a_i)$, $a_i \in \mathcal{N}$, $1 \le i \le N$, $1 \le j \le \ell$, the quantity $\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty$ can be thought of as measuring the unisolvency of the set $\mathcal{W}$ with respect to the kernel. In two dimensions, the unisolvency condition only requires that the points must not all lie on a straight line. Numerical experiments indicate that as the points in $\mathcal{W}$ approach a straight line, $\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \to \infty$, and conversely, as the points become more scattered away from a straight line, $\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \to \mathcal{O}(1)$.

As such, in two dimensions we can draw parallels between $\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty$ and the existence of a minimum angle θ of a triangle with vertices in $\mathcal{W}$: as θ approaches zero, the interpolation points must be approaching a straight line.

For any given h (the *global* fill distance of $\mathcal{A}$), we choose sets of local interpolation points $\mathcal{N} = \mathcal{N}(x)$ of size N, for each $x \in \Omega$ at which we require the value of the local interpolant. Whilst it would be unreasonable to assume that every $\mathcal{N}$ could be mapped via an affine transformation to the reference set $\mathcal{W}$, recall from Lemma 2.6.4 that we may perturb $\mathcal{W}$ and still retain its unisolvency property, so it is more reasonable to assume that each local set can be mapped to a *perturbation* of the reference set.

With these points in mind, we investigate the stability of the interpolation operator in the special case of $\phi(r) = r^2 \ln r$, the so-called *thin-plate spline*.

2.7.2 Proof of stability of the thin-plate spline local interpolant

Let $\mathcal{A} \subset \Omega$ and $\{w_1, \ldots, w_N\} \subset B_R$ be sets of points in $\mathbb{R}^2$, of size n and N, and positive fill distance h and $h_{\mathcal{W}}$, respectively. Assume both sets are unisolvent with respect to Π_1^2 (the space of real-valued linear polynomials in two variables). Let $\delta > 0$ be such that, for any set of points $\mathcal{W} \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta)$, $\mathcal{W}$ is unisolvent with respect to Π_1^2.

Let $x \in \Omega$ and assume that for every $\mathcal{N}(x)$, there exists an affine transformation $\mathcal{N}(x) \to \mathcal{W} = \mathcal{W}(x) \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta)$ and moreover, there exists an angle θ such that every triangle with vertices in $B(w_1, \delta) \times \cdots \times B(w_N, \delta)$ has minimum angle greater than or equal to θ.

Before going on to prove our stability result for the thin-plate spline case, we must first deduce a bound on the vector $\alpha \in \mathbb{R}^{|\mathcal{W}|}$ of the local interpolation problem (2.7.1)–(2.7.2), for an arbitrary $\mathcal{W} \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta)$.

Bounding $\|\alpha\|_\infty$

From our assumptions on the unisolvency of $\mathcal{W}$ and the conditionally strict positive-definiteness of ϕ, there exists a unique solution $(\alpha, \beta)^T \in \mathbb{R}^{N+\ell}$ to (2.7.1)–(2.7.2). Moreover, we have that $\ker(P_{\mathcal{W}}^T)$ is non-trivial and, defining $A_{\mathcal{W}}$ equivalently to $A_{\mathcal{N}}$,

$$\alpha^T A_{\mathcal{W}} \alpha > 0 \qquad (0 \neq \alpha \in \ker P_{\mathcal{W}}^T), \tag{2.7.5}$$

Given that $A_{\mathcal{W}}$ is symmetric, for all $0 \neq \xi \in \mathbb{R}^N$, the following Rayleigh Quotient inequalities hold:

$$\lambda_{\min} \leq \frac{\xi^T A_{\mathcal{W}} \xi}{\xi^T \xi} \leq \lambda_{\max},$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the smallest and largest eigenvalues of $A_{\mathcal{W}}$ (Golub & Van Loan 1996). In particular this holds for all $\alpha \in \ker P_{\mathcal{W}}^T$, thus

$$\lambda_{\min} \|\alpha\|_2^2 \leq \alpha^T A_{\mathcal{W}} \alpha \leq \lambda_{\max} \|\alpha\|_2^2,$$

which immediately displays the importance of the smallest eigenvalue.

By (2.7.5) and the fact that $\|\cdot\|_2 \geq 0$, there must be an eigenvalue $\lambda \geq \lambda_{\min}$ with $\lambda > 0$. Setting $\Lambda = \lambda_{\max}$, we have

$$0 < \lambda \|\alpha\|_2^2 \leq \alpha^T A_{\mathcal{W}} \alpha \leq \Lambda \|\alpha\|_2^2, \tag{2.7.6}$$

for $0 \neq \alpha \in \ker P_{\mathcal{W}}^T$.

The interpolation problem is given by

$$A_{\mathcal{W}} \alpha + P_{\mathcal{W}} \beta = \bar{\hat{g}}, \tag{2.7.7a}$$

$$P_{\mathcal{W}}^T \alpha = 0, \tag{2.7.7b}$$

where $\bar{\hat{g}} = \hat{g}|_{\mathcal{W}}$. Thereby, pre-multiplying (2.7.7a) by α^T gives

$$\alpha^T A_{\mathcal{W}} \alpha + \alpha^T P_{\mathcal{W}} \beta = \alpha^T \bar{\hat{g}},$$

that is,

$$\alpha^T A_{\mathcal{W}} \alpha + (P_{\mathcal{W}}^T \alpha)^T \beta = \alpha^T \bar{\hat{g}}.$$

Hence, employing (2.7.7b), we get

$$\alpha^T A_{\mathcal{W}} \alpha = \alpha^T \bar{\hat{g}}.$$

Using (2.7.6), together with the Cauchy-Schwarz inequality, gives

$$\lambda \|\alpha\|_2^2 \leq \alpha^T \bar{\hat{g}} \leq \|\alpha\|_2 \|\bar{\hat{g}}\|_2;$$

assuming $\alpha \neq 0$, we have

$$\|\alpha\|_\infty \leq \|\alpha\|_2 \leq \lambda^{-1} \|\bar{\hat{g}}\|_2 \leq \lambda^{-1} \sqrt{N} \|\bar{\hat{g}}\|_\infty = \lambda^{-1} \sqrt{N} \|\hat{g}\|_{\infty,\mathcal{W}}. \tag{2.7.8}$$

A lower bound for λ

In order to specify a lower bound for λ, we firstly recall the classical Γ-function.

Definition 2.7.3 (Γ-function). *The Γ-function is defined by*

$$\Gamma(z) = \lim_{n \to \infty} \frac{n! n^z}{z(z+1)\cdots(z+n)},$$

for $z \in \mathbb{C}$.

The Γ-function has many useful properties and details may be found in the book of Lebedev (1965), for example, but the most relevant to our application is

$$\Gamma(n) = (n-1)! \qquad (n \in \mathbb{N}).$$

Defining the two constants,

$$M_d = 12 \left(\frac{\pi \Gamma^2(\frac{d}{2}+1)}{9} \right)^{\frac{1}{d+1}} \quad \text{and} \quad C_d = \frac{1}{2\Gamma(\frac{d}{2}+1)} \left(\frac{M_d}{2^{3/2}} \right)^d,$$

a lower bound for λ is found in the recent book by Wendland (2005, page 214).

Lemma 2.7.4. *Under the foregoing assumptions on the set of points $\mathcal{W}$, the following lower bound on λ holds:*

$$\lambda \geq C_d c_{d,k,\mu} (2M_d)^{-2k-2\mu} h_{\mathcal{W}}^{2k+2\mu-d}, \tag{2.7.9}$$

where

$$c_{d,k,\mu} = \begin{cases} 2^{2k+2\mu-d/2-1}\Gamma(k+\mu)(k+\mu-d/2)!, & 2k+2\mu-d \text{ even}, \\ (-1)^{\lceil k+\mu-d/2 \rceil} 2^{2k+2\mu-d/2}\Gamma(k+\mu)/\Gamma((d/2-k-\mu)), & \text{otherwise.} \end{cases}$$

Stability theorem for two dimensions

Theorem 2.7.5. *Under the foregoing assumptions on the sets $\mathcal{N}(x)$, $x \in \Omega$, and the reference set $\mathcal{W}$,*

$$\|\mathcal{I}_{\mathcal{N}(x)} g\|_{L_\infty(B(x,r))} \leq C_{N,R,h_{\mathcal{W}},\theta} \|g\|_{\infty,\mathcal{N}(x)}, \tag{2.7.10}$$

where

$$C_{N,R,\theta,h_{\mathcal{W}}} = 1 + 2D_{N,R,h_{\mathcal{W}}} + \left(1 + D_{N,R,h_{\mathcal{W}}}\right) \frac{4R\sqrt{2}}{h_{\mathcal{W}} \sin\theta}, \tag{2.7.11}$$

with

$$D_{N,R,h_{\mathcal{W}}} = 128 \frac{9^{1/3}\pi^{2/3}N^{3/2}\Phi_R}{h_{\mathcal{W}}^2} \qquad \text{and} \qquad \Phi_R = \max_{0 \leq y \leq 2R} |\phi(y)|.$$

Proof. Using the definition of the local thin-plate spine interpolant, and the invariance of the Lagrange basis under uniform scalings,

$$\begin{aligned} \|\mathcal{I}_{\mathcal{N}(x)} g\|_{L_\infty(B(x,r))} = \|\mathcal{I}_{\mathcal{W}} \hat{g}\|_{L_\infty(B_R)} &= \max_{y \in B_R} \left| \sum_{w \in \mathcal{W}} \alpha_w \phi(|y-w|) + \sum_{j=1}^{\ell} \beta_j p_j(y) \right| \\ &\leq N \|\alpha\|_\infty \Phi_R + \max_{y \in B_R} \left| \sum_{j=1}^{\ell} \beta_j p_j(y) \right|, \end{aligned} \tag{2.7.12}$$

where $\Phi_R = \max_{0\le y\le 2R}|\phi(y)|$. Here, $\ell = 3$ and for $y = (\mu, \nu)$,

$$p_1(y) = 1, \quad p_2(y) = \mu, \quad p_3(y) = \nu.$$

Let $p(y) = \sum_{j=1}^{3} \beta_j p_j(y)$, and observe that bounding $\max\{|p(y)| : y \in B_R\}$ from above is equivalent to bounding the size of the plane that $p(y)$ describes as y ranges over values in B_R.

We need only three points that lie within it to completely describe the plane, and naturally we choose the points which form the triangle with minimum angle θ; suppose, for convenience, they are labelled $w_1 = (s_1, t_1), w_2 = (s_2, t_2)$ and $w_3 = (s_3, t_3)$. Then we can write the interpolation problem, $P_{\mathcal{W}}\beta = \bar{\bar{g}} - A_{\mathcal{W}}\alpha$, as

$$\begin{aligned}
\beta_1 + \beta_2 s_1 + \beta_3 t_1 &= \hat{g}_1 - \sum_{w\in\mathcal{W}} \alpha_w \phi(|w_1 - w|) =: z_1, \\
\beta_1 + \beta_2 s_2 + \beta_3 t_2 &= \hat{g}_2 - \sum_{w\in\mathcal{W}} \alpha_w \phi(|w_2 - w|) =: z_2, \\
\beta_1 + \beta_2 s_3 + \beta_3 t_3 &= \hat{g}_3 - \sum_{w\in\mathcal{W}} \alpha_w \phi(|w_3 - w|) =: z_3.
\end{aligned}$$

Without loss of generality we can translate the three points w_1, w_2, and w_3 so that w_1 lies at the origin and consider the plane passing through

$$a = (0, 0, z_1), b = (s_2, 0, z_2), c = (s_3, t_3, z_3),$$

where $s_3/t_3 = \tan\theta$; by assumption, the triangle formed by these three points has sides of length at least $2h_{\mathcal{W}}$ and has minimum angle θ.

Taking the vector product $(b - a) \times (c - a)$ we produce a normal vector v to the plane,

$$v = (-t_3(z_2 - z_1), s_3(z_2 - z_1) - s_2(z_3 - z_1), s_2 t_3),$$

and thus the vector equation of the plane is given by $(\mu, \nu, \eta) \cdot v = (0, 0, z_1) \cdot v$, where $\mu, \nu, \eta \in \mathbb{R}$—remember we are thinking of $\eta = p((\mu, \nu))$.

Expanding this gives

$$\mu t_3(z_1 - z_2) + \nu(s_3(z_2 - z_1) - s_2(z_3 - z_1)) + \eta s_2 t_3 = z_1 s_2 t_3,$$

and rearranging gives us the Cartesian equation of the plane,

$$\eta = -\frac{\mu t_3(z_1 - z_2) + \nu(s_3(z_2 - z_1) - s_2(z_3 - z_1))}{s_2 t_3} + z_1.$$

Observe that for $(\mu, \nu) \in B_R$,

$$\max_{(\mu,\nu)\in B_R} |\eta| \leq |\nabla\eta| 2R + |z_1|. \tag{2.7.13}$$

To this end, we compute the size of the gradients of the plane along the μ and ν directions. Firstly,

$$\left|\frac{\partial\eta}{\partial\mu}\right| = \frac{z_1 - z_2}{s_2} \leq \frac{2\max_{i=1,2,3}|z_i|}{2h_{\mathcal{W}}} = \frac{\max_{i=1,2,3}|z_i|}{h_{\mathcal{W}}}.$$

In the ν direction,

$$\begin{aligned}\left|\frac{\partial\eta}{\partial\nu}\right| &= \frac{s_3(z_2 - z_1) - s_2(z_3 - z_1)}{s_2 t_3} \\ &\leq \frac{2\max_{i=1,2,3}|z_i|}{s_2|\tan\theta|} + \frac{2\max_{i=1,2,3}|z_i|}{t_3}.\end{aligned}$$

Now $\frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta} \leq \frac{1}{\sin\theta}$, and $\sin\theta \leq \frac{t_3}{2h_{\mathcal{W}}}$. Therefore

$$\left|\frac{\partial\eta}{\partial\nu}\right| \leq \frac{2\max_{i=1,2,3}|z_i|}{h_{\mathcal{W}}\sin\theta}.$$

The maximum gradient is thus

$$|\nabla\eta| = \left|\left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)\right|$$

$$
\begin{aligned}
&= \left(\left| \frac{\partial z}{\partial x} \right|^2 + \left| \frac{\partial z}{\partial y} \right|^2 \right)^{\frac{1}{2}} \\
&\leq \sqrt{2} \max \left\{ \left| \frac{\partial z}{\partial x} \right|, \left| \frac{\partial z}{\partial y} \right| \right\} \\
&= \frac{2\sqrt{2} \max_{i=1,2,3} |z_i|}{h_{\mathcal{W}} \sin \theta}.
\end{aligned}
$$

Thereby from (2.7.13),

$$
\max_{y \in B_R} |p(y)| = \max_{(\mu,\nu) \in B_R} |\eta| \leq 2R \frac{2\sqrt{2} \max_{i=1,2,3} |z_i|}{h_{\mathcal{W}} \sin \theta} + \max_{i=1,2,3} |z_i| = C_{h_{\mathcal{W}},\theta,R} \max_{i=1,2,3} |z_i|,
$$

where

$$
C_{h_{\mathcal{W}},\theta,R} = 1 + \frac{4R\sqrt{2}}{h_{\mathcal{W}} \sin \theta}.
$$

But by definition, for $i = 1, 2, 3$, we have

$$
\begin{aligned}
|z_i| &= \left| \hat{g}_i - \sum_{w \in \mathcal{W}} \alpha_w \phi(|w_i - w|) \right| \\
&\leq |\hat{g}_i| + N \Phi_R \|\alpha\|_\infty \\
&\leq \|\hat{g}\|_{\infty,\mathcal{W}} + N \Phi_R \|\alpha\|_\infty,
\end{aligned}
$$

the last step employing the interpolation conditions. Thus

$$
\max_{y \in B_R} |p(y)| \leq C_{h_{\mathcal{W}},\theta,R} \Big(\|\hat{g}\|_{\infty,\mathcal{W}} + N \Phi_R \|\alpha\|_\infty \Big).
$$

We have an upper bound on $\|\alpha\|_\infty$ via (2.7.8) and Lemma 2.7.4:

$$
\begin{aligned}
\|\alpha\|_\infty \leq \lambda^{-1} \sqrt{N} \|\hat{g}\|_{\infty,\mathcal{W}} &\leq \frac{\sqrt{N}}{C_d c_{d,k,\mu} (2M_d)^{-2k} h_{\mathcal{W}}^{2k-d}} \|\hat{g}\|_{\infty,\mathcal{W}} \\
&= \frac{(2M_2)^4 \sqrt{N}}{C_2 c_{2,2,0} h_{\mathcal{W}}^2} \|\hat{g}\|_{\infty,\mathcal{W}}
\end{aligned}
$$

in this case since $d = k = 2$ and $\mu = 0$.

Substituting this new bound on $\|\alpha\|_\infty$ and the above bound on $\max_{y\in B_R}|p(y)|$ into (2.7.12),

$$\begin{aligned}\|\mathcal{I}_{\mathcal{W}}\hat{g}\|_{L_\infty(B_R)} &\leq N\|\alpha\|_\infty\Phi_R + \max_{y\in B_R}|p(y)| \\ &\leq N\|\alpha\|_\infty\Phi_R + C_{h_{\mathcal{W}},\theta,R}(\|\hat{g}\|_{\infty,\mathcal{W}} + N\|\alpha\|_\infty\Phi_R) \\ &= (1+C_{h_{\mathcal{W}},\theta,R})N\|\alpha\|_\infty\Phi_R + C_{h_{\mathcal{W}},\theta,R}\|\hat{g}\|_{\infty,\mathcal{W}} \\ &\leq \left((1+C_{h_{\mathcal{W}},\theta,R})\frac{(2M_2)^4N^{\frac{3}{2}}\Phi_R}{C_2c_{2,2,0}h_{\mathcal{W}}^2} + C_{h_{\mathcal{W}},\theta,R}\right)\|\hat{g}\|_{\infty,\mathcal{W}}.\end{aligned}$$

Defining the constant $C_{N,R,\theta,h_{\mathcal{W}}}$ to be the coefficient in parentheses on the right-hand side, some elementary calculation using the definitions of M_d, C_d and $c_{d,k,\mu}$ gives us explicitly that

$$C_{N,R,\theta,h_{\mathcal{W}}} = 1 + 2D_{N,R,h_{\mathcal{W}}} + \left(1 + D_{N,R,h_{\mathcal{W}}}\right)\frac{4R\sqrt{2}}{h_{\mathcal{W}}\sin\theta}, \tag{2.7.14}$$

where

$$D_{N,R,h_{\mathcal{W}}} = 128(9^{\frac{1}{3}}\pi^{\frac{2}{3}}N^{\frac{3}{2}}\Phi_R)/h_{\mathcal{W}}^2.$$

Recalling that

$$\|\mathcal{I}_{\mathcal{W}}\hat{g}\|_{L_\infty(B_R)} = \|\mathcal{I}_{\mathcal{N}(x)}g\|_{L_\infty(B(x,r))} \quad \text{and} \quad \|\hat{g}\|_{\infty,\mathcal{W}} = \|g\|_{\infty,\mathcal{N}(x)}$$

completes the proof. □

Hence for fixed $h_{\mathcal{W}}$, N and R, in the two-dimensional case we can see from (2.7.14) precisely how the angle θ affects the stability of the interpolant. Moreover, for any given local interpolation set of size N in two dimensions, $h_{\mathcal{W}} \sim 2R/N^{\frac{1}{2}}$. Thus, fixing R and θ, we may think of the stability constant here as being

$$C_{N,R,\theta,h_{\mathcal{W}}} = 1 + \mathcal{O}(N^3), \tag{2.7.15}$$

which will be particularly pertinent to the analysis of the advection scheme in the next chapter. The next section extends the two-dimensional stability result to the general case in $\mathbb{R}^d$, $d \geq 1$.

2.7.3 Proof of stability of the general polyharmonic spline local interpolant

We state our general assumption concerning the scattering of the interpolation points as follows:

Assumption 2.7.6. *Let $\mathcal{A} \subset \Omega$ and $\{w_1, \ldots, w_N\} \subset B_R$ be sets of points in $\mathbb{R}^d$, of size n and N, and positive fill distance h and $h_{\mathcal{W}}$, respectively. Choose $k \geq 1$ and $0 \leq \mu < 1$ and assume both sets of points are unisolvent with respect to $\mathcal{K}_{k,\mu}$. Let $\delta > 0$ be such that, for any set of points $\mathcal{W} \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta)$, $\mathcal{W}$ is unisolvent with respect to $\mathcal{K}_{k,\mu}$.*

Fix $r > 0$. For each $x \in \Omega$, let $\mathcal{N} = \mathcal{N}(x) \subset B(x, r)$ be a subset of $\mathcal{A}$ of size N, and assume that there exists an affine transformation $\sigma_x^h : \mathcal{N} \to B(w_1, \delta) \times \cdots \times B(w_N, \delta)$.

We assume that

$$\max_{x \in \Xi} \left\{ \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \, : \, \mathcal{W} = \sigma_x^h(\mathcal{N}) \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta) \right\}$$

is uniformly bounded above.

Let $\Phi_R = \max_{0 \leq y \leq 2R} |\phi(|y|)|$. Recall that we are writing $\hat{g} = g \circ \sigma_x^h$. By definition we have that

$$\begin{aligned} \|\mathcal{I}_{\mathcal{W}} \hat{g}\|_{L_\infty(B_R)} &= \max_{y \in B_R} \left| \sum_{a \in \mathcal{N}} \alpha_a \phi(|y - a|) + \sum_{j=1}^{\ell} \beta_j p_j(y) \right| \\ &\leq \sum_{a \in \mathcal{N}} |\alpha_a| \Phi_R + \max_{y \in B_R} \sum_{j=1}^{\ell} |\beta_j| |p_j(y)| \\ &\leq N \Phi_R \|\alpha\|_\infty + C \|\beta\|_\infty, \end{aligned} \tag{2.7.16}$$

where

$$C = \max_{y \in B_R} \sum_{j=1}^{\ell} |p_j(y)|.$$

We remark that the bounds shown for $\|\alpha\|_\infty$ in (2.7.8) and Lemma 2.7.4 hold for general k and d, so we now proceed to determine a general upper bound on $\|\beta\|_\infty$.

Bounding $\|\beta\|_\infty$

We observe that $\beta^T P_{\mathcal{W}}^T P_{\mathcal{W}} \beta = (P_{\mathcal{W}}\beta)^T \cdot (P_{\mathcal{W}}\beta) = \|P_{\mathcal{W}}\beta\|_2^2 \geq 0$, which implies that the quadratic form $\beta^T P_{\mathcal{W}}^T P_{\mathcal{W}} \beta$ is positive semi-definite. In particular, $\beta^T P_{\mathcal{W}}^T P_{\mathcal{W}} \beta = 0$ if and only if $\|P_{\mathcal{W}}\beta\|_2 = 0$, i.e., when $P_{\mathcal{W}}\beta = 0$. Since $\mathcal{W}$ is unisolvent with respect to $\mathcal{K}_{k,\mu}$, $P_{\mathcal{W}}\beta = 0$ implies that $\beta = 0$, in which case $P_{\mathcal{W}}^T P_{\mathcal{W}}$ is positive-definite, and hence invertible.

With a simple rewriting of (2.7.7a), we have that

$$\begin{aligned}
\|\beta\|_\infty &= \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1} P_{\mathcal{W}}^T (\bar{\hat{g}} - A_{\mathcal{W}}\alpha)\|_\infty \\
&\leq \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty (\|\bar{\hat{g}}\|_\infty + \|A_{\mathcal{W}}\|_\infty \|\alpha\|_\infty) \\
&\leq \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty (\|\bar{\hat{g}}\|_\infty + \|A_{\mathcal{W}}\|_\infty \lambda^{-1}\sqrt{N}\|\bar{\hat{g}}\|_\infty) \\
&= \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty (1 + \|A_{\mathcal{W}}\|_\infty \lambda^{-1}\sqrt{N}) \|\bar{\hat{g}}\|_\infty,
\end{aligned}$$

by the Cauchy-Schwarz inequality and the application of (2.7.8), where $\bar{\hat{g}} = \hat{g}|_{\mathcal{W}}$. Since $A_{\mathcal{W}}$ is symmetric, $\|A_{\mathcal{W}}\|_2 = \Lambda$, where Λ is the largest eigenvalue of $A_{\mathcal{W}}$, and thus $\|A_{\mathcal{W}}\|_\infty \leq \sqrt{N}\Lambda$, so

$$\begin{aligned}
\|\beta\|_\infty &\leq \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty (1 + N\Lambda\lambda^{-1}) \|\bar{\hat{g}}\|_\infty \\
&= \|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty (1 + N\Lambda\lambda^{-1}) \|\hat{g}\|_{\infty,\mathcal{W}}.
\end{aligned}$$

Hence, from (2.7.8), (2.7.16) and the above bound on $\|\beta\|_\infty$, we get

$$\begin{aligned}\|\mathcal{I}_{\mathcal{W}}\hat{g}\|_{L_\infty(B_R)} &\leq N\Phi_R\|\alpha\|_\infty + C\|\beta\|_\infty \\ &\leq N^{\frac{3}{2}}\Phi_R\lambda^{-1}\|\hat{g}\|_{\infty,\mathcal{W}} + E(1+N\Lambda\lambda^{-1})\|\hat{g}\|_{\infty,\mathcal{W}}, \end{aligned} \tag{2.7.17}$$

where $E = C\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty \|P_{\mathcal{W}}^T\|_\infty$.

As noted by Wendland (2005, page 203), an upper bound for Λ is provided as a consequence of Gerschgorin's theorem:

$$\Lambda \leq N \max_{(x,y)\in B_R\times B_R} |\phi(|x-y|)| \leq N\Phi_R. \tag{2.7.18}$$

Finally, we note that we can re-scale via the inverse affine transformation $(\sigma_x^h)^{-1} : \mathcal{W} \to \mathcal{N}$ using of the invariance of the interpolant under this operator, and so by substituting the bounds (2.7.9) and (2.7.18) for λ and Λ, respectively, into (2.7.17) we deduce the following new stability result.

Theorem 2.7.7. *Under the foregoing assumptions on the set of points $\mathcal{W}$ and the neighbour sets $\mathcal{N}(x)$, $x \in \Omega$, for any continuous function g defined on $\mathbb{R}^d$, the following stability bound holds:*

$$\|\mathcal{I}_{\mathcal{N}(x)}g\|_{L_\infty(B(x,r))} = \|\mathcal{I}_{\mathcal{W}}\hat{g}\|_{L_\infty(B_R)} \leq C_S\|\hat{g}\|_{\infty,\mathcal{W}} = C_S\|g\|_{\infty,\mathcal{N}(x)}$$

for all $x \in \Omega$, where $B(x,r)$ denotes the ball containing $\mathcal{N}(x)$, and

$$C_S = E + D\Big(1 + EN^{\frac{1}{2}}\Big), \qquad D = \frac{N^{\frac{3}{2}}\Phi_R}{C_d c_{k,d,\mu}(2M_d)^{-2k-2\mu}h_{\mathcal{W}}^{2k+2\mu-d}},$$

$$E = C\|(P_{\mathcal{W}}^T P_{\mathcal{W}})^{-1}\|_\infty\|P_{\mathcal{W}}^T\|_\infty, \quad C = \max_{x\in B_R}\sum_{j=1}^{\ell}|p_j(x)|.$$

Remark 2.7.8. *We saw from* (2.7.15) *that for the two-dimensional thin-plate spline case, the constant $C_S \equiv C_{N,R,\theta,h_{\mathcal{W}}} = 1 + \mathcal{O}(N^3)$. We shall assume in general that, for a fixed*

radius R of the ball containing $\mathcal{W}$, $C_S = 1+\mathcal{O}(N^\gamma)$ for some $\gamma \geq 1$ for all $d \geq 1$. The proof that this assumption is reasonable in the general case is the subject of intended future work.

2.7.4 Error estimates for local interpolation with polyharmonic splines

If a certain degree of smoothness of the interpoland is assumed, one can show that this smoothness is carried over to the order of convergence of the interpolation procedure by way of a Taylor series argument.

Theorem 2.7.9. *Let $\mathcal{A} \subset \Omega$ have fill distance $h > 0$. Choose $x \in \Omega$ and let $\mathcal{N} = \mathcal{N}(x) \subset B(x,r)$ satisfy Assumption 2.7.6. Choose $k \geq 1$ and $0 \leq \mu < 1$ so that $k + \mu > d/2$ and $k \geq m$, and let $\mathcal{I}_\mathcal{N} g$ be the local polyharmonic spline interpolant from $BL^{k+\mu}(\mathbb{R}^d)$ to a continuous function g over $\mathcal{N}$. Then, assuming that $g \in C^m(B(x,r))$, for every $y \in B(x,r)$ the following approximation result holds:*

$$|g(y) - \mathcal{I}_\mathcal{N} g(y)| \leq C_T h^m |g|_{W^m_\infty(B(x,r))},$$

where C_T is a non-negative constant, independent of h, and,

$$|g|_{W^m_\infty(B(x,r))} := \sup_{z \in B(x,r)} \sum_{|\alpha|=m} |D^\alpha g(z)|.$$

Remark 2.7.10. *Our proof largely follows that of Iske (2003a), but we need to be more careful due to the assumptions on the distribution of our interpotion points (cf. Assumption 2.7.6).*

Proof of Theorem 2.7.9. Observe that g permits the m-th order Taylor polynomial,

$$T^m_{g,y}(z) = \sum_{|\alpha|<m} \frac{1}{\alpha!}(D^\alpha g)(y)(z-y)^\alpha,$$

around $y \in B(x,r)$. This equation can be rewritten to reveal

$$g(y) = T^m_{g,y}(z) - \sum_{\substack{|\alpha|<m,\\ |\alpha|\neq 0}} \frac{1}{\alpha!}(D^\alpha g)(y)(z-y)^\alpha,$$

for all $z \in B(x,r)$, which holds in particular for $z = a \in \mathcal{N}$.

The assumption $k + \mu > d/2$ ensures that the interpolant is at least continuous, so that point evaluation makes sense. The further hypothesis that $k \geq m$ results in polynomial reproduction for, in particular, $\Pi^d_{m-1} \subset \mathcal{K}_{k,\mu}$. As explained in Section 2.5, we may write

$$\mathcal{I}_{\mathcal{N}} g(y) = \sum_{a \in \mathcal{N}} \lambda_a(y) g(a),$$

where the Lagrange basis functions λ_a, $a \in \mathcal{N}$, satisfy $\lambda_a(b) = \delta_{ab}$ for $a, b \in \mathcal{N}$, where δ_{ab} denotes the Kronecker delta function. Moreover, since interpolation reproduces polynomials, in particular the constant polynomial $p = 1$, we have

$$1 = \sum_{a \in \mathcal{N}} \lambda_a(y).1 = \sum_{a \in \mathcal{N}} \lambda_a(y),$$

for all $y \in B(x,r)$. Thus, for any $y \in B(x,r)$,

$$\begin{aligned} g(y) = \sum_{a \in \mathcal{N}} \lambda_a(y) g(y) &= \sum_{a \in \mathcal{N}} \lambda_a(y) \Big(T^m_{g,y}(a) - \sum_{\substack{|\alpha| < m, \\ |\alpha| \neq 0}} \frac{1}{\alpha!} D^\alpha g(y) (a-y)^\alpha \Big) \\ &= \sum_{a \in \mathcal{N}} \lambda_a(y) T^m_{g,y}(a), \end{aligned} \tag{2.7.19}$$

since, by the assumed smoothness of the interpolant, $\sum_{a \in \mathcal{N}} \lambda_a(y)(a-y)^\alpha = (x-y)^\alpha|_{x=y} = 0$ for $|\alpha| < m$. Hence, we deduce that

$$g(y) - \mathcal{I}_{\mathcal{N}} g(y) = \sum_{a \in \mathcal{N}} \lambda_a(y) \big(T^m_{g,y}(a) - g(a) \big), \tag{2.7.20}$$

for any $y \in B(x,r)$.

Now, it is well known that the remainder term of the Taylor expansion is given by

$$T^m_{g,y}(a) - g(a) = \frac{1}{m!} \sum_{|\alpha| = m} (D^\alpha g)(\xi_a)(y-a)^\alpha, \tag{2.7.21}$$

for all $a \in \mathcal{N}$, where ξ_a is some point on the line segment connecting y and a.

Observing that $|y - a| \leq h$, gives

$$\begin{aligned} |T^m_{g,y}(a) - g(a)| &= \left| \frac{1}{m!} \sum_{|\alpha|=m} (D^\alpha g)(\xi_a)(y-a)^\alpha \right| \\ &\leq Ch^m \sum_{|\alpha|=m} |(D^\alpha g)(\xi_a)| \\ &\leq Ch^m |g|_{W^m_\infty(B(x,r))}, \end{aligned}$$

for all $a \in \mathcal{N}$. Defining λ_w for $w \in \mathcal{W}$ analogously to λ_a, $a \in \mathcal{N}$, it can be shown, following ideas from Light & Wayne (1998) for example, that because of the scale-invariance of the Lagrange basis functions, $\sum_{a\in\mathcal{N}} |\lambda_a| = \sum_{w\in\mathcal{W}} |\lambda_w|$, and moreover, this quantity is uniformly bounded over all $\mathcal{W} \subset B(w_1, \delta) \times \cdots \times B(w_N, \delta)$, as $h \to 0$; hence, from (2.7.20), we get

$$|g(y) - \mathcal{I}_\mathcal{N} g(y)| \leq C_T h^m |g|_{W^m_\infty(B(x,r))},$$

as required. □

2.8 Hermite-Birkhoff interpolation

Up to this point we have been considering the recovery of a function using the value of the function at certain points; a natural question to ask is whether we can perform any sort of recovery of a function if only more general information is known — not necessarily the value of the function, but values prescribed by the action of a functional on the function.

We pose the generalised problem of recovering a function from a Hilbert function space $\mathcal{F}$ as follows. Suppose that $\lambda_1, \ldots, \lambda_n$ are a set of linearly independent functionals on $\mathcal{F}$, and let $u_1, \ldots, u_n \in \mathbb{R}$ be corresponding given values for $u \in \mathcal{F}$ — in other words, $\lambda_i(u) = u_i$ for $i = 1, \ldots, n$.

The generalised recovery problem would be to find $\mathcal{S}_u \in \mathcal{F}$ such that

$$\lambda_i(\mathcal{S}_u) = u_i$$

for all $i = 1, \ldots, n$. In the same vein as the variational approach to interpolation that we saw in Section 2.3, we specify that $\mathcal{S}_u$ has minimal norm amongst all functions with this generalised interpolation property.

Example 2.8.1. *Let $\mathcal{A} = \mathcal{A}_\Omega \cup \mathcal{A}_{\partial\Omega} \subset \overline{\Omega} \subset \mathbb{R}^d$ be a set of interpolation points such that $\mathcal{A}_\Omega \in \Omega$ and $\mathcal{A}_{\partial\Omega} \in \partial\Omega$. Then we could have $\lambda_a = \delta_a \circ \Delta$ for $\mathcal{A}_\Omega$ and $\lambda_a = 0$ for $\mathcal{A}_{\delta\Omega}$, as in solving Laplace's equation with Dirichlet boundary conditions.*

Recall that the Riesz representation theorem implies that each functional λ_i has a unique representer $R_{\lambda_i} \in \mathcal{F}$ such that $\lambda_i = \langle \cdot, R_{\lambda_i} \rangle_{\mathcal{F}}$. The following is a generalisation of Theorem 2.3.3, the proof of which is along similar lines to the point-evaluation case, and shown explicitly by Wendland (2005, Theorem 16.5).

Theorem 2.8.2. *Let $\mathcal{F}$ be a Hilbert space, $\lambda_1, \ldots, \lambda_n$ be linearly independent functionals on $\mathcal{F}$ with corresponding representers $R_{\lambda_i} \in \mathcal{F}$, $1 \leq i \leq n$, and let $u_1, \ldots, u_n \in \mathbb{R}$ be given. Then the unique element $\mathcal{S}_u$ of minimal norm satisfying the interpolation conditions*

$$\lambda_i(\mathcal{S}_u) = u_i \qquad (1 \leq i \leq n),$$

is given by

$$\mathcal{S}_u = \sum_{i=1}^{n} \alpha_i R_{\lambda_i},$$

where the coefficients $\alpha_1, \ldots, \alpha_n$ are determined by the interpolation conditions.

Furthermore, $\mathcal{S}_u$ is the best approximation to u from $\operatorname{span}\{R_{\lambda_1}, \ldots, R_{\lambda_n}\}$ (Wendland 2005).

2.8.1 Non-symmetric and symmetric approaches

Although we have generalised the interpolation problem to encompass a wider class of data than merely function values, we would still like to take advantage of the strengths of radial basis function interpolation in solving the problem. Our first instinct would be to utilise collocation in the usual way, by constructing an approximation to a function u via

$$\mathcal{S}_u = \sum_{a \in \mathcal{A}} \alpha_a \Phi(\cdot, a),$$

for a chosen set of interpolation points $\mathcal{A} \subset \Omega$. This function lies in the span of the functions $\Phi(\cdot, a)$, $a \in \mathcal{A}$, where $\Phi = \phi \circ |\cdot|$ for one of our usual radial basis functions ϕ, and gives rise to a system of equations,

$$\lambda_b^1(\mathcal{S}_u) = \lambda_b(u), \tag{2.8.1}$$

$b \in \mathcal{A}$, to be solved in order to determine the coefficients α_a. (We are tacitly assuming that the elements of $\{\Phi(\cdot, a)\}_{a \in \mathcal{A}}$ are linearly independent.) This problem is referred to as *Hermite-Birkhoff interpolation*; for applications where the functionals are point evaluations of (combinations of) the derivatives of the function, the problem is referred to simply as *Hermite interpolation.*

Remark 2.8.3. *We have introduced a superscript notation onto the functionals λ_a in order to clarify as to which variable the operation is with respect. For example, suppose $\lambda_b = \delta_b$ (that is, evaluation at the point b). Then $\lambda_b^1(\Phi(\cdot, a)) = \Phi(b, a)$. Where there is no ambiguity the superscript is unnecessary and will not be displayed.*

This approach was first investigated by Kansa (1990*a*), with ϕ specifically being the multiquadric basis function (see Table 2.1 on page 23 for its definition), and hence is referred to as *Kansa's method*; it is equivalent to the radial basis function interpolation methods described earlier in this chapter if the functionals are all simple point evaluations. However,

in the general case the resulting interpolation matrix with entries

$$\lambda_b^1(\Phi(\cdot, a)),$$

$a, b \in \mathcal{A}$, may be non-symmetric, and for almost twenty years the non-singularity of this matrix was an open problem (Franke & Schaback 1998*a*, Beatson & Mouat 2002). However, recent work by Schaback (2005) appears to provide the conditions under which the matrix is invertible. However, even before this development, it was noted by Hon & Schaback (2001), for example, that the cases in which the possibility that the non-symmetric matrix may be singular are rare and the method performs very well (Kansa 1990*a*, Kansa 1990*b*, Wendland 2005). Even so, we may eliminate this concern entirely by instead constructing a symmetric matrix, the price of which is a second application of the functionals to the chosen basis function.

In other words, setting $\Phi(x, y) = \phi(|x - y|)$ for a conditionally strictly positive-definite radial basis function ϕ of order k, we define the approximation to the solution via

$$\mathcal{I}_{\mathcal{A}} u = \sum_{a \in \mathcal{A}} \alpha_a \lambda_a^1(\Phi(\cdot, \cdot)) + p \tag{2.8.2}$$

where $p \in \Pi_{k-1}^d$, and hence the interpolation conditions are

$$\lambda_b^2(\mathcal{I}_{\mathcal{A}} u) = \lambda_b(u),$$

for all $b \in \mathcal{A}$. Since ϕ is assumed to be only conditionally strictly positive-definite, we need to satisfy the additional conditions

$$\sum_{a \in \mathcal{A}} \alpha_a \lambda_a(p) = 0,$$

for all $p \in \Pi_{k-1}^d$, which again is equivalent to the usual natural conditions when the functionals are all straightforward point evaluations.

Finally, we must generalise the notion of our interpolation point-set $\mathcal{A}$ being unisolvent with respect to Π^d_{k-1}; to this end we insist that

$$\lambda_a(p) = 0 \text{ for all } a \in \mathcal{A} \text{ and } p \in \Pi^d_{k-1} \text{ implies that } p = 0. \tag{2.8.3}$$

The resulting interpolation matrix is thus of the form

$$\begin{pmatrix} A_{\phi,\Lambda} & \Lambda(P) \\ \Lambda(P)^T & 0 \end{pmatrix},$$

where $A_{\phi,\Lambda} = (\lambda^1_a(\lambda^2_b(\Phi(x,y)))_{a,b\in\mathcal{A}} \in \mathbb{R}^{n\times n}$, and $\Lambda(P) = (\lambda_a(p_s))_{a\in\mathcal{A}}$, with span $\{p_s\}_{s=1}^{\ell} = \Pi^d_{k-1}$. We now have an interpolation matrix which is symmetric — an important property in terms of considerably reducing the amount of computation required compared to the non-symmetric matrix above — and is provably non-singular (Wu 1992, Sun 1994, Iske 1995, Fasshauer 1997).

Of course, as noted earlier, the price we have paid is a second application of the functionals λ_a, $a \in \mathcal{A}$, and hence whatever basis function we choose, we must ensure it will permit both applications of the functionals without becoming discontinuous or vanishing. Thankfully, for the applications we have in mind, all of the radial basis functions we have met up to this point are suitable candidates for Hermite-Birkhoff interpolation.

2.8.2 Application of Hermite-Birkhoff interpolation to solving linear partial differential equation problems

Consider a partial differential equation problem of the form

$$\mathcal{L}u = f \quad \text{in } \Omega, \tag{2.8.4a}$$

$$\mathcal{B}u = g \quad \text{on } \partial\Omega, \tag{2.8.4b}$$

for a domain Ω in $\mathbb{R}^d$, where $\mathcal{L}$ and $\mathcal{B}$ are given differential operators, and f and g are also known.

Practical applications and investigations into the feasibility of collocation via radial basis functions for this type of problem are plentiful in the literature (see the work of Kansa (1990*a*), Dubal, Oliveira & Matzner (1992), Moridis & Kansa (1994), Dubal (1994) and Sharan, Kansa & Gupta (1997), for example), and one way of implementing such an approach is with Hermite-Birkhoff interpolation. Discretising Ω and $\partial\Omega$ with finite sets of points $\mathcal{A}_I$ and $\mathcal{A}_B$, respectively, we define the functionals

$$\lambda_b = \begin{cases} \delta_b \circ \mathcal{L}, & b \in \mathcal{A}_I, \\ \delta_b \circ \mathcal{B}, & b \in \mathcal{A}_B, \end{cases}$$

which are used in the generalised interpolant (2.8.2).

Convergence analysis was undertaken by Franke & Schaback (1998*a*) (see also (Franke & Schaback 1998*b*)) for the case where the differential operators are linear, in which they showed that, for all functionals $\lambda \in \mathcal{F}^*$ (the dual of the native space $\mathcal{F}$) the following error estimate holds:

$$|\lambda(u - \mathcal{I}_{\mathcal{A}} u)| \leq P(\lambda) \|u - \mathcal{I}_{\mathcal{A}} u\|_{\mathcal{F}}, \tag{2.8.5}$$

where P is the generalised power function given by

$$P(\lambda) = \inf \left\{ \|\lambda - \mu\|_{\mathcal{F}^*} \, : \, \mu \in \operatorname{span} \{\lambda_a\}_{a \in \mathcal{A}} \right\}.$$

In the special case of point evaluation functionals, P coincides with the power function that we met in Section 2.6.1, and thus Franke & Schaback's (1998*a*) work generalised the available results to a wider class of applications. However, it must be noted that, as the authors themselves acknowledge, a high degree of regularity of the solution is an essential assumption. For the general existence and uniqueness of the Hermite-Birkhoff interpolant

solution to the linear partial differential problem (2.8.4), we state the requirements as follows.

Definition 2.8.4 (Linear differential operator of order k). *An operator $\mathcal{T} : C^k(\Omega) \to C(\Omega)$ is called a* linear differential operator of order k *if it has the form*

$$\mathcal{T} = \sum_{|\alpha| \leq k} d_\alpha D^\alpha, \qquad \text{with } d_\alpha \in C(\Omega).$$

Example 2.8.5. *The well-known Laplacian operator Δ is a linear differential operator of order 2, since for example in two dimensions it may be written as*

$$\Delta \equiv D^{(2,0)} + D^{(0,2)}.$$

Theorem 2.8.6. *Let $\mathcal{L}, \mathcal{B}$ be linear differential operators of order k over $\Omega \subset \mathbb{R}^d$. Choose ϕ, a conditionally strictly positive-definite radial basis function of order k such that $\Phi \in C^{2k}(\Omega \times \Omega)$, where $\Phi(x, y) := \phi(|x - y|)$, and such that the native space is mapped via both $\mathcal{L}$ and $\mathcal{B}$ to functions that are at least continuous on the domain. Let $\mathcal{A} = \mathcal{A}_I \cup \mathcal{A}_B$ be a finite set of interpolation points satisfying the generalised unisolvency condition* (2.8.3). *Set*

$$\lambda_b = \begin{cases} \delta_b \circ \mathcal{L}, & b \in \mathcal{A}_I, \\ \delta_b \circ \mathcal{B}, & b \in \mathcal{A}_B, \end{cases}$$

and assume these functionals are linearly independent. Then the Hermite-Birkhoff interpolant

$$\mathcal{I}_\mathcal{A} u = \sum_{a \in \mathcal{A}} \alpha_a \lambda_a^1(\Phi(\cdot, \cdot)) + p,$$

uniquely satisfies the conditions

$$\begin{aligned} \lambda_b^2(\mathcal{I}_\mathcal{A} u) &= \lambda_b(u), && (b \in \mathcal{A}), \\ \sum_{a \in \mathcal{A}} \alpha_a \lambda_a(p) &= 0, && (p \in \Pi_{k-1}^d). \end{aligned}$$

This theorem is a special case of the general theory neatly described by Wendland (2005, Chapter 16). We will be applying Hermite-Birkhoff interpolation theory when we come to be investigating the numerical solution of advection-diffusion problems in Chapter 5. For the purely advective case, we are able to apply the classic theory of radial basis function interpolation, specifically that of local interpolation via polyharmonic splines which we saw in Section 2.7.4, and it is this class of problem we deal with first of all, in the following chapter.

Chapter 3

Linear Advection Problems

In this chapter we discretise and numerically solve the problem (1.1.1) with $\varepsilon = 0$, which is a purely advective system. The numerical method, which was introduced by Behrens & Iske (2002), is a combination of a semi-Lagrangian method and local radial basis function interpolation. The authors of the original paper incorporated an adaptive step into their algorithm, which automatically coarsens or refines the interpolation point set depending on pre-set criteria. Here, we do not consider adaptivity, but rather concentrate on the analysis for an arbitrary point set of fill distance h. Under certain assumptions on the smoothness of the solution u, as well as some other technical hypotheses, we are able to prove that the algorithm converges as $h \to 0$.

3.1 Model problem and discretisation

Given a final time $T > 0$, we consider the unsteady advection problem

$$\frac{\partial u}{\partial t} + a(x,t) \cdot \nabla u = f(x,t), \qquad (x,t) \in \mathbb{R}^d \times (0,T], \tag{3.1.1a}$$

$$u(x,0) = u_0(x), \qquad x \in \mathbb{R}^d, \tag{3.1.1b}$$

where $a : \mathbb{R}^d \times (0,T] \to \mathbb{R}^d$, $f : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ and $u_0 : \mathbb{R}^d \to \mathbb{R}$ are the *velocity field*, *forcing function* and *initial condition* respectively. For simplicity, we assume that u_0 has compact support in $\mathbb{R}^d$.

The semi-Lagrangian radial basis function method is constructed by first rewriting (3.1.1a) in terms of Lagrangian coordinates. To this end, we first introduce the following notation: let $X(x,s;t)$ be the position at time t of a point (or particle) located at position x at time $s \in [0,T]$ — thus $X(x,s;s) = x$. The particle trajectories, or *characteristics*, associated with problem (3.1.1a) are defined as the solution of the following system of ordinary differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}X(x,s;t) = a(X(x,s;t),t), \tag{3.1.2a}$$

$$X(x,s;s) = x. \tag{3.1.2b}$$

We note that for $a \in C^1(\mathbb{R}^d \times [0,T])^d$ there exists a unique solution to (3.1.2), see Lefschetz (1977), for example. Moreover, this assumption guarantees that the particle trajectories X are Lipschitz continuous in time, that is, there exists a positive constant L_X such that

$$|X(x,s;s) - X(x,s;t)| \leq L_X|t-s|, \tag{3.1.3}$$

for any $x \in \mathbb{R}^d$ and $s,t \in (0,T]$.

Additionally, we assume that (minimally) $f(\cdot,t) \in C^1(\mathbb{R}^d)$ for all $t \in (0,T]$ and $u_0 \in C(\mathbb{R}^d)$. We shall also need to assume that the solution $u(\cdot,t)$ is Lipschitz continuous in space, that is

$$|u(x,t) - u(y,t)| \leq L_u|x-y|, \tag{3.1.4}$$

for $x,y \in \mathbb{R}^d$, $t \in (0,T]$.

With this notation, we define the material derivative $D_t u$ as follows:

$$D_t u(x,t) := \frac{\mathrm{d}}{\mathrm{d}t}u(X(x,s;t),t)_{|_{t=s}} = u_t(x,t) + a(x,t)\cdot\nabla u(x,t) = f(x,t), \tag{3.1.5}$$

by application of the chain rule and (3.1.1a).

We let $0 = t^0 < t^1 < \ldots < t^M < t^{M+1} = T$ be a uniform subdivision of $[0,T]$, with time-step Δt; that is, $t^n = n\Delta t$, for $0 \leq n \leq M+1$. Integrating (3.1.5) with respect to t

between the time-steps t^n and t^{n+1}, we get the following exact representation formula for the analytical solution u:

$$u(X(x,t^{n+1};t^{n+1}),t^{n+1}) - u(X(x,t^{n+1};t^n),t^n) = \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\ \mathrm{d}t. \qquad (3.1.6)$$

If we now define y to be the "upstream" point at time t^n from which a point located at x at time t^{n+1} has originated, that is,

$$y = X(x,t^{n+1};t^n),$$

then (3.1.6) may be written in the following simplified form:

$$u(x,t^{n+1}) = u(y,t^n) + \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\ \mathrm{d}t \qquad (3.1.7)$$

for all $x \in \mathbf{R}^d$.

3.1.1 Temporal discretisation

We now write $u_{\Delta t}$ to denote the *semi-discrete temporal approximation* to u; this function is discrete in time, but continuous in space, with $u_{\Delta t}(\,\cdot\,,t^0) := u_0$. Using the rectangle rule and the fact that $t^{n+1} - t^n = \Delta t$, $0 \le n \le M$, we define $u_{\Delta t}$ as follows:

$$u_{\Delta t}(x,t^{n+1}) = u_{\Delta t}(y,t^n) + \Delta t f(x,t^{n+1}) \qquad (n = 0,\ldots,M).$$

In general, the precise location of y will not be typically known analytically, and thereby, it must be numerically approximated instead. For the purposes of this thesis, any appropriate numerical ordinary differential equation solver of order $\rho > 1$ may be employed; in other words, we assume that the scheme employed for the numerical approximation of (3.1.2) between two consecutive time levels satisfies

$$|y - \tilde{y}| \le L_\rho \Delta t^{\rho+1}, \qquad (3.1.8)$$

where $\tilde{y}$ denotes the approximation to the upstream point y and L_ρ is a non-negative constant which is independent of the time step Δt. As an example, an explicit Runge-Kutta method of order ρ may be employed for this task (see, for example, the book of Morton & Mayers (2005)). Using the same notation as before (for simplicity), we introduce the following alternative semi-discrete scheme:

$$u_{\Delta t}(x, t^{n+1}) = u_{\Delta t}(\tilde{y}, t^n) + \Delta t f(x, t^{n+1}), \qquad (n = 0, \ldots, M). \tag{3.1.9}$$

This describes the semi-discretisation of u with respect to time, so that $u_{\Delta t}(\,\cdot\,, t^n) \approx u(\,\cdot\,, t^n)$ for all time-levels $t^n = n\Delta t$, $n = 1, \ldots, M+1$. To complete the definition of the proposed discretisation method, we now proceed to discretise in space.

3.1.2 Spatial discretisation

We define a set of scattered, pairwise distinct interpolation points $\mathcal{A} = \{x_1, \ldots, x_{|\mathcal{A}|}\} \subset \mathbb{R}^d$ upon which we seek to approximate the value of $u_{\Delta t}(\,\cdot\,, t^n)$, $n = 0, \ldots, M+1$; here, we have used the notation $|\mathcal{A}|$ to denote the cardinality of the set $\mathcal{A}$. From equation (3.1.9) we see that in order to determine $u_{\Delta t}(x, t^{n+1})$, $x \in \mathcal{A}$, we need to exploit knowledge of the value of $u_{\Delta t}$ computed at the previous time-step t^n, at the upstream point $\tilde{y}$. In general, this point will not be contained in the point cloud $\mathcal{A}$, except in very special circumstances; thereby, we compute an approximation to $u_{\Delta t}(\tilde{y}, t^n)$ based on performing an interpolation over a subset of points in $\mathcal{A}$, naturally utilising the radial basis function theory we introduced in the previous chapter — specifically, local interpolation via polyharmonic splines. Recall that these have the form

$$\phi_{d,k,\mu}(r) = \begin{cases} n_{d,k,\mu} r^{2k+2\mu-d} \ln(r), & 2k + 2\mu - d \text{ even}, \\ n_{d,k,\mu} r^{2k+2\mu-d}, & \text{otherwise}, \end{cases}$$

where $k \geq 0$, $0 \leq \mu < 1$ and $n_{d,k,\mu}$ is a known constant. Their natural setting is the Beppo Levi space $BL^{k+\mu}(\mathbb{R}^d)$ which has the μ-dependent kernel $\mathcal{K}_{k,\mu}$ given by (2.4.2).

Firstly, for all $N \geq d+1$, assume $\{w_1, \ldots, w_N\}$ is a set of points unisolvent with respect to $\mathcal{K}_{k,\mu}$, and δ is chosen such that all sets $\{c_1, \ldots, c_N\} \in B(w_1, \delta) \times \cdots \times B(w_N, \delta)$ are unisolvent with respect to $\mathcal{K}_{k,\mu}$ (recall that such a δ exists by virtue of Lemma 2.6.4). Let the fill distance of a set $\mathcal{W} \in B(w_1, \delta) \times \cdots \times B(w_N, \delta)$ be $h_{\mathcal{W}}$, and let $\mathcal{A}$ be a point cloud in $\mathbb{R}^d$ of fill distance h. Now we assume that the following hypotheses hold:

(i) For each $x \in \mathcal{A}$, we assume that there exists a set of nearest neighbours $\mathcal{N}(x) \subset \mathcal{A}$ to the numerical approximation $\tilde{y}$ to the solution of the backwards characteristic equations (3.1.2). In particular, we assume that for each $x \in \mathcal{A}$ and $h > 0$ there exists an affine transformation $\sigma_x^h : \mathcal{N}(x) \to \mathcal{W}(x)$, where $\mathcal{W}(x) \in B(w_1, \delta) \times \cdots \times B(w_N, \delta)$.

(ii) Let $\gamma \geq 1$ be as defined in Remark 2.7.8, i.e. so that the stability constant of the interpolation operator $C_S = 1 + \mathcal{O}(N^\gamma)$. Then we assume that $N \sim \Delta t^{\frac{1}{\gamma}}$ —that is, there exists a constant K such that $C_S = 1 + K\Delta t$.

(iii) The polyharmonic spline variables k and μ are chosen such that $k + \mu > d/2$ and $k \geq m \geq 1$, where m is the degree of smoothness of the analytical solution u to (3.1.1).

Remark 3.1.1. *We note that hypothesis (i) is essential to guarantee that the resulting (local) interpolation problem (cf.* (2.7.1)–(2.7.2)*) is uniquely solvable. Hypothesis (ii) ensures that the semi-Lagrangian radial basis function scheme remains stable over time, by demanding that the size of the neighbour sets $\mathcal{N}(x)$, $x \in \mathcal{A}$, depends on the time-step Δt: for smaller time-steps we take smaller neighbour sets. We shall say more on this assumption following the proof of the scheme, in Remark 3.4.1.*

Hypothesis (iii) ensures that, in the first instance, the interpolant is at least continuous, and moreover, that a high enough degree of polynomial is reproduced in order to apply our intended Taylor series argument.

For further discussion on the selection of a nearest neighbour set, from a computational point of view, we refer the reader to Appendix A.

Given the above hypotheses, we write $u_{\Delta t,h}$ to denote the numerical approximation to u in both time and space. Following on from the temporal approximation (3.1.9), for the time-level t^{n+1} we define the numerical solution as follows:

$$u_{\Delta t,h}(x,t^{n+1}) := (\mathcal{I}_{\mathcal{N}(x)}u_{\Delta t,h}(\cdot\,,t^n))(\tilde{y}) + \Delta t f(x,t^{n+1}) \qquad (n=0,\ldots,M), \qquad (3.1.10)$$

for $x \in \mathcal{A}$, where $\tilde{y}$ is the numerical approximation to $y = X(x,t^{n+1};t^n)$. At time $t^0 = 0$, we set $u_{\Delta t,h}(x,t^0) = u_0(x)$ for all $x \in \mathcal{A}$.

3.2 Main results

In this section we state *a priori* error bounds for both the (continuous in space) semi-discrete scheme (3.1.9), as well as for the radial basis function semi-Lagrangian numerical method (3.1.10).

Lemma 3.2.1. *Let $u : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ be the analytical solution of the advection problem* (3.1.1) *satisfying $u \in C(\mathbb{R}^d \times (0,T])$. Let $u_{\Delta t}$ be the semi-discrete temporal approximation obtained by* (3.1.9). *Then, the following* a priori *error bound holds:*

$$\|u(\cdot\,,t^{n+1}) - u_{\Delta t}(\cdot\,,t^{n+1})\|_\infty \leq t^{n+1}(L_u L_\rho \Delta t^\rho + C_{f,X}\Delta t),$$

where L_u is defined in (3.1.4), *L_ρ and ρ are defined in* (3.1.8) *and $C_{f,X} = L_X \|f\|_{C((0,T];C^1(\mathbb{R}^d))}$ (cf.* (1.3.4)*), where L_X is as in* (3.1.3).

Theorem 3.2.2. *Let $u : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ be the analytical solution of the advection problem* (3.1.1) *satisfying $u \in C((0,T];C^m(\mathbb{R}^d))$. Let $u_{\Delta t,h}$ be the semi-Lagrangian numerical approximation obtained by* (3.1.10) *by employing polyharmonic splines, assuming that hypotheses (i), (ii) and (iii) are satisfied. Then, the following* a priori *error bound holds:*

$$\begin{aligned}\|u(\cdot\,,t^{n+1}) - u_{\Delta t,h}(\cdot\,,t^{n+1})\|_{\infty,\mathcal{A}} \\ \leq e^{(K+1)t^{n+1}}\Big(L_u L_\rho \Delta t^\rho + C_{f,X}\Delta t + C_T h^m \Delta t^{-1}\|u\|_{\ell^\infty((t^0,t^n];W^m_\infty(\mathbb{R}^d))}\Big),\end{aligned}$$

where

$$\|u\|_{\ell^\infty((t^0,t^n];W^m_\infty(\mathbb{R}^d))} := \max_{0\le i\le n}\|u(\cdot,t^i)\|_{W^m_\infty(\mathbb{R}^d)} := \max_{0\le i\le n}\sum_{|\alpha|=m}\|D^\alpha u(\cdot,t^i)\|_\infty,$$

and the constants K, L_u, L_ρ, C_T and $C_{f,X}$ are defined in hypothesis (ii), (3.1.4), (3.1.8), *Theorem 2.7.9, and Lemma 3.2.1, respectively.*

The proofs of Lemma 3.2.1 and Theorem 3.2.2 will now be presented in the following sections, respectively.

3.3 Analysis of the semi-discrete scheme

In this section we estimate the error between the analytical and numerical solutions with respect to time only, so we consider the difference between the two solutions u and $u_{\Delta t}$ defined by (3.1.7) and (3.1.9), respectively. To this end, subtracting (3.1.7) and (3.1.9) gives

$$\begin{aligned} u(x,t^{n+1}) - u_{\Delta t}(x,t^{n+1}) &= u(y,t^n) - u_{\Delta t}(\tilde{y},t^n) \\ &\quad + \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\,\mathrm{d}t - \Delta t f(x,t^{n+1}) \\ &= u(y,t^n) - u(\tilde{y},t^n) + u(\tilde{y},t^n) - u_{\Delta t}(\tilde{y},t^n) \\ &\quad - \Delta t f(x,t^{n+1}) + \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\,\mathrm{d}t. \end{aligned} \tag{3.3.1}$$

We note from Taylor's Theorem that

$$\begin{aligned} f(X(x,t^{n+1};t),t) =& f(X(x,t^{n+1};t^{n+1}),t^{n+1}) + \\ & (X(x,t^{n+1};t) - X(x,t^{n+1};t^{n+1}))\cdot\nabla_x f(\xi(t),t), \end{aligned}$$

where $\xi(t)$ is on the line segment $(X(x,t^{n+1};t),X(x,t^{n+1};t^{n+1}))$. Thereby, writing

$$E_n = \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\ \mathrm{d}t - \Delta t f(x,t^{n+1}). \tag{3.3.2}$$

we deduce that

$$E_n = \int_{t^n}^{t^{n+1}} (X(x,t^{n+1};t) - X(x,t^{n+1};t^{n+1}))\cdot \nabla_x f(\xi(t),t)\ \mathrm{d}t.$$

Exploiting the Lipschitz continuity of X (3.1.3), gives

$$|E_n| \leq L_X \Delta t \int_{t^n}^{t^{n+1}} |\nabla_x f(\xi(t),t)|\ \mathrm{d}t \leq C_{f,X,n+1}\Delta t^2, \tag{3.3.3}$$

where $C_{f,X,n+1} = L_X \|f\|_{C(t^n,t^{n+1};C^1(\mathbb{R}^d))} := L_X \sup_{t\in[t^n,t^{n+1}]}|\nabla_x f(\xi(t),t^{n+1})|$. Additionally, using the Lipschitz continuity of u, cf. (3.1.4), together with (3.1.8), we get

$$|u(y,t^n) - u(\tilde{y},t^n)| \leq L_u L_\rho \Delta t^{\rho+1}. \tag{3.3.4}$$

Employing (3.3.3) and (3.3.4) we may bound (3.3.1) as follows

$$|u(x,t^{n+1}) - u_{\Delta t}(x,t^{n+1})| \leq |u(\tilde{y},t^n) - u_{\Delta t}(\tilde{y},t^n)| + L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2.$$

Taking the maximum over $x \in \mathbb{R}^d$, we deduce that

$$\|u(\cdot,t^{n+1}) - u_{\Delta t}(\cdot,t^{n+1})\|_\infty \leq \|u(\cdot,t^n) - u_{\Delta t}(\cdot,t^n)\|_\infty + L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2. \tag{3.3.5}$$

At time $t = 0$ (that is, when $t = t^0$), $\|u(\cdot, t^0) - u_{\Delta t}(\cdot, t^0)\|_\infty = 0$; thereby, recursively applying (3.3.5) gives

$$\begin{aligned}
\|u(\cdot, t^{n+1}) - u_{\Delta t}(\cdot, t^{n+1})\|_\infty &\leq \|u(\cdot, t^n) - u_{\Delta t}(\cdot, t^n)\|_\infty + L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 \\
&\leq \|u(\cdot, t^{n-1}) - u_{\Delta t}(\cdot, t^{n-1})\|_\infty \\
&\quad + 2L_u L_\rho \Delta t^{\rho+1} + (C_{f,X,n+1} + C_{f,X,n})\Delta t^2 \\
&\vdots \\
&\leq (n+1)L_u L_\rho \Delta t^{\rho+1} + \sum_{i=1}^{n+1} C_{f,X,i}\Delta t^2.
\end{aligned}$$

Noting that $\sum_{i=1}^{n+1} C_{f,X,i} \leq (n+1)C_{f,X}$, and using the fact that $t^{n+1} = (n+1)\Delta t$, we deduce the statement of Lemma 3.2.1.

Remark 3.3.1. *In particular, from this result, we have*

$$|u(x, t^n) - u_{\Delta t}(x, t^n)| \leq \|u(\cdot, t^n) - u_{\Delta t}(\cdot, t^n)\|_\infty \leq t^{n+1}(L_u L_\rho \Delta t^\rho + C_{f,X}\Delta t)$$

for each $x \in \mathcal{A}$.

Remark 3.3.2. *Note that when* $f \equiv 0$, $C_{f,X} = 0$; *moreover,* $L_\rho = 0$ *when the ODE method used to solve the backwards characteristics system* (3.1.2) *is exact. In this case the right-hand side of the error bound stated in Lemma 3.2.1 is identically zero, which reflects the fact that there is no error in the semi-discrete scheme in this case.*

3.4 Analysis of the radial basis function semi-Lagrangian numerical method

Finally, in this section we present the proof of Theorem 3.2.2; that is, we analyse the difference between the analytical solution and its numerical approximation in space and

time (3.1.10) at an arbitrary point x in our discrete set $\mathcal{A}$ at time-level t^{n+1}:

$$\begin{aligned}
|u(x,t^{n+1}) - u_{\Delta t,h}(x,t^{n+1})| &\leq |u(y,t^n) - (\mathcal{I}_{\mathcal{N}(x)}u_{\Delta t,h}(\,\cdot\,,t^n))(\tilde{y})| \\
&\quad + \left| \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\ \mathrm{d}t - \Delta t f(x,t^{n+1}) \right| \\
&\leq |u(y,t^n) - u(\tilde{y},t^n)| + |u(\tilde{y},t^n) - (\mathcal{I}_{\mathcal{N}(x)}u(\,\cdot\,,t^n))(\tilde{y})| \\
&\quad + |(\mathcal{I}_{\mathcal{N}(x)}u(\,\cdot\,,t^n))(\tilde{y}) - (\mathcal{I}_{\mathcal{N}(x)}u_{\Delta t,h}(\,\cdot\,,t^n))(\tilde{y})| \\
&\quad + \left| \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\ \mathrm{d}t - \Delta t f(x,t^{n+1}) \right| \\
&\leq L_u L_\rho \Delta t^{\rho+1} + C_T h^m |u(\,\cdot\,,t^n)|_{W^m_\infty(\mathbb{R}^d)} \\
&\quad + C_S \|u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n)\|_{\infty,\mathcal{A}} + C_{f,X,n+1}\Delta t^2, \quad (3.4.1)
\end{aligned}$$

by application of the bound (3.3.3), Theorems 2.7.9 and 2.7.7, and the bound (3.3.4).

Recall from hypothesis (ii) that $C_S = 1 + K\Delta t$. Thus,

$$\begin{aligned}
|u(x,t^{n+1}) - u_{\Delta t,h}(x,t^{n+1})| &\leq L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 + C_T h^m |u(\,\cdot\,,t^n)|_{W^m_\infty(\mathbb{R}^d)} \\
&\quad + (1 + K\Delta t)\|u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n)\|_{\infty,\mathcal{A}}.
\end{aligned}$$

Taking the maximum over all $x \in \mathcal{A}$ provides us with the error estimate in terms of the previous time-level:

$$\begin{aligned}
\|u(\,\cdot\,,t^{n+1}) - u_{\Delta t,h}(\,\cdot\,,t^{n+1})\|_{\infty,\mathcal{A}} &\leq L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 + C_T h^m |u(\,\cdot\,,t^n)|_{W^m_\infty(\mathbb{R}^d)} \\
&\quad + (1 + K\Delta t)\|u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n)\|_{\infty,\mathcal{A}}.
\end{aligned}$$

Telescoping this bound through all time-levels, noting that $\|u(\,\cdot\,,t^0) - u_{\Delta t,h}(\,\cdot\,,t^0)\|_{\infty,\mathcal{A}} = 0$,

$$\begin{aligned}
&\|u(\,\cdot\,,t^{n+1}) - u_{\Delta t,h}(\,\cdot\,,t^{n+1})\|_{\infty,\mathcal{A}} \\
&\quad \leq L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 + C_T h^m \|u(\,\cdot\,,t^n)\|_{W^m_\infty(\mathbb{R}^d)} \\
&\qquad + (1 + K\Delta t)\|u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n)\|_{\infty,\mathcal{A}}
\end{aligned}$$

$$
\begin{aligned}
&\leq L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 + C_T h^m \|u(\cdot,t^n)\|_{W_\infty^m(\mathbb{R}^d)} \\
&\quad + (1+K\Delta t)\Big(L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n}\Delta t^2 + C_T h^m \|u(\cdot,t^{n-1})\|_{W_\infty^m(\mathbb{R}^d)} \\
&\quad\quad + (1+K\Delta t)\|u(\cdot,t^{n-1}) - u_{\Delta t,h}(\cdot,t^{n-1})\|_{\infty,\mathcal{A}}\Big) \\
&= L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n+1}\Delta t^2 + C_T h^m \|u(\cdot,t^n)\|_{W_\infty^m(\mathbb{R}^d)} \\
&\quad + (1+K\Delta t)\Big(L_u L_\rho \Delta t^{\rho+1} + C_{f,X,n}\Delta t^2 + C_T h^m \|u(\cdot,t^{n-1})\|_{W_\infty^m(\mathbb{R}^d)}\Big) \\
&\quad\quad + (1+K\Delta t)^2\|u(\cdot,t^{n-1}) - u_{\Delta t,h}(\cdot,t^{n-1})\|_{\infty,\mathcal{A}} \\
&\;\;\vdots \\
&\leq \sum_{i=0}^{n}(1+K\Delta t)^{n-i}\Big(L_u L_\rho \Delta t^{\rho+1} + C_{f,X,i+1}\Delta t^2 + C_T h^m \|u(\cdot,t^i)\|_{W_\infty^m(\mathbb{R}^d)}\Big) \\
&\leq (n+1)(1+K\Delta t)^{n+1}\Big(L_u L_\rho \Delta t^{\rho+1} + C_{f,X}\Delta t^2 + C_T h^m \|u\|_{\ell^\infty(t^0,t^n;W_\infty^m(\mathbb{R}^d))}\Big) \\
&= t^{n+1}(1+K\Delta t)^{n+1}\Big(L_u L_\rho \Delta t^{\rho} + C_{f,X}\Delta t + C_T h^m \Delta t^{-1}\|u\|_{\ell^\infty(t^0,t^n;W_\infty^m(\mathbb{R}^d))}\Big),
\end{aligned}
$$

since $(n+1)\Delta t = t^{n+1}$. The facts that

$$
(1+K\Delta t)^{n+1} \leq (e^{K\Delta t})^{n+1} = e^{(n+1)K\Delta t} = e^{Kt^{n+1}}
$$

and

$$
t^{n+1} = e^{\ln(t^{n+1})} \leq e^{t^{n+1}}
$$

complete the result.

Remark 3.4.1. *Recall hypothesis (ii), which assumed that the size N of a neighbour set $\mathcal{N}$ is related to the time-step Δt. In the two dimensional thin-plate spline case we saw that we required $N^3 = K\Delta t$ for some constant K (cf. (2.7.15)). From Theorem 3.2.2 we can see that for $\rho \geq 1$ the first two terms of the error estimate are balanced by $\Delta t^{-1}h^m$, so this latter quantity must be uniformly bounded above to ensure convergence of the scheme. Thus, in essence, N should be of order $h^{m/6}$ as the mesh is uniformly refined, indicating that for rougher functions we should take N to be larger than for smoother functions.*

Remark 3.4.2. *In the special case of thin-plate spline interpolation in two dimensions, when $N = 3$ the interpolation operator reduces to that of linear polynomial interpolation, in which case it can be shown via an alternative method of proof that the scheme converges precisely as predicted by Theorem 3.2.2 but with N independent of Δt.*

Chapter 4

Linear Advection Numerical Experiments

In this chapter we present the results of some two-dimensional numerical experiments which illustrate the performance of the method and confirm our analysis. In Section 4.1 we consider the oft-used rotating cone problem. The exact solutions for both the ordinary differential equation problem of finding the upstream points (cf. (3.1.2) on page 76) and the partial differential equation problem itself are known, hence we can investigate the accuracy of the scheme, and as the analytical solution possesses the required level of regularity, we may apply the theory of the previous chapter.

In Section 4.2 we consider the rotating cylinder problem, which is similar to the rotating cone except that in this case the solution is not continuous everywhere at any time t and hence our analysis cannot be applied; however, the numerical experiments show that it still performs very well for this problem.

Finally, in Section 4.3 we present numerical results for a Gaussian cone in a vortex, for which an analytical solution is not available, to further demonstrate the utility of the scheme.

4.1 Experiment 1: Rotating cone problem

In this first example, we let $f = 0$, $a(x, y, t) = -2\pi(2y - 1, 1 - 2x)^T$ and

$$u_0(x, y) = \begin{cases} \cos^2(2\pi r), & r \leq \frac{1}{4}, \\ 0, & \text{otherwise}, \end{cases}$$

where $r^2 = (2x - \frac{1}{2})^2 + (2y - 1)^2$.

The analytical solution u is simply the initial cone u_0 rotating about the point $(\frac{1}{2}, \frac{1}{2})$ with period $\frac{1}{2}$ (in other words, $u(x, y, \frac{j}{2}) = u_0(x, y)$ for $j = 0, 1, 2, \ldots$), and hence we can restrict our domain of computation to $[0, 1]^2$. This problem is a commonly-used test case for linear advection schemes, also seen in the work of Süli & Ware (1991) and Priestley (1994), for example.

We remark that the solution lies in the fractional-order Sobolev space $W_2^{5/2-\epsilon}((0, 1)^2)$, $\epsilon > 0$; this space comprises the same equivalence classes of functions as $BL^{5/2-\epsilon}((0, 1)^2)$, which is contained in $C^1([0, 1]^2)$ by Lemma 2.4.7. Thereby, this model problem satisfies the hypotheses of our analysis — recall from Theorem 3.2.2 (page 80) that we require the smoothness of the solution u to be $m \geq 1$. Thus, we proceed to demonstrate the practical performance of the numerical scheme (3.1.10) on a sequence of increasingly finer point clouds $\mathcal{A}$ of regular gridpoints covering $[0, 1]^2$. To this end, in Tables 4.1 and 4.2 we present numerical experiments, based on employing the thin-plate spline radial basis function $\phi(r) = r^2 \ln r$, that is, $k = 2$ and $\mu = 0$, for two different time-steps, $\Delta t = 0.01$ and $\Delta t = 0.001$, respectively. In both cases, we show the maximum error computed over all of the nodes in the point cloud $\mathcal{A}$ at the final time $T = 0.5$ and their respective orders of convergence, using an increasing number of neighbours for the computation of the approximation of the solution at the upstream point, as well as using a global interpolant, where computationally feasible.

Since here $m = 1$, Theorem 3.2.2 would predict an order of convergence of $\mathcal{O}(h)$ as h tends to zero. However, from Tables 4.1 and 4.2, we clearly observe that $\mathcal{O}(h^2)$ convergence

	N		
	10	20	40
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order
225	5.228e−1 / −	4.534e−1 / −	4.725e−1 / −
900	1.708e−1 / 1.61	9.660e−2 / 2.23	1.255e−1 / 1.91
3600	4.621e−2 / 1.89	3.867e−2 / 1.32	1.547e−2 / 3.02
8100	2.629e−2 / 1.39	2.070e−2 / 1.54	6.369e−3 / 2.19
14400	1.632e−2 / 1.66	1.209e−2 / 1.87	3.394e−3 / 2.19
57600	4.333e−3 / 1.91	3.179e−3 / 1.93	7.943e−4 / 2.10

	N		
	80	160	$\lvert\mathcal{A}\rvert$
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order
225	4.752e−1 / −	4.752e−1 / −	4.752e−1 / −
900	1.269e−1 / 1.90	1.276e−1 / 1.90	1.276e−1 / 1.90
3600	1.260e−2 / 3.33	1.271e−2 / 3.33	1.273e−2 / 3.33
8100	4.847e−3 / 2.36	4.770e−3 / 2.42	4.772e−3 / 2.42
14400	2.553e−3 / 2.23	2.472e−3 / 2.29	− / −
57600	5.855e−4 / 2.12	5.578e−4 / 2.15	− / −

Table 4.1: $L_{\infty,\mathcal{A}}$ errors at $T = 0.5$ for the rotating cone problem with $\Delta t = 0.01$ and $k = 2$.

	N			
	20	40	80	160
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order	Error / Order
225	4.247e−1 / −	4.889e−1 / −	4.924e−1 / −	4.922e−1 / 0.00
900	3.355e−1 / 0.34	2.511e−1 / 0.96	2.408e−1 / 1.03	2.429e−1 / 1.02
3600	3.696e−1 / −0.14	1.027e−1 / 1.29	4.229e−2 / 2.51	4.267e−2 / 2.51
14400	1.475e−1 / 1.33	3.886e−2 / 1.40	1.035e−2 / 2.03	9.666e−3 / 2.14
57600	4.004e−2 / 1.88	7.397e−3 / 2.39	2.506e−3 / 2.05	1.995e−3 / 2.28

Table 4.2: $L_{\infty,\mathcal{A}}$ errors at $T = 0.5$ for the rotating cone problem with $\Delta t = 0.001$ and $k = 2$.

is attained in practice as the number of points in $\mathcal{A}$ is uniformly increased. This is further observed in Figure 4.1, where we plot the results presented in Tables 4.1 and 4.2. Secondly, we note that on coarse meshes the error is relatively insensitive to an increase in the number of points employed in the nearest neighbour set $\mathcal{N}$ for the local interpolation algorithm. However, on finer meshes, an increase in the size of $\mathcal{N}$ does indeed lead to a significant decrease in the error. It is particularly noteworthy that, whilst for the analysis we required that N be explicitly dependent on the mesh-spacing, we see here that the method converges as h tends to zero for fixed N. In practice, there is a balance between the computational effort involved in solving these larger local problems on coarser meshes, as opposed to using smaller neighbour sets $\mathcal{N}$ on finer meshes. The latter approach is particularly attractive from a parallel perspective, as these local interpolation problems are entirely independent of each other. Finally, from Tables 4.1 and 4.2, we notice that on finer meshes, a decrease in the time step actually leads to an increase in the error. This is due to the fact that since $f = 0$ and that the upstream points are evaluated exactly, in this problem there is no error due to Δt—in other words, the first two terms on the right-hand side of the error bound stated in Theorem 3.2.2 are zero—thereby, a smaller value for Δt must be compensated for by a finer node set $\mathcal{A}$.

In Tables 4.3 and 4.4 we present results based on employing smoother polyharmonic spline interpolation functions, with $k = 3$ and $k = 4$, respectively, with $\Delta t = 0.01$; cf. Figure 4.2, also. Here, we observe that the order of convergence remains unchanged, as we would expect, and moreover that the actual magnitude of the errors are very comparable for each value of k. Finally, in Figures 4.3 and 4.4 we show the computed numerical solution at the final time $T = 0.5$, with $\Delta t = 0.01$ and $k = 2$, for different choices of $\mathcal{A}$ and $\mathcal{N}$. Here, we note that although some small oscillations are observable at the base of the cone, together with some slight loss in the height of the profile, the general quality of the computed solution is extremely good.

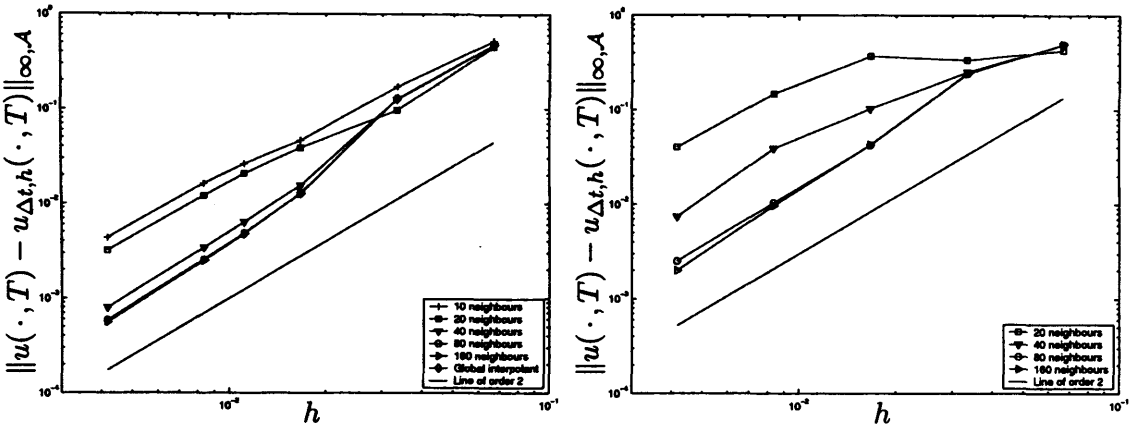


Figure 4.1: Convergence plots for the rotating cone problem at $T = 0.5$, with $k = 2$: (a) $\Delta t = 0.01$; (b) $\Delta t = 0.001$.

	N			
	10	20	40	80
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order	Error / Order
225	2.403e−1 / −	1.878e−1 / −	2.239e−1 / −	2.228e−1 / −
900	6.531e−2 / 1.88	3.364e−2 / 2.48	2.380e−2 / 3.23	1.780e−2 / 3.65
3600	1.633e−2 / 2.00	6.231e−3 / 2.43	3.582e−3 / 2.73	3.436e−3 / 2.37
14400	2.908e−3 / 2.49	1.090e−3 / 2.52	8.209e−4 / 2.13	7.855e−4 / 2.13
57600	5.822e−4 / 2.32	2.352e−4 / 2.21	1.924e−4 / 2.09	1.930e−4 / 2.02

Table 4.3: $L_{\infty,\mathcal{A}}$ errors at $T = 0.5$ for the rotating cone problem with $\Delta t = 0.01$ and $k = 3$.

	N			
	10	20	40	80
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order	Error / Order
225	3.726e−1 / −	1.235e−1 / −	9.810e−2 / −	1.272e−1 / −
900	7.766e−2 / 2.26	3.677e−2 / 1.75	1.210e−2 / 3.02	9.066e−3 / 3.81
3600	8.124e−3 / 3.26	4.086e−3 / 3.17	2.156e−3 / 2.49	2.333e−3 / 1.96
14400	1.808e−3 / 2.17	7.875e−4 / 2.38	4.453e−4 / 2.28	4.968e−4 / 2.23
57600	4.337e−4 / 2.06	1.932e−4 / 2.03	1.143e−4 / 1.96	1.315e−4 / 1.92

Table 4.4: $L_{\infty,\mathcal{A}}$ errors at $T = 0.5$ for the rotating cone problem with $\Delta t = 0.01$ and $k = 4$.

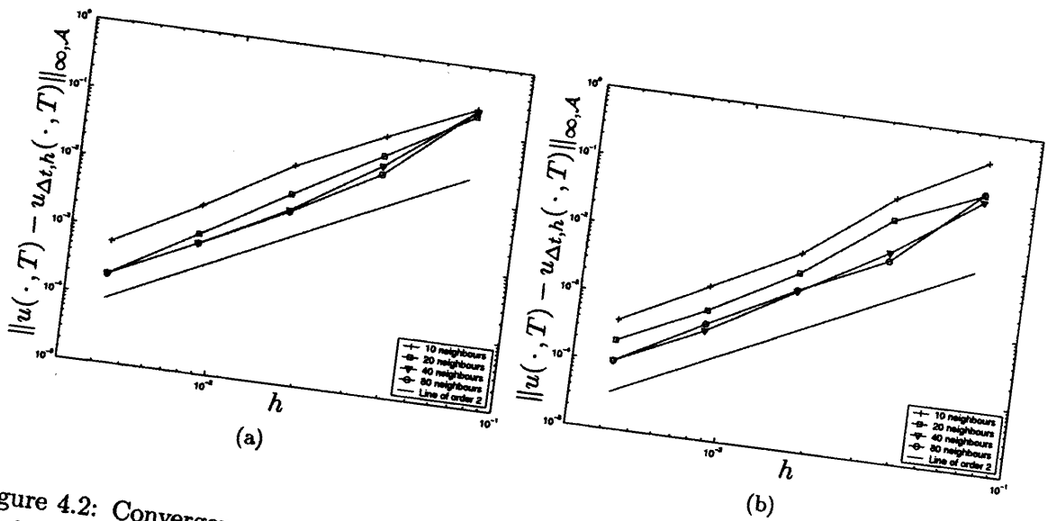


Figure 4.2: Convergence plots for the rotating cone problem at $T = 0.5$, with $\Delta t = 0.01$: (a) $k = 3$; (b) $k = 4$.

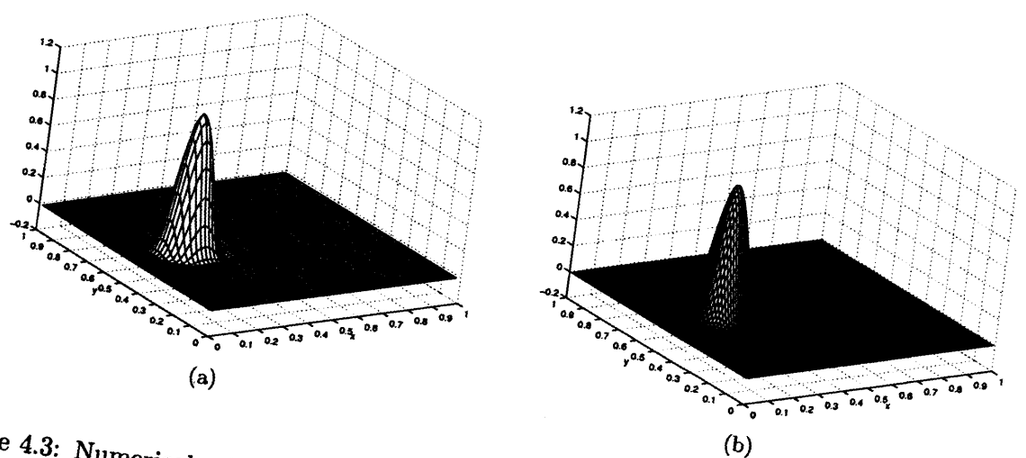


Figure 4.3: Numerical solutions of rotating cone problem at $T = 0.5$ with $\Delta t = 0.01$ and $k = 2$: (a) $|\mathcal{A}| = 3600$, $N = 20$; (b) $|\mathcal{A}| = 14400$, $N = 10$.

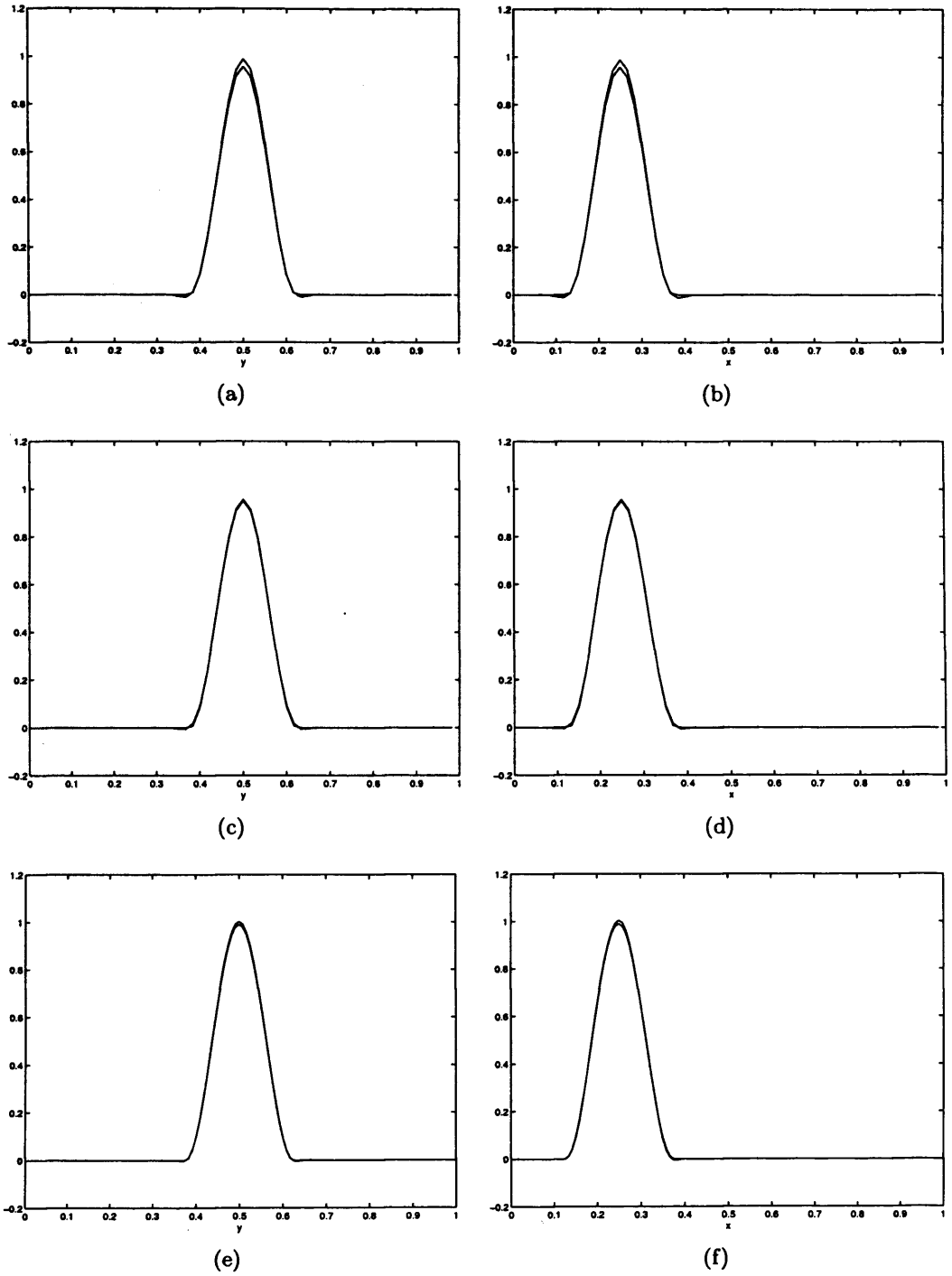


Figure 4.4: Cut through solutions for rotating cone problem at $T = 0.5$ with $\Delta t = 0.01$ and $k = 2$: (a) $|\mathcal{A}| = 3600$, $N = 20$, where $x = 0.25$ and $0 \leq y \leq 1$; (b) $|\mathcal{A}| = 3600$, $N = 20$, where $y = 0.5$ and $0 \leq x \leq 1$; (c) $|\mathcal{A}| = 3600$, $N = 80$, where $x = 0.25$ and $0 \leq y \leq 1$; (d) $|\mathcal{A}| = 3600$, $N = 80$, where $y = 0.5$ and $0 \leq x \leq 1$; (e) $|\mathcal{A}| = 14400$, $N = 10$, where $x = 0.25$ and $0 \leq y \leq 1$; (f) $|\mathcal{A}| = 14400$, $N = 10$, where $y = 0.5$ and $0 \leq x \leq 1$.

4.2 Experiment 2: Rotating cylinder problem

In this second example, we consider the rotating cylinder problem; this another popular test case which demonstrates the effectiveness of the scheme even when the solution is discontinuous. Here, the set-up is almost identical to that of the rotating cone problem presented in Section 4.1; the only difference being the initial condition is now the cylinder defined by

$$u_0(x,y) = \begin{cases} 1, & r \leq \frac{1}{4}, \\ 0, & \text{otherwise}, \end{cases}$$

where $r^2 = (2x - \frac{1}{2})^2 + (2y - 1)^2$.

This problem can also be found in the results of Bermejo (1990), for example. In Figure 4.5 we show slices of the analytical and numerical solutions at $T = 0.5$, with $\Delta t = 0.01$ and $k = 2$, for varying sizes of point- and neighbour-sets. Here, we again observe some oscillatory behaviour of the computed numerical solution in the vicinity of the discontinuities in the analytical solution. However, these oscillations are extremely localised, and do not lead to a general deterioration of the accuracy of the underlying approximation.

4.3 Experiment 3: Gaussian cone in a vortex

In this section we present numerical results which serve to further validate the scheme, with a more complex velocity field than the previous two examples. The domain of computation is again $[0,1]^2$. The initial condition is given by

$$u_0(x,y) = 10\exp\left(-128\left(\left(x - \frac{3}{4}\right)^2 + \left(y - \frac{3}{4}\right)^2\right)\right),$$

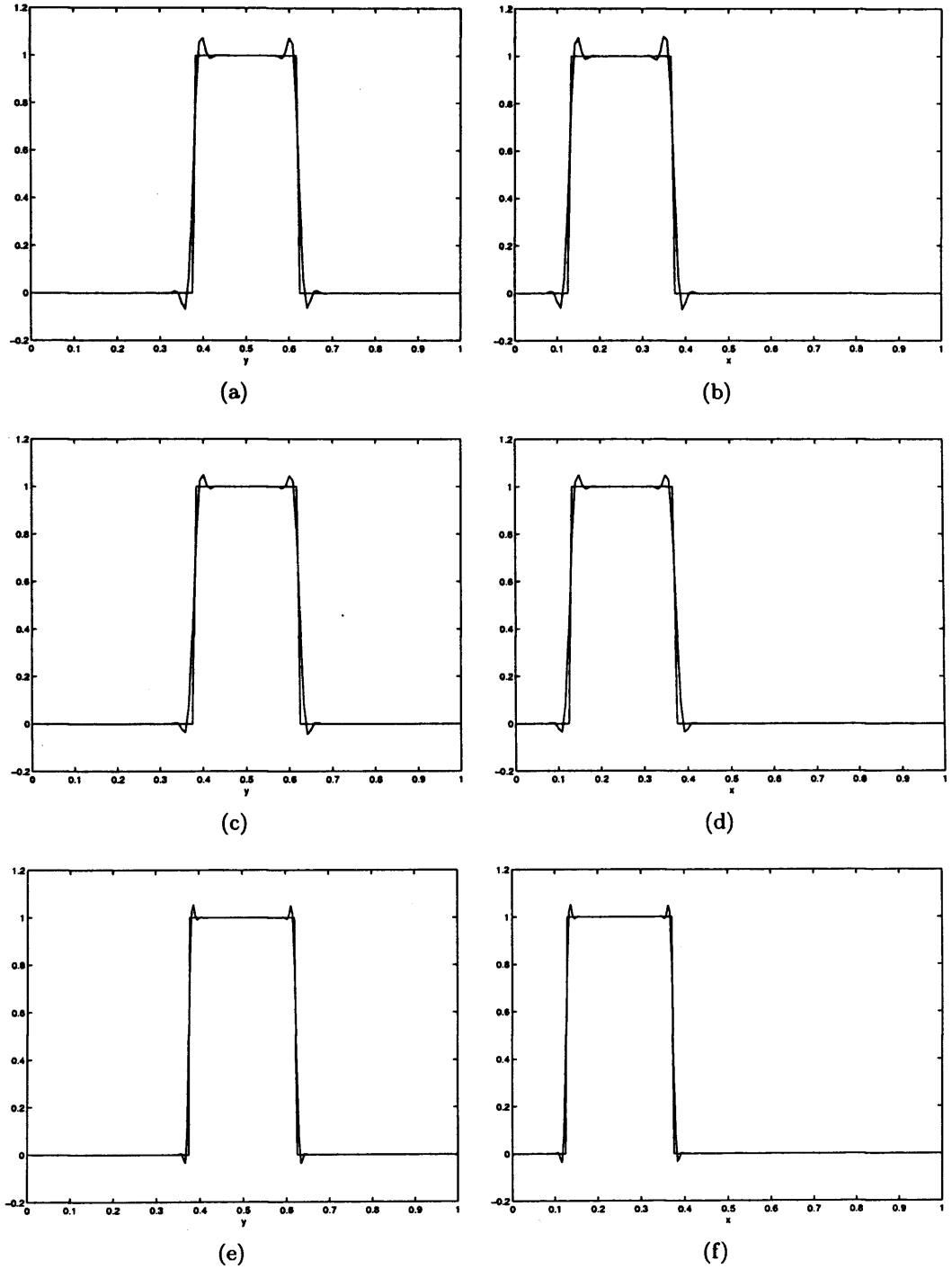


Figure 4.5: Cut through solutions for rotating cylinder problem at $T = 0.5$ with $\Delta t = 0.01$: (a) $|\mathcal{A}| = 14400$, $N = 20$ where $x = 0.25$ and $0 \leq y \leq 1$; (b) $|\mathcal{A}| = 14400$, $N = 20$ where $y = 0.5$ and $0 \leq x \leq 1$; (c) $|\mathcal{A}| = 14400$, $N = 80$ where $x = 0.25$ and $0 \leq y \leq 1$; (d) $|\mathcal{A}| = 14400$, $N = 80$ where $y = 0.5$ and $0 \leq x \leq 1$; (e) $|\mathcal{A}| = 57600$, $N = 10$ where $x = 0.25$ and $0 \leq y \leq 1$; (f) $|\mathcal{A}| = 57600$, $N = 80$ where $y = 0.5$ and $0 \leq x \leq 1$.

in other words, a Gaussian cone of height 10 centred at $\left(\frac{3}{4}, \frac{3}{4}\right)$, and the velocity field is

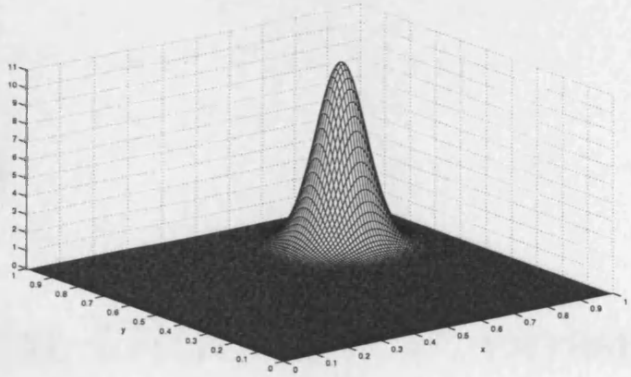
$$a(x,y,t) = \frac{7r(x,y)\sinh(r(x,y))}{160\cosh^3(r(x,y))}\left(\frac{1}{\frac{1}{2}-y}, -\frac{1}{\frac{1}{2}-x}\right)^T,$$

where

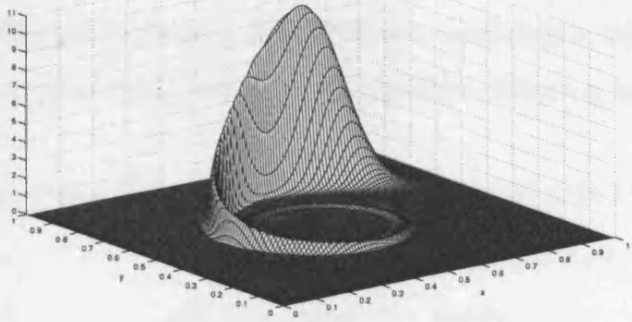
$$r(x,y) := 8\sqrt{\left(\frac{1}{2}-x\right)^2 + \left(\frac{1}{2}-y\right)^2},$$

which essentially "sucks" the cone into the centre of the domain in a spiralling, whirlpool-like motion. Here, the forcing function is $f = 0$.

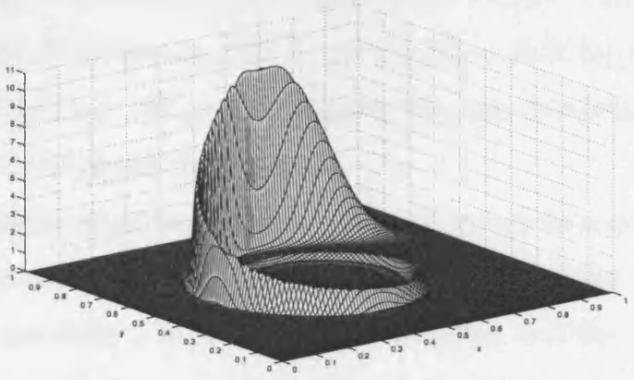
Figure 4.6 shows the results for $|\mathcal{A}| = 14400$, $N = 20$, $\Delta t = 0.01$ at times $t = 0$, $t = 2.5$ and $t = 5$. In this example we do not have an exact solution for the backwards characteristic problem (3.1.2) for computing the upstream points y, so instead a Runge-Kutta method of order 4 (sometimes known as RK4) was deployed to determine the approximate point $\tilde{y}$ in each instance. This problem appears to have received little or no attention in the current literature, so drawing conclusions as to the performance of our particular scheme is difficult. However, we can see from Figure 4.6 that although we can observe some slight degradation of the numerical solution along the spine (due to the resolution of the point cloud $\mathcal{A}$), the scheme appears to perform very well for this problem, displaying the behaviour we would expect of the given velocity field on the initial data.



(a) $t = 0$



(b) $t = 2.5$



(c) $t = 5$

Figure 4.6: Numerical solution for Experiment 3: Gaussian cone in a vortex.

Chapter 5

Advection-Diffusion Problems

We now increase the complexity of the partial differential equation problem that we have been numerically solving by reintroducing the diffusion coefficient ε, and thus we present a new scheme for solving advection-diffusion problems of the following form:

$$u_t(x,t) + a \cdot \nabla u - \varepsilon \Delta u = f, \qquad (x,t) \in \Omega \times (0,T], \tag{5.0.1a}$$

$$u = g, \qquad (x,t) \in \partial\Omega \times (0,T], \tag{5.0.1b}$$

$$u(x,0) = u_0(x), \qquad x \in \Omega, \tag{5.0.1c}$$

where $\Omega \subset \mathbb{R}^d$ and $\partial\Omega$ is the boundary of Ω. Here, $a : \mathbb{R}^d \times (0,T] \to \mathbb{R}^d$ is the *velocity field*, $\varepsilon \in \mathbb{R}$ is the coefficient of diffusion ($\varepsilon > 0$), $f : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ is the *forcing function*, and $g : \partial\Omega \times (0,T] \to \mathbb{R}$ and $u_0 : \mathbb{R}^d \to \mathbb{R}$ are known functions providing the *boundary* and *initial conditions* of the problem, respectively.

The numerical solution of problems of this type has generally been approached with finite difference, finite element or boundary element methods developed for strictly-parabolic or diffusion-dominated problems—see Zerroukat et al. (2000) and the references contained therein. Each of these methods requires a mesh to be generated and maintained in order to carry out the computations, placing a significant burden on the scheme; moreover, it is well known that standard finite difference or finite element solutions of advection-dominated

diffusion problems exhibit problems with non-physical oscillations or excessive numerical diffusion.

With these concerns in mind, the development of methods which do not require a mesh, and the obvious higher-dimensional benefits that such schemes would bring, has been of increased interest in recent times. As we have seen, radial basis functions are essentially dimensionally-independent but powerful interpolatory tools, thus their successful usage would be of significant advantage in advection-diffusion problems.

Consequently, there are a variety of methods utilising radial basis functions to numerically solve advection-diffusion-type problems. Boztosun et al. (2002) describe a thin-plate spline-based scheme which uses a Crank-Nicholson method (sometimes called a θ-weighted method) to discretise in time, with thin-plate splines discretising the spatial domain by applying both the Laplacian and gradient operators to the basis functions directly, in contrast to the finite difference method with which they draw comparisons; the latter method discretises the two differential operators with the second-order central difference and backward difference, respectively.

An alternative strategy is to solve via a combination of the method of fundamental solution with the dual reciprocity method, as favoured by Ingber et al. (2004) and Chen, Kuhn, Li & Mishuris (2003), for example. Broadly speaking, the solution is assumed to comprise two functions: the solution to the corresponding homogeneous problem, and a particular solution which is approximated by radial basis functions via collocation.

Whilst the above schemes choose to discretise time using an Eulerian method, the new scheme we present in this chapter follows the lead of that described in Chapter 3, by rewriting the advection-diffusion problem (5.0.1) in Lagrangian form and discretising in time along the direction of the characteristics, giving the solution at the current time-level in terms of the solution at the previous level. We then proceed to discretise in space by defining a set of interpolation points $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ such that the points in $\mathcal{A}_1$ lie in the interior of Ω whereas those in $\mathcal{A}_2$ are wholly on $\partial\Omega$, the boundary of the domain. The value of the solution at the previous time-step is given via solving a global *Hermite interpolation* problem based on the known values at this time-level, for which we refer back

to Section 2.8; the numerical solution at the present time-level is determined from these approximate values. To this end, we first proceed to temporally discretise.

5.1 Temporal discretisation

Rewriting the advection-diffusion equation (5.0.1a) in Lagrangian form, we have

$$D_t u(x,t) - \varepsilon \Delta u(x,t) = f(x,t),$$

where $D_t u$ denotes the material derivative of u (cf. (3.1.5), page 76). Let $0 = t^0 < t^1 < \ldots < t^M < t^{M+1} = T$ be a uniform subdivision of $[0,T]$ with $t^{n+1} - t^n =: \Delta t$ for all $0 \leq n \leq M$. Integrating along the characteristic between two time-levels t^{n+1} and t^n, we have

$$\int_{t^n}^{t^{n+1}} D_t u(x,t)\,\mathrm{d}t - \varepsilon \int_{t_n}^{t^{n+1}} \Delta u(x,t)\,\mathrm{d}t = \int_{t^n}^{t^{n+1}} f(x,t)\,\mathrm{d}t,$$

that is,

$$u(x,t^{n+1}) - u(y,t^n) - \varepsilon \int_{t_n}^{t^{n+1}} \Delta u(x,t)\,\mathrm{d}t = \int_{t^n}^{t^{n+1}} f(x,t)\,\mathrm{d}t, \tag{5.1.1}$$

where $y = X(x,t^{n+1};t^n)$, and it is clear that, unlike the purely advective case (cf. (3.1.7)), we cannot localise this problem spatially due to the global nature of the Laplacian operator acting on the solution u in the second term on the left-hand side.

By applying the right-hand rectangle rule, we have that

$$\int_{t^n}^{t^{n+1}} f(x,\cdot) \approx \Delta t f(x,t^{n+1}),$$

and

$$\int_{t^n}^{t^{n+1}} \Delta u(x,\cdot) \approx \Delta t \Delta u(x,t^{n+1}).$$

As in the advection case, it will not generally be possible to precisely find the upstream point, and instead an approximation, $\tilde{y}$, to y is introduced, for which we assume $|y - \tilde{y}| \leq L_\rho \Delta t^{\rho+1}$, where $\rho > 0$ and L_ρ is a non-negative constant which is independent of the time-step Δt.

Thus, following (5.1.1), we define a semi-discrete approximation of u with respect to time, denoted $u_{\Delta t} : \Omega \times [0, T] \to \mathbb{R}$, to be the function satisfying

$$u_{\Delta t}(x, t^{n+1}) - \varepsilon \Delta t (\Delta u_{\Delta t})(x, t^{n+1}) = u_{\Delta t}(\tilde{y}, t^n) + \Delta t f(x, t^{n+1}), \qquad x \in \Omega, \tag{5.1.2a}$$

$$u_{\Delta t}(x, t^{n+1}) = g(x, t^{n+1}), \qquad x \in \partial\Omega, \tag{5.1.2b}$$

where $\tilde{y}$ is the numerical approximation to y, the upstream point at time t^n corresponding to x, and with $u_{\Delta t}(x, 0) := u_0(x)$ for all $x \in \Omega$.

Defining the operator

$$\mathcal{L} := I - \varepsilon \Delta t \Delta, \tag{5.1.3}$$

where I denotes the identity operator, we may write the scheme as

$$(\mathcal{L} u_{\Delta t})(x, t^{n+1}) = u_{\Delta t}(\tilde{y}, t^n) + \Delta t f(x, t^{n+1}), \qquad x \in \Omega,$$

$$u_{\Delta t}(x, t^{n+1}) = g(x, t^{n+1}), \qquad x \in \partial\Omega.$$

5.2 Spatial discretisation

Define a set of scattered, pairwise-distinct interpolation points

$$\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 = \{x_1, \ldots, x_I, x_{I+1}, \ldots, x_N\} \subset \Omega,$$

upon which we seek to approximate the value of $u_{\Delta t}(\cdot, t^n)$, where $\mathcal{A}_1 = \{x_1, \ldots, x_I\}$ are points in the interior of Ω and $\mathcal{A}_2 = \{x_{I+1}, \ldots, x_M\}$ are points which lie on $\partial\Omega$, the boundary of Ω.

We define $u_{\Delta t,h} : \mathcal{A} \times [0,T] \to \mathbf{R}$, the numerical approximation to u in both time and space, to be the function satisfying

$$(\mathcal{L}u_{\Delta t,h})(x,t^{n+1}) = u_{\Delta t,h}(\tilde{y},t^n) + \Delta t f(x,t^{n+1}), \qquad (x,t^{n+1}) \in \mathcal{A}_1 \times (0,T], \tag{5.2.1a}$$

$$u_{\Delta t,h}(x,t^{n+1}) = g(x,t^{n+1}), \qquad (x,t^{n+1}) \in \mathcal{A}_B \times (0,T], \tag{5.2.1b}$$

where, for all $x \in \Omega$ and all time $t \in (0,T]$, $u_{\Delta t,h}$ has the form

$$u_{\Delta t,h}(x,t) := \sum_{y \in \mathcal{A}_1} \alpha_y (\mathcal{L}\phi)(|x-y|) + \sum_{y \in \mathcal{A}_2} \beta_y \phi(|x-y|) + \sum_{j=1}^{\ell} \gamma_j p_j(x), \tag{5.2.2}$$

where $\{p_j\}_{j=1}^{\ell}$ forms a basis for Π_{k-1}^d and ϕ is a conditionally strictly positive-definite function of order k on $\mathbb{R}^d$ which permits the application of $\mathcal{L}^2$ without vanishing or creating singularities. This means in particular that it must allow Δ^2 to be applied; in the case of polyharmonic splines, this requires $2k+2\mu-d > 4$, whereas infinitely-smooth basis functions such as the multiquadrics require no such forethought in this regard.

Defining M functionals λ_x, $x \in \mathcal{A}$, by

$$\lambda_x = \begin{cases} \delta_x \circ \mathcal{L}, & x \in \mathcal{A}_1, \\ \delta_x, & x \in \mathcal{A}_2, \end{cases} \tag{5.2.3}$$

it is clear that $u_{\Delta t,h}$ has the form of the Hermite-Birkoff interpolants that we met in Section 2.8 (cf. (2.8.2) on page 70):

$$\begin{aligned} u_{\Delta t,h}(x,t) &= \sum_{y \in \mathcal{A}} \alpha_y \lambda_y^2(\phi(|x-\cdot|)) + \sum_{j=1}^{\ell} \gamma_j p_j(x) \\ &= \sum_{y \in \mathcal{A}_1} \alpha_y \lambda_y^2(\phi(|x-\cdot|)) + \sum_{y \in \mathcal{A}_2} \beta_y \lambda_y^2(\phi(|x-\cdot|)) + \sum_{j=1}^{\ell} \gamma_j p_j(x), \end{aligned}$$

where the superscript 2 on the functionals denotes action with respect to the second variable. (In fact, since the functionals involve combinations of derivatives, $u_{\Delta t,h}$ is a *Hermite interpolant.*)

In order to construct a symmetric system, we also demand that

$$\sum_{y \in \mathcal{A}} \alpha_y \lambda_y(p_j) = 0, \tag{5.2.4}$$

that is,

$$\sum_{y \in \mathcal{A}_1} \alpha_y (\mathcal{L}p_j)(y) + \sum_{y \in \mathcal{A}_2} \beta_y p_j(y) = 0, \tag{5.2.5}$$

for $j = 1 \ldots \ell$.

As with the temporal discretisation, $u_{\Delta t,h}(x,0) := u_0(x)$ for all $x \in \mathcal{A}$. From Theorem 2.8.6 we know there exists a unique solution of the form (5.2.2) for all $t \in \{t^0, \ldots, t^{n+1}\}$ subject to the conditions (5.2.1) and (5.2.5).

Remark 5.2.1. *In contrast to the numerical advection scheme* (3.1.10)*, which was defined only on the point cloud, here the function $u_{\Delta t,h}$ is globally supported.*

5.3 Analysis of the semi-discrete scheme

The error analysis for the semi-discrete approximation (5.1.2) follows in an analogous manner to the analysis carried out for the purely advective scheme (cf. Section 3.3). Firstly, we state our assumptions:

- We assume that the solution $u \in C((0,T]; C^2(\mathbb{R}^d))$ and, moreover, $u(\cdot, t)$ is Lipschitz continuous in space, that is

$$|u(x,t) - u(y,t)| \leq L_u |x - y|, \tag{5.3.1}$$

for $x, y \in \mathbb{R}^d$, $t \in (0,T]$.

- The velocity field $a \in C^1(\mathbb{R}^d \times [0,T])^d$, providing a unique solution to the backwards characteristic problem (3.1.2) and guaranteeing that the particle trajectories X are Lipschitz continuous in time—that is, there exists a positive constant L_X such that

$$|X(x,s;s) - X(x,s;t)| \leq L_X|t-s|, \tag{5.3.2}$$

  for any $x \in \mathbb{R}^d$ and $s,t \in (0,T]$.

- For the backwards characteristic problem, we further assume

$$|y - \tilde{y}| \leq L_\rho \Delta t^{\rho+1}, \tag{5.3.3}$$

  where $\tilde{y}$ denotes the approximation to the upstream point y, $\rho > 0$ and L_ρ is a non-negative constant which is independent of the time step Δt.

By (5.1.1) and (5.1.2),

$$\begin{aligned}
|u(x,t^{n+1}) - u_{\Delta t}(x,t^{n+1})| \leq & |u(y,t^n) - u(\tilde{y},t^n)| \\
& + \left|\varepsilon\left(\int_{t^n}^{t^{n+1}} \Delta u(x,t)\,\mathrm{d}t - \Delta t \Delta u(x,t^{n+1})\right)\right| \\
& + \left|\int_{t^n}^{t^{n+1}} f(x,t)\,\mathrm{d}t - \Delta t f(x,t^{n+1})\right| \\
& + \left|u(\tilde{y},t^n) - u_{\Delta t}(\tilde{y},t^n)\right|.
\end{aligned}$$

By the previous analysis (cf. (3.3.3), page 82) we know that there exists a constant $C_{f,X,n+1} = L_X \|f\|_{C(t^n,t^{n+1};C^1(\mathbb{R}^d))}$ such that

$$\left|\int_{t^n}^{t^{n+1}} f(x,t)\,\mathrm{d}t - \Delta t f(x,t^{n+1})\right| \leq C_{f,X,n+1}\Delta t^2; \tag{5.3.4}$$

by similar arguments we have

$$\left| \int_{t^n}^{t^{n+1}} \Delta u(x,t)\,\mathrm{d}t - \Delta t \Delta u(x,t^{n+1}) \right| \leq C_{u,X,n+1}\Delta t^2, \tag{5.3.5}$$

for some constant $C_{u,X,n+1}$ defined analogously to $C_{f,X,n+1}$.

Applying these bounds together with (5.3.1) and (5.3.3), then taking the maximum over all $x \in \mathbb{R}^d$, we have

$$\begin{aligned}\|u(\cdot,t^{n+1}) - u_{\Delta t}(\cdot,t^{n+1})\|_\infty \leq L_u L_\rho \Delta t^{\rho+1} + (C_{f,X,n+1} + \varepsilon C_{u,X,n+1})\Delta t^2 \\ + \|u(\cdot,t^n) - u_{\Delta t}(\cdot,t^n)\|_\infty.\end{aligned}$$

Recursively applying this bound and noting that at time $t = 0$, $\|u(\cdot,0) - u_{\Delta t}(\cdot,0)\|_\infty = 0$, gives

$$\begin{aligned}\|u(\cdot,t^{n+1}) - u_{\Delta t}(\cdot,t^{n+1})\|_\infty &\leq L_u L_\rho \Delta t^{\rho+1} + (C_{f,X,n+1} + \varepsilon C_{u,X,n+1})\Delta t^2 \\ &\quad + \|u(\cdot,t^n) - u_{\Delta t}(\cdot,t^n)\|_\infty \\ &\leq 2L_u L_\rho \Delta t^{\rho+1} + (C_{f,X,n+1} + C_{f,X,n} + \varepsilon C_{u,X,n+1} + \varepsilon C_{u,X,n})\Delta t^2 \\ &\quad + \|u(\cdot,t^{n-1}) - u_{\Delta t}(\cdot,t^{n-1})\|_\infty \\ &\vdots \\ &\leq (n+1)L_u L_\rho \Delta t^{\rho+1} + (n+1)(C_{f,X} + \varepsilon C_{u,X})\Delta t^2,\end{aligned}$$

where

$$C_{f,X} = L_X \|f\|_{C(0,T;C^1(\mathbb{R}^d))}, \qquad C_{u,X} = L_X \|u\|_{C(0,T;C^1(\mathbb{R}^d))}, \tag{5.3.6}$$

with L_X defined in (5.3.2)

Using the definition of Δt, we conclude the following lemma.

Lemma 5.3.1. *Let $u : \mathbb{R}^d \times (0,T] \to \mathbb{R}$ be the analytical solution of the advection-diffusion problem* (5.0.1) *satisfying $u \in C((0,T] \times C^2(\mathbb{R}^d))$. Let $u_{\Delta t}$ be the semi-discrete temporal*

approximation obtained by (5.1.2). *Then, the following* a priori *error bound holds:*

$$\|u(\cdot, t^{n+1}) - u_{\Delta t}(\cdot, t^{n+1})\|_\infty \leq t^{n+1}\big(L_u L_\rho \Delta t^\rho + (C_{f,X} + \varepsilon C_{u,X})\Delta t\big),$$

where L_u *and* L_ρ *are defined in* (5.3.1) *and* (5.3.3), *respectively, and* $C_{f,X}$ *and* $C_{u,X}$ *are given by* (5.3.6).

5.4 Remarks on the analysis of the full discretisation scheme

In this section we provide some comments on the issues that may arise when attempting the analysis of the numerical advection-diffusion scheme (5.2.1).

Consider the difference between the numerical scheme and the analytical solution (5.0.1) at an arbitrary point $x \in \mathcal{A}_1$ at time t^{n+1}. Using the definition of the functionals λ_x, $x \in \mathcal{A}$, and applying the theory of Franke & Schaback (1998*a*) (cf. (2.8.5) on page 72), we would have

$$|(\mathcal{L}(u - u_{\Delta t,h}))(x, t^{n+1})| = |(\lambda_x(u - u_{\Delta t,h}))(\cdot, t^{n+1})| \leq P_{\Phi,\Lambda}(\lambda_x)\|(u - u_{\Delta t,h})(\cdot, t^{n+1})\|_{\mathcal{F}},$$

where $\|\cdot\|_{\mathcal{F}}$ denotes the norm associated with the native space for ϕ and $P_{\Phi,\Lambda}$ denotes the power function depending on Φ, where $\Phi(x, y) := \phi(|x-y|)$, and $\Lambda := \{\lambda_x : x \in \mathcal{A}_1\}$. It can be shown that the power function involving $\lambda_x = \delta_x \circ \mathcal{L}$ and Φ can be reduced to the usual point-evaluation power function involving $\lambda_x^1 \lambda_y^2 \Phi(\cdot, \cdot)$. However, even given appropriate stability results for the power function and Hermite interpolation in general (see (Narcowich & Ward 1994), for example), one cannot simply telescope this bound down to $t = t^0$ due to the presence of the operator $\mathcal{L}$ on the left-hand side, and more work must be required in order to develop a useful error estimate for the scheme. This is the subject of future work we intend to carry out.

5.5 Numerical experiments for advection-diffusion problems

Here we present the results of numerical implementations of the proposed scheme to advection-diffusion problems.

5.5.1 Experiment 1: Shear flow

The first test of the computational performance of the scheme is Problem 4 from Baptista, Adams & Gresho (1995), which models the transport of a small source in a plane shear flow. Let $\Omega = (0, 24000) \times (-3400, 3400)$, $f = 0$, $a(x, y, t) = (a_0 + \alpha y, 0)^T$, where $a_0 = 0.5$ and $\alpha = 5.0 \times 10^{-4}$.

The initial condition u_0 is a point source of mass m located at $(x_0, y_0) = (7200, 0)$; however, in order to allow the numerics to begin with a finite source size, the calculations begin at $t = t_0 = 2400$, with $m = 4\pi\epsilon t_0(1 + \alpha^2 t_0^2/12)^{1/2}$.

The solution to this problem is given by

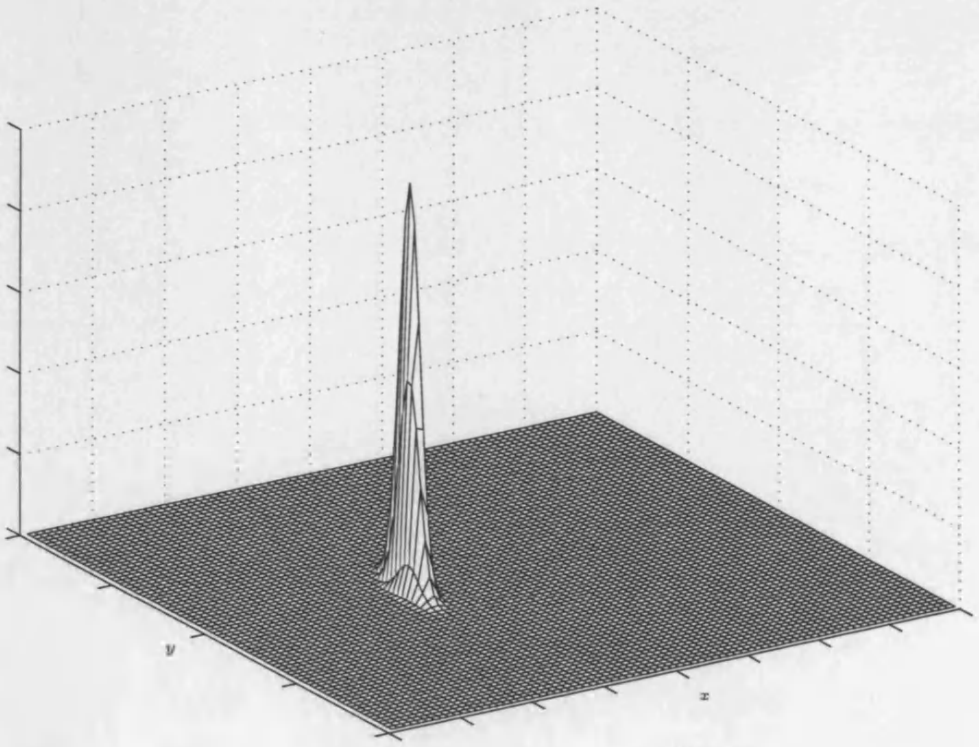
$$u(x, y, t) = \frac{m}{4\pi\epsilon t(1 + \alpha^2 t^2/12)^{1/2}} \exp(-\xi),$$

where

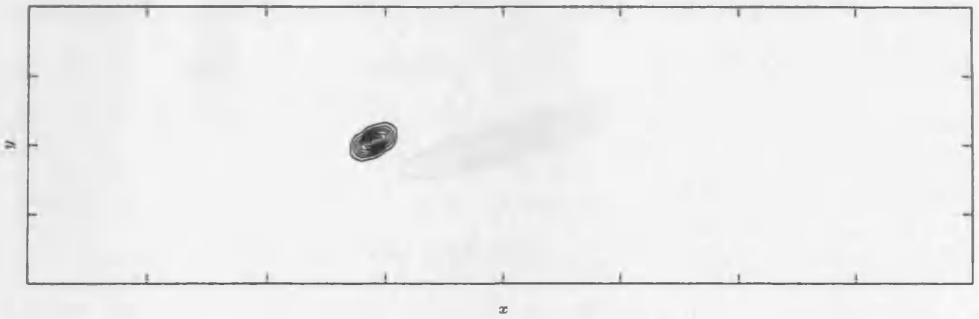
$$\xi = \frac{(x - x_0 - a_0 t - \alpha y t/2)^2}{4\epsilon t(1 + \alpha^2 t^2/12)} + \frac{y^2}{4\epsilon t}.$$

Setting $\epsilon = 10$ and $T = 9600$, we ran the algorithm over several different time-steps and resolutions, with three different radial basis functions: inverse multiquadrics, multiquadrics, and a class of function we have not met before, the *Matern family* of radial basis functions, which we will define shortly. Figure 5.1 shows the initial data (that is, time $t = t_0 = 2400$) as both a three-dimensional elevation and as a contour plot.

The analytical solution at $t = T = 9600$ is shown in Figure 5.2, again in both three dimensions and as a contour plot. For comparison purposes, we note that this problem was also attempted by Houston & Süli (2001) using an adaptive Lagrange-Galerkin method.

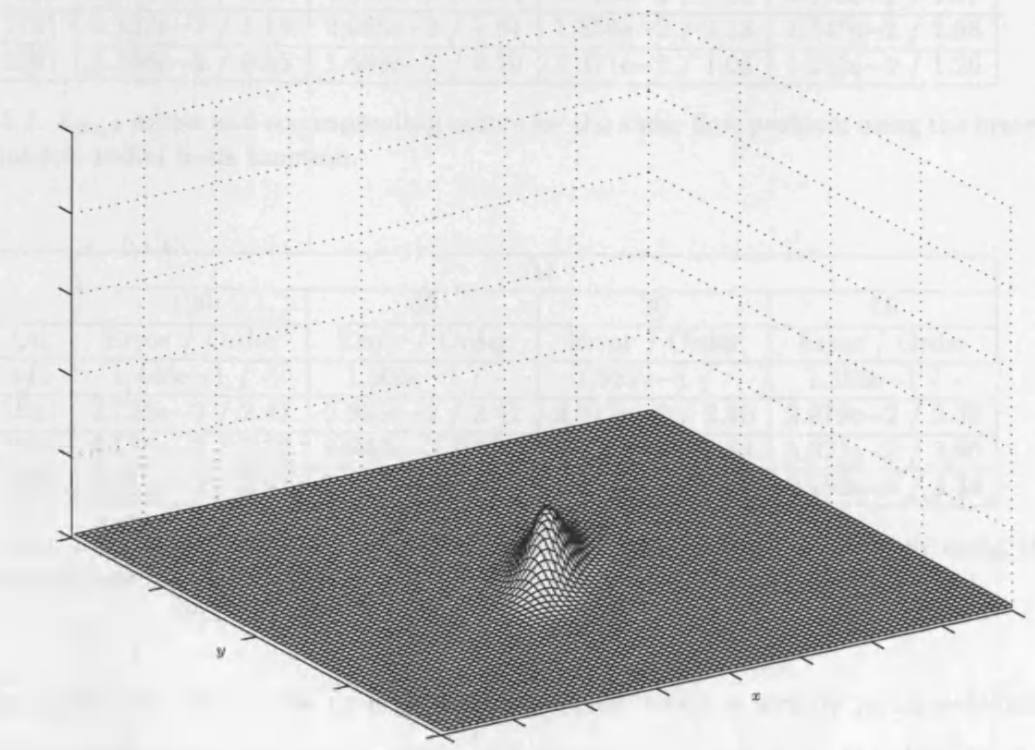


(a)

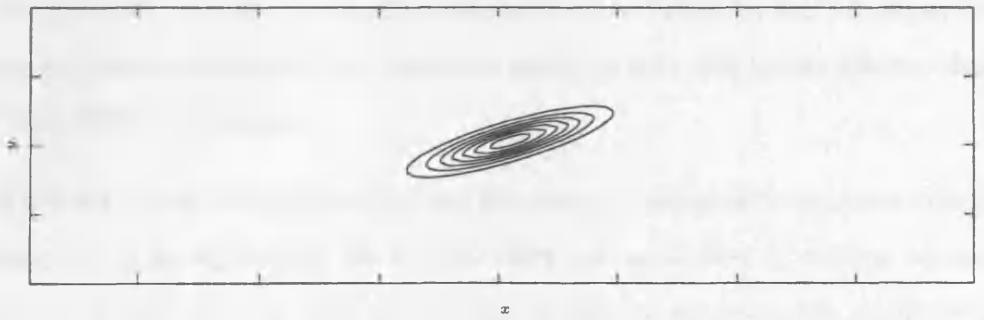


(b) Contour plot of initial data, displaying correct aspect-ratio of domain of computation.

Figure 5.1: Initial data ($t \equiv t_0 = 2400$) for the shear flow problem on the domain $[0, 24000] \times [-3400, 3400]$.



(a)



(b)

Figure 5.2: Analytical solution to the shear flow problem at time $T = 9600$ on the domain $[0, 24000] \times [-3400, 3400]$.

	Δt			
	120	60	30	15
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order	Error / Order
441	1.456e−1 / -	1.538e−1 / -	1.580e−1 / -	1.602e−1 / -
1681	3.873e−2 / 1.91	4.296e−2 / 1.84	4.488e−2 / 1.82	4.578e−2 / 1.81
3721	2.432e−2 / 1.15	2.065e−2 / 1.81	1.856e−2 / 2.18	1.747e−2 / 2.38
6561	2.266e−2 / 0.25	1.689e−2 / 0.70	1.371e−2 / 1.05	1.215e−2 / 1.26

Table 5.1: $L_{\infty,\mathcal{A}}$ errors and corresponding orders for the shear flow problem using the inverse multiquadric radial basis function.

	Δt			
	120	60	30	15
$\lvert\mathcal{A}\rvert$	Error / Order	Error / Order	Error / Order	Error / Order
441	1.446e−1 / -	1.506e−1 / -	1.537e−1 / -	1.552e−1 / -
1681	2.725e−2 / 2.41	2.833e−2 / 2.41	2.917e−2 / 2.40	2.979e−2 / 2.38
3721	6.539e−3 / 3.52	6.043e−3 / 3.81	6.031e−3 / 3.89	5.971e−3 / 3.96
6561	3.036e−3 / 2.67	2.632e−3 / 2.89	2.386e−3 / 3.22	2.285e−3 / 3.34

Table 5.2: $L_{\infty,\mathcal{A}}$ errors and corresponding orders for the shear flow problem using the multiquadric radial basis function.

The results for the inverse multiquadric, which we recall is strictly positive-definite, and for the multiquadric, which is conditionally strictly positive-definite of order 1 (and so requires a constant polynomial modifier), are presented in Tables 5.1 and 5.2, respectively. It can be seen that the multiquadric in particular performs very well in this scheme, displaying better than $\mathcal{O}(h^2)$ convergence.

Remark 5.5.1. *Both the multiquadric and the inverse multiquadric require a value for the parameter c to be set before they can be used; there are, at the time of writing, no hard-and-fast rules as to the optimum value of c in order to balance accuracy with conditioning.*

It has been shown (Madych 1992) that multiquadric interpolation becomes more accurate as c increases — however, as c gets larger the function "flattens out", which leads to ill-conditioning. Some headway has been made, by Kansa & Hon (2000) for example, in attempts to circumvent this problem, although in many applications, trial and error has been used to determine the "optimal" value for c (Fasshauer 1997). In our setting, for simplicity we choose to set $c = h$.

Figures 5.3 and 5.4 show, respectively, the numerical solution at time T for the inverse multiquadric and multiquadric at two different resolutions ($|\mathcal{A}| = 1681$ and $|\mathcal{A}| = 6561$, respectively). In both cases, the coarser grid (seen in Figures 5.3(a) and 5.4(a)) displays some oscillatory artifacts. It is clear that these artificial oscillations reduce significantly in the case of the finer grid (Figures 5.3(b) and 5.4(b)).

The method was also tested using a radial basis function from the family known as *Matern functions*. These are, it seems, less frequently-used basis functions, but it has been demonstrated by Beatson & Mouat (2002) that they can prove more effective than multiquadrics, for example, at tackling partial differential equation problems.

Definition 5.5.2 (Matern family of radial basis function). *The* Matern family *is given by*

$$\phi_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}(cr)^\nu K_\nu(cr), \tag{5.5.1}$$

where K_ν is a modified Bessel function of order $\nu > 0$ (Wendland 2005, Definition 5.10) and $c > 0$.

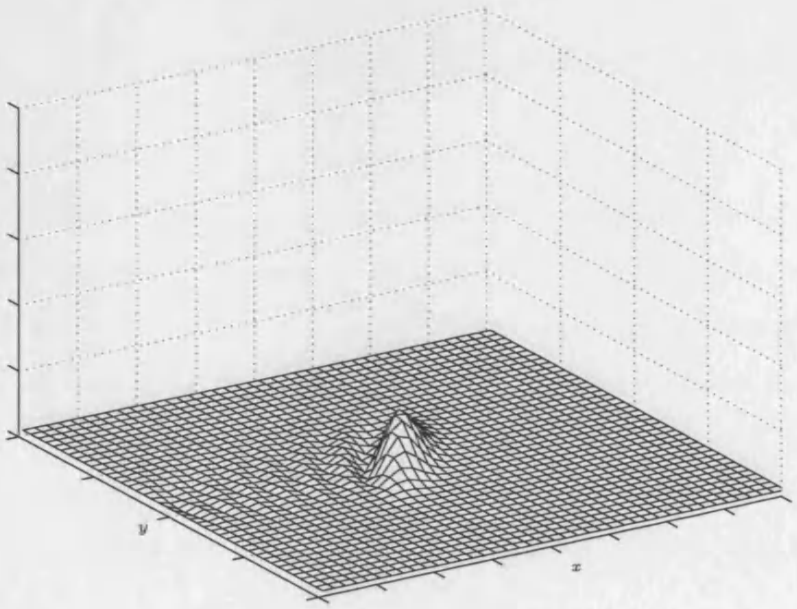
For $\nu = n \in \mathbb{Z}_+$, (5.5.1) simplifies to

$$\phi_{n+\frac{1}{2}}(r) = \frac{\exp(-cr)(cr)^n}{(2n-1)!!}\sum_{k=0}^{n}\frac{(n+k)!}{k!(n-k)!(2cr)^k},$$

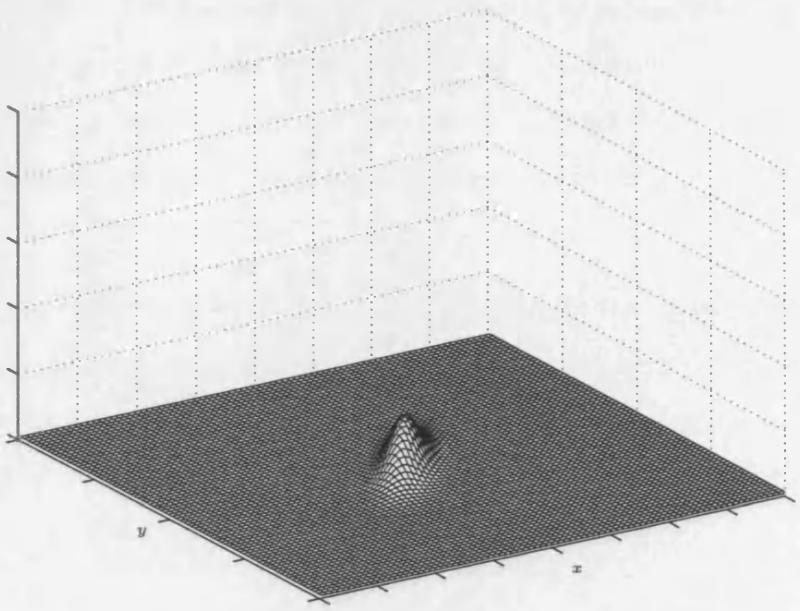
where for any $m \in \mathbb{Z}_+$,

$$m!! = \begin{cases} 1, & m = 0, \\ m\cdot(m-2)\cdots 4\cdot 2, & 0 \neq m \text{ even}, \\ m\cdot(m-2)\cdots 3\cdot 1, & m \text{ odd}. \end{cases}$$

Remark 5.5.3. *It can be shown by appealing to Fourier transform arguments that the Matern family of radial basis functions are strictly positive-definite; essentially, Matern functions are the Fourier transforms of inverse multiquadrics. Rudin (1991, Theorem 7.7) and Wendland (2005, Theorems 6.11 and 6.13) detail the salient results; see also (Beatson & Bui 2003) and the references cited therein.*

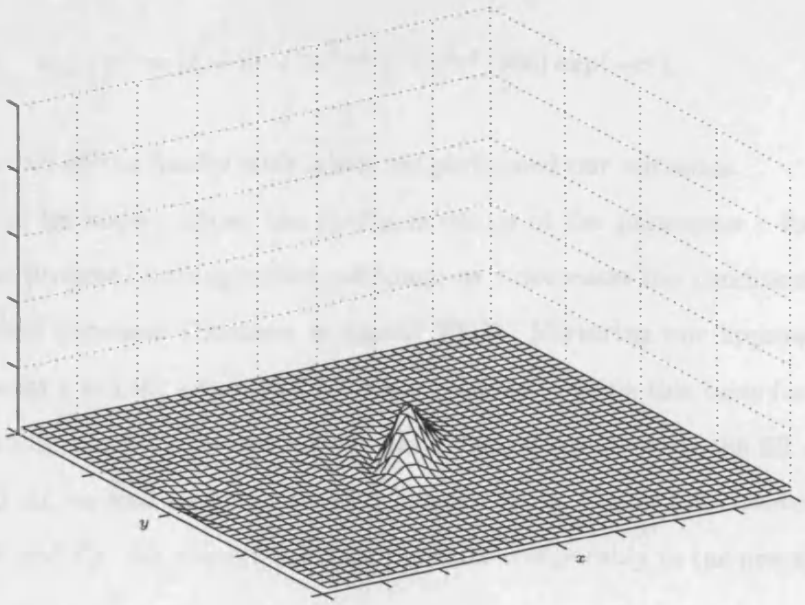


(a) 1681 regularly-spaced points.

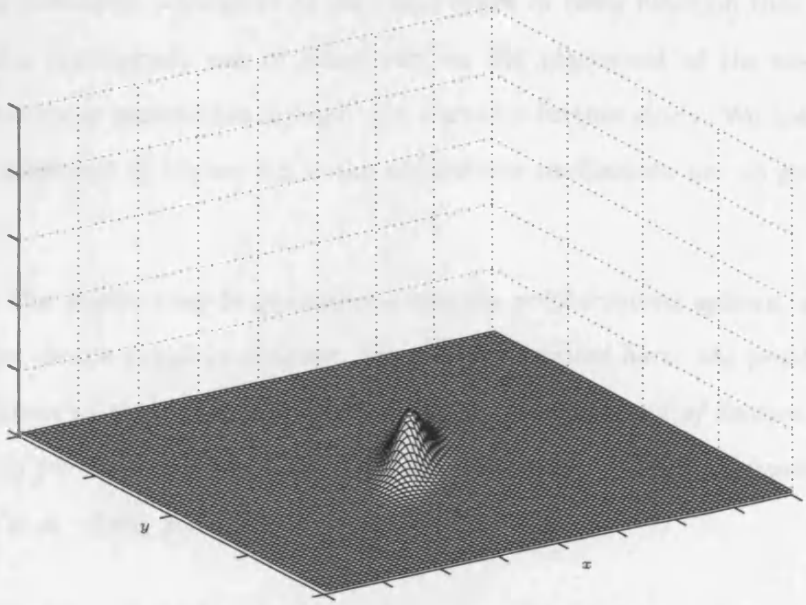


(b) 6561 regularly-spaced points.

Figure 5.3: Numerical solution for the shear flow problem using the inverse multiquadric radial basis function with $\Delta t = 15$ on the domain $[0, 24000] \times [-3400, 3400]$.



(a) 1681 regularly-spaced points.



(b) 6561 regularly-spaced points.

Figure 5.4: Numerical solution for the shear flow problem using the multiquadric radial basis function with $\Delta t = 15$ on the domain $[0, 24000] \times [-3400, 3400]$.

For $\nu = 9/2$ (that is, $n = 4$),

$$\phi_{4+\frac{1}{2}}(r) = (1 + cr + 3c^2r^2/7 + c^4r^4/105)\exp(-cr),$$

and it is this member of the family with which we performed our numerics.

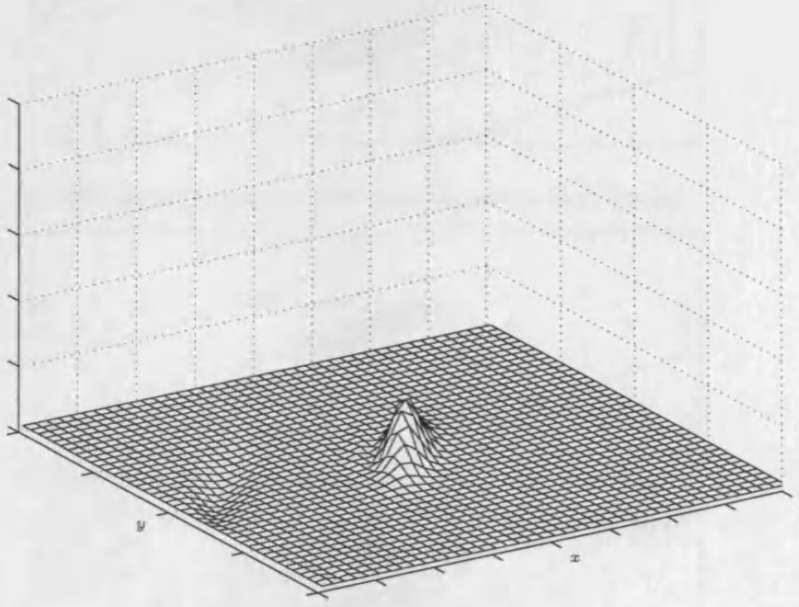
Less appears to be known about the optimum values of the parameter c for Materns functions than for (inverse) multiquadrics, although as c decreases the conditioning of the interpolation matrix increases (Beatson & Mouat 2002). Mirroring our approach for the multiquadrics, we set $c = 1/h$, although with mixed success — using this basis function, the scheme appears to be particularly sensitive to the relationship between the fill distance h and the time-step Δt, as well as the value of c. We can see in Figure 5.5 however, that for certain values of h and Δt, the Matern function performs comparably to the previously-used basis functions.

Figure 5.5.1 displays more clearly the dramatic reduction in oscillatory artifacts as the resolution of $\mathcal{A}$ is increased, whichever of the three types of basis function that we tested is being used. An appropriate use of adaptivity on the placement of the nodes would significantly reduce these oscillations further, and warrants further study. We note however that in all cases displayed in Figure 5.5.1, the superfluous oscillations are no greater than 10^{-4} in modulus.

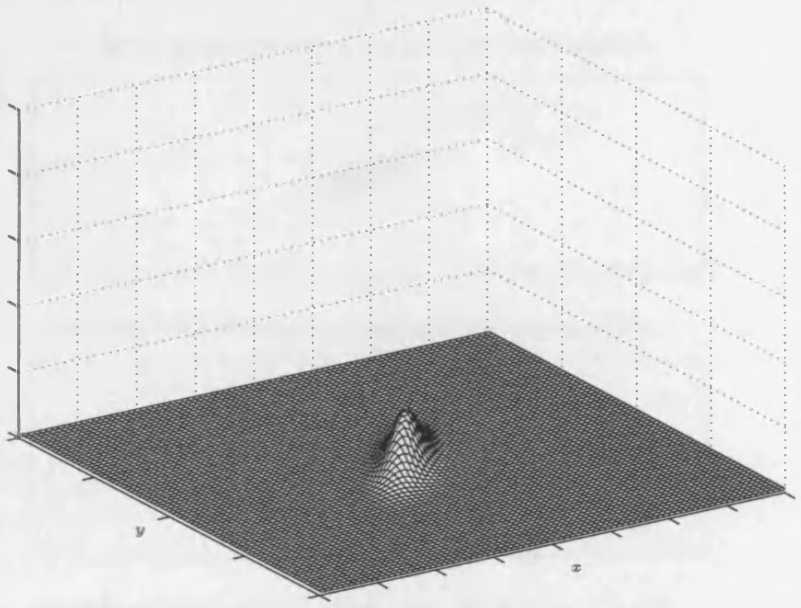
Remark 5.5.4. *The reader may be questioning why the polyharmonic splines, which were used to great effect in the previous chapter, have not been tested here: the problem lies in the double-application of the Laplacian operator. As noted at the end of Section 2.8.1, the basis function used for Hermite-Birkhoff interpolation must permit two applications of the functionals λ_x, $x \in \mathcal{A}$. Here, for $x, y \in \mathcal{A}_1$,*

$$\lambda_y^2(\lambda_x^1(\Phi)) = \Phi(x,y) - 2\varepsilon\Delta t(\Delta\Phi)(x,y) + \varepsilon^2\Delta t^2(\Delta^2\Phi)(x,y).$$

In the case of the polyharmonic spline, where for even dimensions $\Phi(x,y) \doteq |x - y|^{2k-d}\ln|x - y|$ and for odd dimensions $\Phi(x,y) \doteq |x - y|^{2k-d}$, we must choose k large

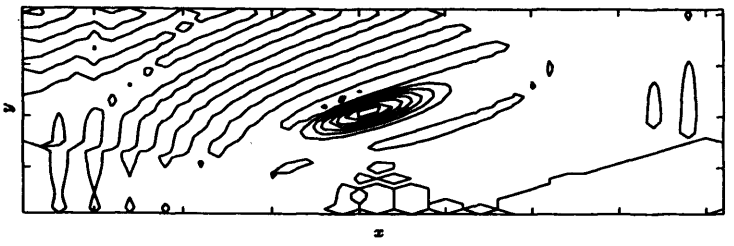


(a) 1681 regularly-spaced points.

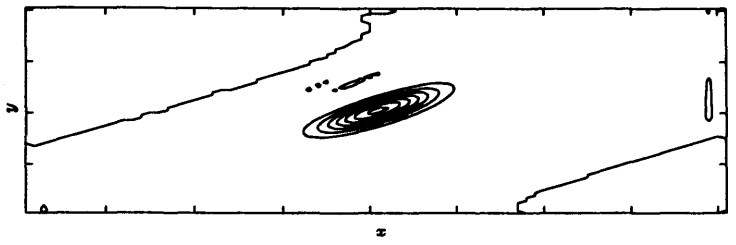


(b) 6561 regularly-spaced points.

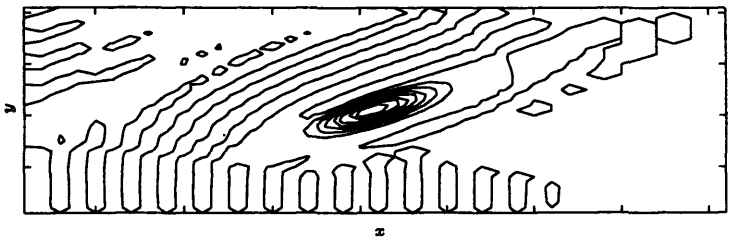
Figure 5.5: Numerical solution for the shear flow problem using the Matern radial basis function with $\nu = 9/2$ and $\Delta t = 120$ on the domain $[0, 24000] \times [-3400, 3400]$.



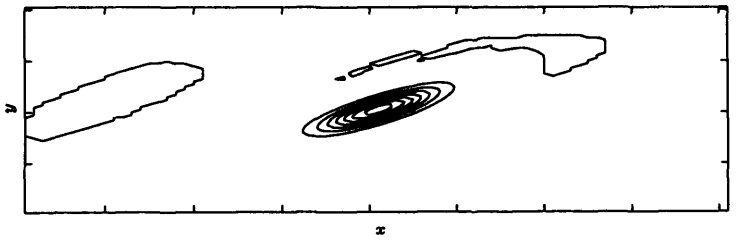
(a) 1681 regularly-spaced points using the inverse multiquadric.



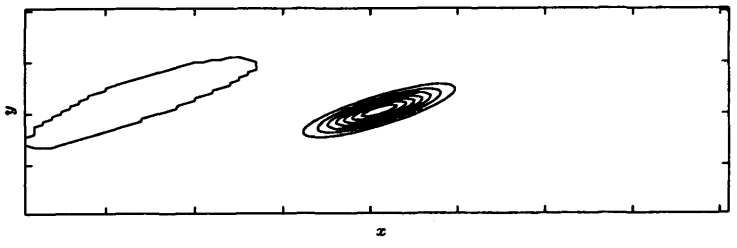
(b) 6561 regularly-spaced points using the inverse multiquadric.



(c) 1681 regularly-spaced points using the multiquadric.



(d) 6561 regularly-spaced points using the multiquadric.



(e) 6561 regularly-spaced points using the Matern, $\nu = 9/2$.

Figure 5.6: Contour plots of the numerical solution to the shear flow problem at $T = 9600$, with $\Delta t = 15$ in all but the final plot, where $\Delta t = 120$, on the domain $[0, 24000] \times [-3400, 3400]$.

enough so that $\Delta^2\Phi$ does not become discontinuous or vanish — for example, for $d = 2$ we require at least $k = 4$ — which can result in a badly-scaled interpolation matrix.

Preconditioning of the matrix — that is, performing preprocessing in order to improve its conditioning — may allow the successful use of polyharmonic splines with this scheme, and we feel this deserves further investigation as this class of basis function has the strength of requiring no additional parameters to be set.

Conditioning

As we saw in Section 2.7.1, the conditioning of the standard interpolation matrix is linked to the condition number of the matrix, and this is also the case with the Hermite-Birkhoff interpolation that our method uses.

The interpolation matrix is of the form

$$\begin{pmatrix} A_{\phi,\Lambda} & \Lambda(P) \\ \Lambda(P)^T & 0 \end{pmatrix},$$

where $A_{\phi,\Lambda} = (\lambda_a^1(\lambda_b^2(\Phi(x,y)))_{a,b\in\mathcal{A}} \in \mathbb{R}^{n\times n}$, and $\Lambda(P) = (\lambda_a(p_s))_{a\in\mathcal{A},1\leq s\leq \ell}$, with λ_a given by (5.2.3) and span $\{p_s\}_{s=1}^{\ell} = \Pi_{k-1}^d$; it is the condition number of $A_{\phi,\Lambda}$ with which we are concerned. Table 5.3 displays the condition numbers for the previously-described experiment using each of the three basis functions we tested, and suggest that we may like to seek to improve the conditioning of the method via preconditioning or with further experiments on the parameter c of each basis function.

5.5.2 Experiment 2: Rotating cylinder with small diffusion coefficient

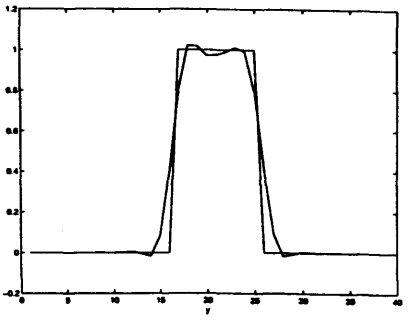
For this experiment we return to the advection-type rotating cylinder problem that we met in Chapter 4, Section 4.2, but on this occasion we treat the problem as advection-diffusion by setting $\varepsilon = 1.0 \times 10^{-8}$ and applying our new method.

As before, a time-step of $\Delta t = 0.01$ was used and the final time was set to $t \equiv T = 0.5$, corresponding to one full revolution by the cylinder. Figures 5.7 and 5.8 display cut-

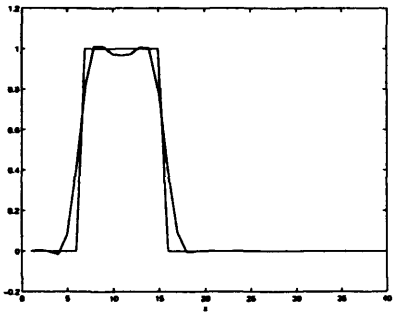
	Basis function		
	Multiquadric, $\phi(r) = (h^2 + r^2)^{1/2}$	Inverse multiquadric, $\phi(r) = (h^2 + r^2)^{-1/2}$	Matern, $\phi(r) = (1 + r/h + r^2/3h^2)e^{(-r/h)}$
Nodes	cond($A_{\phi,\Lambda}$)	cond($A_{\phi,\Lambda}$)	cond($A_{\phi,\Lambda}$)
441	1.285e+6	1.285e+6	6.095e+10
1681	1.246e+7	1.246e+7	1.264e+11
3721	6.392e+7	6.302e+7	3.653e+11
6561	1.990e+8	1.990e+8	9.709e+11

Table 5.3: Condition numbers for the shear flow problem Hermite-Birkhoff interpolation matrix (or sub-matrix) $A_{\phi,\Lambda}$ for gridded interpolation points with $\Delta t = 15$.

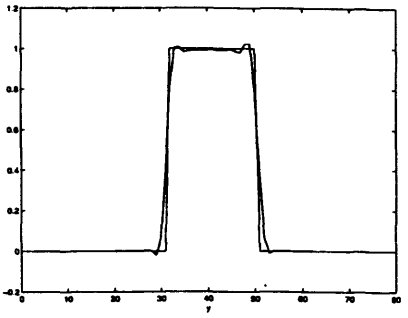
throughs of the numerical and (approximate) analytical solutions at the final time using the inverse multiquadric and the multiquadric, respectively. In both cases the basis function parameter c was set to h. The same experiment was carried out by Houston (1996, Section 5.4.2) using an adaptive Lagrange-Galerkin finite element method; comparing the profiles shown in Houston (1996, Fig. 5:9) with our own results, it is clear that, although the localised oscillations that can be seen around the base and top of the cylinder could surely be improved with adaptivity, for example, the overall profile of the numerical solution not only matches that of the analytical very well, but further demonstrates that this new method is competitive with existing methods.



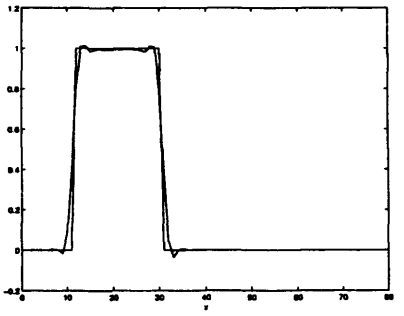
(a) $x = 0.25$ and $0 \leq y \leq 1$ on a regularly-spaced grid of 1681 points.



(b) $y = 0.5$ and $0 \leq x \leq 1$ on a regularly-spaced grid of 1681 points.

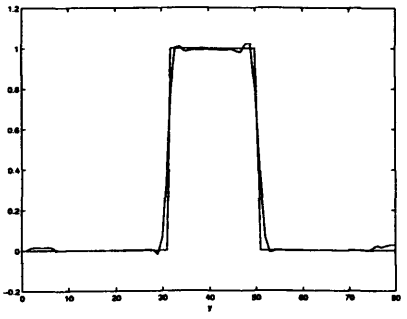


(c) $x = 0.25$ and $0 \leq y \leq 1$ on a regularly-spaced grid of 6561 points.

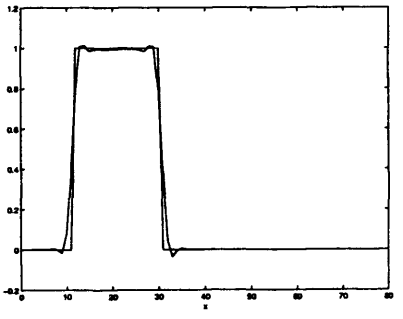


(d) $y = 0.5$ and $0 \leq x \leq 1$ on a regularly-spaced grid of 6561 points.

Figure 5.7: Cut-throughs of the numerical and analytical solutions to the advection-diffusion rotating cylinder problem with $\Delta t = 0.01$ at $t = T = 0.5$ using the inverse multiquadric basis function.



(a) $x = 0.25$, $0 \leq y \leq 1$.



(b) $y = 0.5$, $0 \leq x \leq 1$.

Figure 5.8: Cross-sections of numerical and analytical solutions for rotating cylinder problem with $\Delta t = 0.01$ at $t = T = 0.5$, using the multiquadric basis function on a regularly-space grid of 6561 points.

Chapter 6

Conclusions and Future Work

6.1 Introduction

In this chapter we summarise the results presented in this thesis, and discuss the scope for future developments of both the advection scheme analysed and tested in Chapters 3 and 4, respectively, and the advection-diffusion scheme introduced and tested in Chapter 5.

6.2 Conclusions

The focus of this thesis has been the marrying of the interpolatory power of radial basis functions with the semi-Lagrangian approach employed for the numerical approximation of advection-diffusion problems. In particular, we have proved *a priori* error estimates for the numerical scheme of Behrens & Iske (2002), which as far as we know is the first proof of convergence for this class of numerical methods.

We firstly introduced the concept of radial basis function interpolation in Chapter 2. This provided solid theoretical grounding for the global interpolation problem, including error estimates for polyharmonic splines, before proving results for *local* interpolation by polyharmonic splines: the stability of the local interpolation operator in the special case of two-dimensional thin-plate splines (Theorem 2.7.5) and in the general multi-dimensional case (Theorem 2.7.7), followed by the essential local error estimate result of Theorem 2.7.9.

The discussion was extended to consider the generalised interpolation problem of data prescribed by functionals as opposed to function values; Theorem 2.8.2 and, in particular, Theorem 2.8.6, suggest that radial basis functions could be employed for the numerical solution of linear partial differential equation problems.

The work initiated in Chapter 3 derived *a priori* error estimates for the numerical schemes (3.1.9) and (3.1.10), which is one of the central achievements of this thesis. The stability bounds, Theorems 2.7.5 and 2.7.7, were essential ingredients of the latter result.

The numerical advection scheme was validated in a series of experiments described in Chapter 4, which showed that the scheme can perform better than the analysis would predict even for "rough" solutions, providing some motivation for extending the analysis to encompass functions that do not possess at least C^1-regularity.

Chapter 5 extended the discussion to advection-diffusion problems, beginning with an outline of the advantages radial basis functions bring to this arena, followed by a brief description of some existing numerical advection-diffusion schemes based on interpolation with radial basis functions. Taking the lead from the work of Chapter 3, we proceeded to describe, in Sections 5.1 and 5.2, a new semi-Lagrangian radial basis function method for advection-diffusion problems, based on the Hermite-Birkhoff interpolation of Section 2.8.

In Section 5.3, the semi-discrete numerical scheme (5.1.2) was analysed in an analogous manner to the advection scheme (3.1.9), the result of which is summarised in Lemma 5.3.1. Section 5.4 describes issues pertaining to the analysis of the fully-discrete scheme (5.2.1).

Finally in this chapter, we validated the scheme with two numerical experiments: firstly, a shear flow problem detailed in Section 5.5.1, for which an analytical solution is available, using three different radial basis functions (inverse multiquadric, multiquadric, and Matern) and describing the issues surrounding the possible use of polyharmonic splines with this scheme. The conditioning of the matrix for the scheme was investigated in the context of this problem by examining the condition numbers of the interpolation matrix.

For the second numerical experiment we returned to the rotating cylinder problem of Section 4.2, but with a small diffusion coefficient added in order to apply the new scheme using both the inverse multiquadric and the multiquadric basis functions.

6.3 Future work

In this section we outline possible directions in which further study could be carried out, based on the work contained in this thesis.

6.3.1 Analysis of the linear advection scheme

The main result of Chapter 3, Theorem 3.2.2, proves convergence of the numerical scheme (3.1.10) provided that the solution u is of at least C^1-regularity in space. In order to show convergence for a wider class of functions, it would be necessary to relax this restriction whilst still retaining control of the interpolation error. We have seen that the difference between the analytical and numerical solutions can be written as follows (cf. (3.4.1)):

$$\begin{aligned}
|u(x,t^{n+1}) - u_{\Delta t,h}(x,t^{n+1})| \leq\ & |u(y,t^n) - u(\tilde{y},t^n)| \\
& + |u(\tilde{y},t^n) + (\mathcal{I}_{\mathcal{N}}u(\,\cdot\,,t^n))(\tilde{y})| \\
& + |\mathcal{I}_{\mathcal{N}}(u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n))(\tilde{y})| \\
& + \left| \int_{t^n}^{t^{n+1}} f(X(x,t^{n+1};t),t)\,\mathrm{d}t - \Delta t f(x) \right| \\
\leq\ & L_u L_\rho \Delta t^{\rho+1} \\
& + |u(\tilde{y},t^n) + (\mathcal{I}_{\mathcal{N}}u(\,\cdot\,,t^n))(\tilde{y})| \\
& + |\mathcal{I}_{\mathcal{N}}(u(\,\cdot\,,t^n) - u_{\Delta t,h}(\,\cdot\,,t^n))(\tilde{y})| \\
& + C_{f,X,n+1}\Delta t^2,
\end{aligned}$$

where the coefficients and constants are as in Lemma 3.2.1 and, as before, $x \in \mathcal{A}$, $y = X(x,t^{n+1};t^n)$ is the upstream point of x, and $\tilde{y}$ is the numerical approximation to y.

Thus a bound is required on the second and third terms on the right hand side. Naturally one could choose from the plethora of radial basis function error estimates (see the books of Buhmann (2003), Cheney & Light (2000) and Wendland (2005), for example), and in our analysis we opted to utilise explicitly the smoothness of the underlying solution u, assuming

it possessed at least C^1-regularity. But the less we can assume about the regularity of the underlying solution, the better. Brownlee & Light (2004) have already taken the first steps in proving error estimates for interpolation of rough functions, by providing bounds on the error when the interpoland lies outside the native space of the interpolant, in the case of interpolation by polyharmonic splines. The incorporation of their results would be a significant step forward in the analysis of schemes of this class, as would similar bounds for functions outside the native spaces of the other commonly-used radial basis functions such as the multiquadrics and Gaussians (it is recognised that the latter basis function in particular does not enjoy a particularly rich native space).

However, it is clear that if such results are to be utilised, the stability of the interpolation operator, seen in the third term above, is of central importance to this endeavour. Based on our stability result for the thin-plate spline in two dimensions, Theorem 2.7.5, we assumed the associated constant to be of size $1 + \mathcal{O}(\Delta t)$ in general, as it is via this term that we telescope the bound back to $t = t^0$. Proof that the constant is indeed of the assumed size in the case of interpolation via polyharmonic splines would validate this assumption.

The *a posteriori* error analysis for this scheme is also a possible avenue for further research, which would involve firstly rewriting the scheme in a variational setting.

6.3.2 The advection-diffusion scheme

The numerical advection-diffusion scheme defined by (5.2.1), (5.2.2) and (5.2.5) has many important applications, in particular to problems of high dimension that are common in financial applications, and polymeric flows, for example. In these latter problems, the Fokker-Planck equation is coupled with the incompressible Navier-Stokes equations. For treating the lower-dimensional Navier-Stokes problem, finite element methods have a proven record of accuracy and stability (Girault & Raviart 1986), whereas for the higher-dimensional Fokker-Planck component, the strengths of radial basis functions could be called into play via our mesh-less numerical scheme, resulting in combination of these two numerical methods.

In this view, improvements to the efficiency and accuracy of our scheme would be an important direction to consider, and there are a variety of possibilities: the areas of pre-conditioning of the interpolation matrix and localising the problem (or decomposing the computational domain) would both induce greater efficiency on the scheme if it can be demonstrated that they can be applied effectively—see (Beatson & Light 1997, Beatson, Light & Billings 2000, Mateescu, Ribbens & Watson 2003) for possible starting-points. From a numerical-accuracy perspective, sensible adaptivity of the point cloud—placing more nodes in regions of steep gradient, for example, or areas where the solution is more complex—would be of great benefit, and it is but for the constraints of time that we were not able to include such a step ourselves in the testing of the scheme.

As noted in Remark 5.5.1 and the comments following Remark 5.5.3, the use of (inverse) multiquadrics and Matern functions requires careful control or selection of the parameter c; for the shear flow experiment in particular, multiquadrics and Matern functions displayed promising results. Polyharmonic splines require no additional parameters to be set, and this strength together with the results of Brownlee & Light (2004) described above provide justification for further investigation into their application in this scheme.

When using globally-supported radial basis functions such as those just mentioned, the interpolation matrices of the scheme are full, so it may be worth testing the effectiveness of the *compactly supported* basis functions of Wendland (2005, Chapter 9). Continuous compactly-supported functions enjoy the property that, if they are conditionally positive-definite of order m, then $m = 0$—that is, they are positive-definite. But it is well known that there exists a trade-off between the accuracy of an interpolant and the conditioning of the matrix (the so-called *uncertainty principle* coined by Schaback (1995)), so whilst the compactly-supported functions may generate sparser matrices than their globally-supported cousins, in order to maintain a required degree of accuracy it is necessary to keep the support fixed as the fill distance tends to zero, which will obviously affect the conditioning.

Aside from these computational-efficiency-focussed directions there are of course the *a priori* and *a posteriori* error analyses of the scheme, both of which are yet to be attempted

directly and, as far as is known, are not easy consequences of work already in the established literature. Section 5.4 described some issues concerning the analysis of this scheme; to this we add that, similarly to the advection case, error estimates for Hermite (or Hermite-Birkhoff) interpolation of functions that do not lie in the native space of the interpolant could be of considerable advantage in this endeavour.

Appendix A

Nearest Neighbour Algorithms

In this appendix we enter into a brief discussion of nearest neighbour algorithms in the context of the analysis of the linear advection scheme presented in Chapter 3; whilst the details are not vital as far as our analysis is concerned, the process is of crucial importance when implementing the scheme, for the choice of a slow nearest neighbour algorithm could seriously inhibit the overall computational efficiency of the method.

The size of the neighbour set is an important choice, independent of the method used to determine its members. As we saw in Chapter 4, choosing a large set would increase the accuracy of the interpolant which determines the value of the solution at $\tilde{y}$, but this adds to the computational expense of the scheme, as each interpolatory step is solved for a larger interpolation matrix. On the other hand, too small a set of neighbours would adversely affect the error of the interpolant. For our purposes the choice of $|\mathcal{N}(x)|$ is left somewhat arbitrary, but we assume it is fixed over all $x \in \Xi$ for all time.

Behrens & Iske (2002) provide one possible method for determining the set of nearest neighbours. Firstly, the nodes of Ξ are used to construct a triangulation $\mathscr{T}$ of Ω. Strictly speaking it is $\tilde{y}$, the numerical approximation, to which we should be determining the nearest neighbours, but without loss of generality we can consider $y = X(x, t^{n+1}; t^n)$. Locate the triangle $T \in \mathscr{T}$ containing y, and let $\mathcal{N}(x) = \bigcup_{v \in \mathscr{V}_T} \mathscr{C}_v$, where $\mathscr{V}_T$ is the set of three vertices of the triangle T, and for any vertex $v \in \mathscr{T}$, $\mathscr{C}_v$ is the vertex set of its cell.

An alternative method has been described by Mielikäinen (n.d.), which has the strength of being exhaustive whilst being significantly faster than a naïve "brute force" approach.

Consider a set $\Xi = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ and a point $y \in \mathbb{R}^d$, and suppose we wish to find the nearest ℓ neighbouring points. The brute force method would be to compute the distances $|y - x_i|$ for $i = 1, \ldots, N$, sort in ascending order and pick the first ℓ elements. Now suppose, as in our advection algorithm, we need to find the nearest neighbours from Ξ for not just one point, but many—say, all members of $\{y_1, \ldots, y_M\}$ for some M not necessarily equal to N. For each $j = 1, \ldots, M$ we would need to compute all $|y_j - x_i|$, $i = 1, \ldots, N$. Clearly this is highly inefficient and thus finding a fast algorithm is crucial to the performance of our scheme.

A.0.1 Fast exhaustive nearest neighbour algorithm

In order to speed up the process, we perform some preprocessing on $\Xi = \{x_1, \ldots, x_N\}$. Let D be the $N \times N$ matrix whose entries are given by

$$D(i, j) = |x_i - x_j|$$

for $i, j = 1, \ldots, N$. That is, D is the matrix of all distances between the points in Ξ. Since this matrix is symmetric with zeros along the diagonal, we need only perform $\sum_{i=1}^{N-1} i$ calculations. For each y_j we now proceed as follows.

Let nn be a $l \times 3$ matrix whose rows are given by

$$nn(i, :) = [i, x_i, |y_j - x_i|]$$

for $i = 1, \ldots, l$. The second column of this matrix is our ℓ-tuple of candidate nearest neighbours, in the first instance comprising the first ℓ elements in Ξ; obviously it is highly unlikely this vector will be the correct set of nearest neighbours to y_j, so we need to update it as described in Algorithm 1.

Algorithm 1 Update the candidate nearest-neighbour matrix nn.

Require: $y \in \mathbb{R}^d$, $D \in \mathbb{R}^{N \times N}$, $\Xi = \{x_1, \ldots, x_N\}$
  for $i = 1 : \ell$ **do**
    $nn(i,:) = [i, x_i, |y - x_i|]$
  end for
  Sort nn in ascending order of $nn(:,3)$.
  $N_n = nn(\ell, 1)$
  $D_n = nn(\ell, 3)$
  for $i = \ell + 1 : N$ **do**
    $d_e = |D(i, N_n) - D_n|$
    if $d_e \leq D_n$ **then**
      $d_a = |y_j - x_i|$
      if $d_a < D_n$ **then**
        $nn(\ell,:) = [i, x_i, d_a]$
        Sort the rows of nn in ascending order of column 3
        $N_n = nn(\ell, 1)$
        $D_n = nn(\ell, 3)$
      end if
    end if
  end for
  return nn

One may convince oneself this is faster and cheaper than brute force because we perform simple look-ups (of the values stored in D) and subtractions, and only perform the more expensive norm calculation when we suspect we have a better candidate. The matrix D can be stored and used for as long as the set Ξ remains unchanged — if there is no adapting of the nodes, D need only be computed once.

A.0.2 Comparison of fast full-search equivalent algorithm with brute force search

We performed a simple comparison to demonstrate that, although the algorithm just described is exhaustive, it performs much better than simple brute force for larger values of $|\Xi|$ in terms of computational speed. For each point $x \in \{x_1, \ldots, x_N\} \subset \mathbb{R}^2$, we find the nearest 5 neighbours (treating Ξ as a periodic set). Table A.1 displays the CPU times for each algorithm against N; these results are plotted in Figure A.1. For a fair comparison we have included the pre-processing time it takes to generate the matrix D, hence for small sets of points, brute force out-performs the Mielikäinen method.

Moreover, it is clear that if, alternatively, Ξ is contained in some ball $B \subset \mathbb{R}^d$ and $\chi \subset B$ with $|\chi| \gg |\Xi|$, the algorithm will perform at a much greater rate of efficiency than brute force at finding the nearest neighbours from Ξ to each point in χ.

A.0.3 Alternative nearest neighbour algorithms

There are several popular data structures that one may impose on the point cloud Ξ in order to be able to quickly generate nearest-neighbour sets, and the reader is directed to (Wendland 2005, Chapter 14), and the references cited therein, for further discussion.

N	T_{BF}	T_M
8	0.12	0.13
16	0.12	0.24
32	0.34	0.43
64	0.97	1.26
128	3.75	3.85
256	14.81	12.29
512	59.22	41.14
1024	235.45	143.94
2048	939.04	522.30
4096	3754.39	1980.01

Table A.1: CPU time for the brute force algorithm (T_{BF}) and the method of Mielikäinen (T_M).

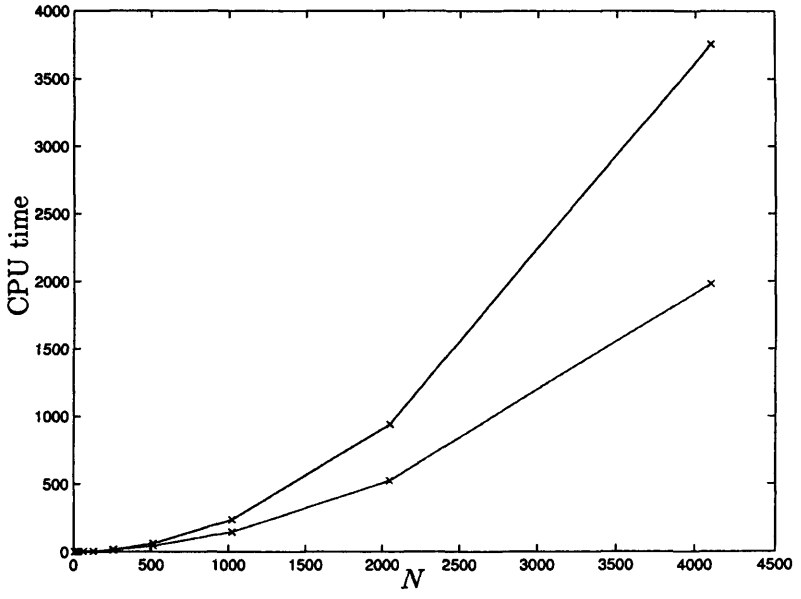


Figure A.1: Plot of the data in Table A.1, showing that Mielikäinen's algorithm rapidly out-performs the brute-force method.

Bibliography

Adams, R. A. (1978), *Sobolov Spaces*, Academic Press, San Diego.

Baptista, A. M., Adams, E. E. & Gresho, P. (1995), Benchmarks for the transport equation: The convection-diffusion forum and beyond, *in* Lynch & Davies, eds, 'Quantitative Skill Assessment for Coastal Ocean Models', Vol. 47, AGU Coastal and Estuarine Studies, pp. 241–268.

Beatson, R. K. & Bui, H.-Q. (2003), Mollification formulas and implicit smoothing.

Beatson, R. K. & Light, W. A. (1997), 'Fast evaluation of radial basis functions: methods for two-dimensional polyharmonic splines', *IMA J. Numer. Anal.* **17**(3), 343–372.

Beatson, R. K., Light, W. A. & Billings, S. (2000), 'Fast solution of the radial basis function interpolation equations: domain decomposition methods', *SIAM J. Sci. Comput.* **22**(5), 1717–1740 (electronic).

Beatson, R. K. & Mouat, C. T. (2002), RBF collocation.

Behrens, J. & Iske, A. (2002), 'Grid-free adaptive semi-lagrangian advection using radial basis functions', *Computers and Mathematics with Applications* **43:3—5**, 319—327.

Behrens, J., Iske, A. & Käser, M. (2003), Adaptive meshfree method of backward characteristics for nonlinear transport equations, *in* 'Meshfree methods for partial differential equations (Bonn, 2001)', Vol. 26 of *Lect. Notes Comput. Sci. Eng.*, Springer, Berlin, pp. 21–36.

Bercovier, M. & Pironneau, O. (1982), Characteristics and the finite element method, *in* 'Finite element flow analysis; Proceedings of the Fourth International Symposium on Finite Element Methods in Flow Problems, Tokyo, Japan, July 26-29, 1982 (A84-25833 10-34)', University of Tokyo Press/North-Holland Publishing Co., Tokyo/Amsterdam, pp. 67–73.

Bermejo, R. (1990), 'On the equivalence of semi-Lagrangian schemes and particle-in-cell finite element methods', **118**, 979–987.

Boztosun, I., Charafi, A., Zerroukat, M. & Djidjeli, K. (2002), 'Thin-plate spline radial basis function scheme for advection-diffusion problems', *Electronic Journal of Boundary Elements* (2), 267–282.

Brownlee, R. A. & Light, W. A. (2004), 'Approximation orders for interpolation by surface splines to rough functions', *IMA J. Numer. Anal.* **24**(2), 179–192.

Buhmann, M. D. (2003), *Radial basis functions: theory and implementations*, Vol. 12 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge.

Chen, C. S., Kuhn, G., Li, J. & Mishuris, G. (2003), 'Radial basis functions for solving near singular Poisson problems', *Comm. Numer. Methods Engrg.* **19**(5), 333–347.

Cheney, E. W. (1966), *Introduction to approximation theory*, McGraw-Hill Book Co., New York.

Cheney, E. W. & Light, W. A. (2000), *A Course in Approximation Theory*, Brooks/Cole Series in Advanced Mathematics, Brooks/Cole, New York.

Douglas, Jr., J. & Russell, T. F. (1982), 'Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures', *SIAM J. Numer. Anal.* **19**(5), 871–885.

Dubal, M. R. (1994), 'Domain decomposition and local refinement for multiquadric approximations. I. Second-order equations in one-dimension', *J. Appl. Sci. Comput.* **1**(1), 146–171.

Dubal, M. R., Oliveira, S. R. & Matzner, R. A. (1992), Solution of elliptic equations in numerical relativity using multiquadrics, *in* 'Approaches to numerical relativity (Southampton, 1991)', Cambridge Univ. Press, Cambridge, pp. 265–280.

Duchon, J. (1978), 'Sur l'erreur d'interpolation des fonctions de plusieurs variables par les d^m-splines', *R.A.I.R.O. Anal. Numér.* **12**(4), 325–334.

Falcone, M. & Ferretti, R. (1998), 'Convergence analysis for a class of high-order semi-lagrangian advection schemes', *SIAM J. Numer. Anal.* **35**(3), 909–940.

Fasshauer, G. E. (1997), Solving partial differential equations by collocation with radial basis functions, *in* A. Le Mehaute, C. Rabut & L. L. Schumaker, eds, 'Surface Fitting and Multiresolution Methods', Vanderbilt University Press, pp. 131–138.

Franke, C. & Schaback, R. (1998*a*), 'Convergence order estimates of meshless collocation methods using radial basis functions', *Advances in Computational Mathematics* **8**(4), 381–399.

Franke, C. & Schaback, R. (1998*b*), 'Solving partial differential equations by collocation using radial basis functions', *Appl. Math. Comput.* **93**(1), 73–82.

Girault, V. & Raviart, P.-A. (1986), *Finite element methods for Navier-Stokes equations*, Vol. 5 of *Springer Series in Computational Mathematics*, Springer-Verlag, Berlin. Theory and algorithms.

Golub, G. H. & Van Loan, C. F. (1996), *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, third edn, Johns Hopkins University Press, Baltimore, MD.

Guo, K., Hu, S. & Sun, X. (1993), 'Conditionally positive definite functions and Laplace-Stieltjes integrals', *J. Approx. Theory* **74**(3), 249–265.

Hon, Y. C. & Schaback, R. (2001), 'On unsymmetric collocation by radial basis functions', *Appl. Math. Comput.* **119**(2-3), 177–186.

Houston, P. (1996), Lagrange-Galerkin Methods for Unsteady Convection-Diffusion Problems: A Posteriori Error Analysis and Adaptivity, Ph.d. Thesis, University of Oxford.

Houston, P. & Süli, E. (2001), 'Adaptive Lagrange-Galerkin methods for unsteady convection-diffusion problems', *Math. Comp.* **70**(233), 77–106.

Ingber, M. S., Chen, C. S. & Tanski, J. A. (2004), 'A mesh free approach using radial basis functions and parallel domain decomposition for solving three-dimensional diffusion equations', *Internat. J. Numer. Methods Engrg.* **60**(13), 2183–2201.

Iske, A. (1995), Reconstruction of functions from generalized Hermite-Birkhoff data, *in* 'Approximation theory VIII, Vol. 1 (College Station, TX, 1995)', Vol. 6 of *Ser. Approx. Decompos.*, World Sci. Publishing, River Edge, NJ, pp. 257–264.

Iske, A. (2003*a*), On the approximation order and numerical stability of local Lagrange interpolation by polyharmonic splines, *in* 'Modern developments in multivariate approximation', Vol. 145 of *Internat. Ser. Numer. Math.*, Birkhäuser, Basel, pp. 153–165.

Iske, A. (2003*b*), 'Radial basis functions: basics, advanced topics and meshfree methods for transport problems', *Rend. Sem. Mat. Univ. Politec. Torino* **61**(3), 247–285 (2004). Splines, radial basis functions and applications.

Kansa, E. J. (1990*a*), 'Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics. I. Surface approximations and partial derivative estimates', *Comput. Math. Appl.* **19**(8-9), 127–145.

Kansa, E. J. (1990*b*), 'Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics. II. Solutions to parabolic, hyperbolic and elliptic partial differential equations', *Comput. Math. Appl.* **19**(8-9), 147–161.

Kansa, E. J. & Hon, Y. C. (2000), 'Circumventing the ill-conditioning problem with multiquadric radial basis functions: applications to elliptic partial differential equations', *Computers and Mathematics with Applications* **39**, 123–137.

Lebedev, N. N. (1965), *Special functions and their applications*, Revised English edition. Translated and edited by Richard A. Silverman, Prentice-Hall Inc., Englewood Cliffs, N.J.

Lefschetz, S. (1977), *Differential equations: geometric theory*, second edn, Dover Publications, Inc., New York.

Levesley, J. & Light, W. A. (1999), 'Direct form seminorms arising in the theory of interpolation by translates of a basis function', *Adv. Comput. Math.* **11**(2-3), 161–182.

Light, W. A. & Wayne, H. (1998), 'On power functions and error estimates for radial basis function interpolation', *J. Approx.* **92**(2), 245–265.

Light, W. A. & Wayne, H. (1999), 'Spaces of distributions, interpolation by translates of a basis function and error estimates', *Numer. Math.* **81**(3), 415–450.

Light, W., ed. (1992), *Advances in numerical analysis. Vol. II*, Oxford Science Publications, The Clarendon Press Oxford University Press, New York. Wavelets, subdivision algorithms, and radial basis functions.

Lorentz, R. A., Narcowich, F. J. & Ward, J. D. (2003), 'Collocation discretizations of the transport equation with radial basis functions', *Appl. Math. Comput.* **145**(1), 97–116.

Madych, W. R. (1992), 'Miscellaneous error bounds for multiquadric and related interpolators', *Computers and Mathematics with Applications* **24**, 121–138.

Mairhuber, J. C. (1956), 'On Haar's theorem concerning Chebychev approximation problems having unique solutions', *Proc. Amer. Math. Soc.* **7**, 609–615.

Mateescu, G., Ribbens, C. J. & Watson, L. T. (2003), 'A domain decomposition preconditioner for Hermite collocation problems', *Numer. Methods Partial Differential Equations* **19**(2), 135–151.

Micchelli, C. A. (1986), 'Interpolation of scattered data: Distance matrices and conditionally positive definite functions', *Constr. Approx.* **2**, 11–22.

Mielikäinen, J. (n.d.), Fast full-search equivalent nearest-neighbor search algorithm.

Moridis, G. J. & Kansa, E. J. (1994), 'The Laplace transform multiquadrics method: a highly accurate scheme for the numerical solution of linear partial differential equations', *J. Appl. Sci. Comput.* **1**(2), 375–407.

Morton, K. W. & Mayers, D. F. (2005), *Numerical solution of partial differential equations*, second edn, Cambridge University Press, Cambridge. An introduction.

Morton, K. W., Priestley, A. & Süli, E. (1988), 'Stability of the Lagrange-Galerkin method with nonexact integration', *RAIRO Modél. Math. Anal. Numér.* **22**(4), 625–653.

Morton, K. W. & Süli, E. (1995), 'Evolution-Galerkin methods and their supraconvergence', *Numer. Math.* **71**(3), 331–355.

Narcowich, F. J. & Ward, J. D. (1994), 'Generalized Hermite interpolation via matrix-valued conditionally positive-definite functions', *Math. Comp.* **63**(208), 661–687.

Pironneau, O. (1981/82), 'On the transport-diffusion algorithm and its applications to the Navier-Stokes equations', *Numer. Math.* **38**(3), 309–332.

Priestley, A. (1994), 'Exact projections and the Lagrange-Galerkin method: A realistic alternative to quadrature', **112**, 316–333.

Robert, A. (1981), 'A stable numerical integration scheme for the primitive meteorological equations', *Atmosphere—Ocean* **19**(1), 35–46.

Rudin, W. (1991), *Functional Analysis*, second edn, McGraw Hill, New York.

Schaback, R. (1995), 'Error estimates and condition numbers for radial basis function interpolation', *Adv. Comput. Math.* **3**(3), 251–264.

Schaback, R. (2005), Convergence of unsymmetric kernel-based meshless collocation methods.

Sharan, M., Kansa, E. J. & Gupta, S. (1997), 'Application of the multiquadric method for numerical solution of elliptic partial differential equations', *Appl. Math. Comput.* **84**(2-3), 275–302.

Süli, E. & Ware, A. (1991), 'A spectral method of characteristics for hyperbolic problems', *SIAM J. Numer. Anal.* **28**(2), 423–445.

Sun, X. (1994), 'Scattered Hermite interpolation using radial basis functions', *Linear Algebra Appl.* **207**, 135–146.

Wayne, H. (1996), Towards a Theory of Multivariate Interpolation Using Spaces of Distributions, Ph.d Thesis, University of Leicester.

Wendland, H. (2005), *Scattered Data Approximation*, Vol. 17 of *Cambridge Monographs on Applied and Computational Mathematics*, Cambridge University Press, Cambridge.

Wu, Z. M. (1992), 'Hermite-Birkhoff interpolation of scattered data by radial basis functions', *Approx. Theory Appl.* **8**(2), 1–10.

Zerroukat, M., Djidjeli, K. & Charafi, A. (2000), 'Explicit and implicit meshless methods for linear advection-diffusion-type partial differential equations', *Internat. J. Numer. Methods Engrg.* **48**(1), 19–35.

Zerroukat, M., Power, H. & Chen, C. S. (1998), 'A numerical method for heat transfer problems using collocation and radial basis functions', *International Journal for Numerical Methods in Engineering* **42**(7), 1263–1278.